Extended Handover Keying and Modified IEEE 802.21 Resource Query Approach for Improving Vertical Handoff Performance

Ahmed H. Zahran and Cormac J. Sreenan
Computer Science Dept.
University College Cork
Cork, Ireland.
Email: [a.zahran, cjs]@cs.ucc.ie

Abstract—Inter-technology handoff is a complex process that involves several procedures including resource management, mobility management, and security functions. These procedures increase the handoff delay as the involved signaling occurs between different administrative domains. Hence, seamless handoff across different access technologies is conditioned on developing efficient architectures and mechanisms that reduce this delay. In this paper, we propose a broker-based architecture for integrated heterogeneous networks. The proposed architecture employs novel resource query mechanism for IEEE 802.21. Additionally, we propose an extended handover keying to reduce the user authentication delay. Furthermore, other techniques such as proactive configuration, data buffering and controlled forwarding are used to reduce the handoff delay and ensure seamless transitions. The overall architecture and proposed mechanisms show a significant reduction in both the average handoff delay and the the probability of handoff interruption.

Index Terms—Vertical handoff, IEEE 802.21, Media Independent Handover, Performance analysis, Extended Handover Keying.

I. INTRODUCTION

Enabling inter-technology roaming, commonly known as vertical handoff (VHO), is motivated by improving the system resource utilization, enhancing the user wireless experience, and reducing the communication cost. However, the inherent characteristic heterogeneity of both access technologies and provided services complicates the handoff process. For example, both VHO discovery and decision can not depend on the received signal strength as a single parameter. Other factors such as user preferences, power consumption, and application requirements [1] are considered basic decision parameters.

Media Independent handover (MIH) standard [2], also known as IEEE 802.21, is recently proposed as a generic framework that enables information exchange of several parameters, such as network condition, service cost, and application quality-of-service (QoS) requirements, in order to facilitate VHO discovery and decision phases. Once the decision is taken, the execution of any VHO involves other procedures including security and mobility functions as the mobile terminal (MT) typically crosses different administrative domains. All these procedures increase VHO delay and increase the probability of VHO interruption.

In this paper, we develop a distributed broker architecture that employ the state of the art protocols and MIH functions to achieve seamless VHO across heterogeneous domains. In our architecture, we propose a novel resource query strategy for IEEE 802.21. Additionally, we propose a new extension for handover keying (HOKEY) [3] to reduce the required authentication delay. These proposals reduce both VHO delay and handoff interruption probability as shown in our performance analysis.

In Section II, we present a brief background on the state of the art architectures and protocols for integrated systems. The proposed architecture is then presented in Section III. In Section IV, we derive the expected VHO delay and VHO interruption probability and compare them to existing architecture corresponding metrics. The conclusion is then presented in Section V.

II. BACKGROUND AND RELATED WORK

VHO is considered a three-phase process including discovery, decision, and execution. During the discovery phase, the MT scans its neighborhood for better alternative access. The VHO decision involves determining the required resources for active applications, checking for these resources in the discovered candidate networks, and choosing one of these candidates according to a pre-defined VHO policy. Once the decision is taken, the MT should establish a new path for its data through the chosen network. Hence, VHO execution implies several functions such as user authentication and traffic diversion to the new point of attachment (PoA). These procedures involve the interaction between different nodes whose locations are architecture dependent. Hence, the architecture design has a direct impact on VHO delay.

Generally, integrated network architectures can be categorized into tightly-coupled and loosely-coupled networks [4], [5] depending on the relation between the integrated networks. In the former, different PoAs emulate the functions of a single master network such that VHOs appear to the MT as traditional handoffs in the master network. In loosely-coupled architectures, integrated networks operate independently using different set of protocols except during user authentication

and authorization. Loosely-coupled networks can further be classified into direct and indirect [6], [7], [8] architectures, depending on the type of service level agreements (SLAs) between network operators. In the former, SLAs are established between the network operators themselves, while in the latter, the SLAs are established between individual operators and an intermediate entity, e.g. a broker network. Generally, tightly-coupled architectures feature smaller VHO delay but their applicability is limited to single-ownership networks. Hence, loosely-coupled architectures are preferred when large number of operators exists in the system. For a similar reason, indirect architectures are preferred due to their scalability to the number of existing operators. However, these architecture require further optimization to reduce their handoff delay requirements.

MIH provides several functions to facilitate VHO and reduce its associated delay. First, MIH *information service* represents a means for obtaining information that assist the MT in VHO discovery and decision. The configuration parameters of neighbor PoAs are examples of information that can be exchanged to reduce the VHO delay. For example, the active frequency bands of neighbor PoAs can be exchanged to avoid excessive delays and power consumption in a full band scan. Second, *event service* enables communicating VHO relevant information to higher layers. Losing the connection to a specific PoA, the availability of new PoAs, and network condition degradation are examples for events that can be communicated to higher layers. MIH also supports a *command service* by which higher layers can gather information, manage the device, and control handover through instructing specific actions to lower layers. Obtaining current network condition, interface switching, and resource reservation in the candidate network are all examples for possible command services. Many of these MIH functions are employed to reduce VHO delay. These services are enabled through several logical architectural entities. In addition to the MIH-enabled MTs, these entities include MIH point of service (MIH_PoS) if they interact with the mobile client, MIH non-point of service (MIH non-PoS) if it interacts *only* with other MIH-enabled *network* entity, and MIH Information Server (MIH_IS).

On executing VHOs, the user authentication is associated with key distribution using well-known protocols such as Extensible authentication protocol (EAP). These processes involve an EAP client (e.g. MT) and an EAP home server through the visited domain authenticator using an Authentication, Authorization, and Accounting (AAA) protocol such as RADIUS. A typical EAP authentication process takes one [9] or more round-trips between the EAP client and server depending on the adopted authentication algorithm. Hence, user authentication could result in a significant delay to the VHO process, especially in loosely-coupled architectures in which EAP messages are transmitted over the Internet. HOKEY [3] is recently proposed as an extension to EAP to speed handoff process. HOKEY introduces two new re-authentication messages and a new key hierarchy to EAP. HOKEY reduces the handover delay by allowing the MT to
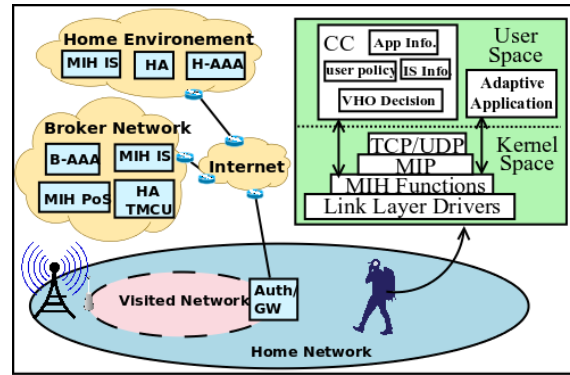

Figure 1. Proposed integrated heterogeneous system architecture.

authenticate itself to a local server when it moves between different authenticators within the same domain. It is worth noting that mobility signaling also contributes to VHO delay as their messages crosses the Internet. Hence, fast handover techniques, e.g. [10], are encouraged to reduce the total delay.

In the proposed architecture, we exploit the state of the art MIH and EAP protocols to improve VHO performance. Additionally, we propose a few modifications to the existing standards to reduce the signaling load and the MT configuration time during VHOs as presented in the following section.

III. PROPOSED ARCHITECTURE

Figure 1 shows our proposed IEEE 802.21-enabled architecture. The architecture adopts a loosely-coupled distributed broker-based coupling approach. The distributed nature enables implementing the broker function close to the MT. Additionally, it ensures the architecture scalability and facilitates limited initial deployment and a smooth path for upgrades. The following subsections introduce different components required for supporting seamless VHO in our IEEE 802.21-enabled integrated architecture followed by improved resource query mechanism and fast authentication procedure.

A. Main Architecture Components

1) MIH-enabled Mobile Client: Figure 1 shows the protocol stack for a multi-interface MIH-enabled mobile client. The Connection Coordinator (CC) represents the heart of VHO management by controlling different functions including mobility management, power management, and resource management. The CC, as an MIH user, registers and sets thresholds for specific events such as new link availability, current connection deterioration, and losing the connection to the current PoA. Additionally, the CC gathers information from the MIH IS about neighbor PoAs including operating bands and Medium Access Control addresses. Such information combined with collected application requirements and resource availability in candidate networks enable the CC to take the VHO decision according to a predefined policy.

2) The MT Home Environment: The home environment is assumed to have an MIH IS (H-IS) and an AAA server (H-AAA). The H-IS may participate in the MT bootstrapping

as detailed in Section III-D. The H-AAA participates in authenticating the user in the home network and the transfer of security credentials in the visited networks.

3) Broker Network: The broker network consists of a set of distributed sub-networks that fully cover the integrated system area. Each sub-network should have the following logical components. A broker MIH PoS (B-PoS) interacts with the mobile terminal during resource query, reservation, and releasing. The broker also employs an MIH IS (B-IS) to store relevant information about the access technologies located within its coverage. This information include network-specific information such as service cost, QoS capabilities, and PoA-specific information such as location, addressing information, supported data rates, and channel parameters. A Broker Home Agent (B-HA) is dynamically assigned to the mobile terminal during its bootstrapping. The B-HA is responsible for forwarding data to the MT at its current PoA and preparing detailed usage reports for billing. Note that the B-HA always participates in data forwarding independent of the user current location. Hence, this design choice simplifies the billing management process. In order to reduce packet losses during VHO, the broker employs a Traffic Control and Monitoring Unit (B-TCMU). The B-TCMU allocates buffering spaces for users in the middle of VHOs. The buffered data is then forwarded to the MT in the next network at a controlled rate to avoid packet loss resulting from limited buffering spaces and resource availability on moving to slower networks. Finally, a Broker-AAA (B-AAA) server is used to authenticate the MT whenever it roams to a foreign environment. The authentication process is based on a broker specific derived security credentials as presented in subsection III-C.

B. Improved Resource Query Mechanism

MIH enables the MT to query the candidate networks, obtained from MIH IS, for resource availability through the MIH PoS. According to the IEEE 802.21 standard, the MIH PoS should query every corresponding MIH PoS in the candidate networks to check for the resource availability. This procedure involves a huge signaling load and processing requirements as the number of VHOs increases. Additionally, it may involve a significant delay as the sending MIH PoS has to wait for the response coming over the Internet from all the candidate networks. In our architecture, we propose that the B-IS maintains an indicator for the interest of the covered PoAs in receiving VHO traffic. Typically, this indicator is only reset when a PoA loading exceeds a specific predetermined threshold. Hence, during the resource query phase, the MIH PoS, i.e. B-PoS, sends the query request to the local B-IS including the candidate PoAs. The B-IS replies with the queried PoA indicators. The MIH PoS then compiles the resource query response to the MT. Note that the proposed approach employs only one message to a *local* server in comparison to sending multiple queries to candidate MIH PoAs over the Internet. Hence, it will significantly reduce both signaling load and delay.
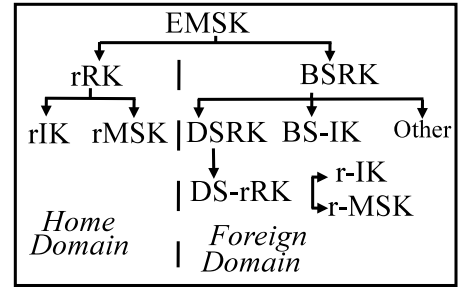

Figure 2. Re-authentication key hierarchy.

C. Extended HOKEY

The main motivation for the proposed authentication mechanism is to reduce the VHO delay whenever the MT is roaming to a foreign network. The proposed mechanism takes advantage of the conveniently located broker sub-network to reduce VHO delay by eliminating the need to contact the usually far home network whenever the MT roams to a foreign network. This is attained by introducing an additional level in the key distribution hierarchy as shown in Figure 2 . In this figure, a special broker specific root key (BSRK) is derived for the designated broker network at the beginning of the session independent of whether the session starts in the home domain or in a foreign domain.

Typically, the BSRK is derived from a shared secret between the MT and its H-AAA. In the figure, BSRK is derived from an extended master session key (EMSK) that is generated at the beginning of the user session. Note that the EMSK is not disclosed to any other entity in the system and is reserved for various scenarios including re-authentication during handover process. The BSRK is used to derive different security credentials required to ensure the integrity and confidentiality of the communication between the MT and different entities in the broker network. On visiting a new domain, the local-AAA (L-AAA) server contacts the B-AAA to authenticate the MT using an appropriate EAP method and obtaining a valid domain specific root key (DSRK). The L-AAA uses the DSRK to derive the domain specific re-authentication root key (DS-rRK), which is used to generate both re-authentication-Integrity key (rIK) and re-authentication master session key (r-MSK). rIK is used to protect any re-authentication procedure within this local domain while r-MSK is used to derive other session keys using a suitable secure association protocol, e.g. IEEE 802.11i for WLANs, to ensure the communication security. Based on this key hierarchy, when a MT roams to a foreign network, it performs full EAP authentication with the B-AAA instead of H-AAA.

D. MT Bootstrapping

In the proposed architecture, the MT bootstrapping implies device configuration at the beginning of each session. This configuration attaches the MT to a broker sub-network through setting the MT HA, MIH PoS, and MIH IS to be the respective entities of this broker sub-network. Additionally, the MT

should obtain a new home address (HoA) and derive all the keys required for securing its communication with different nodes. The first step in the MT bootstrapping is obtaining the address of the local MIH IS, whose address can be obtained using different mechanisms such as using dynamic host configuration protocol (DHCP) or using an EAP feature such as Type-Length-Value (TLV) or Attribute Value Pair (AVP) after a successful authentication on the local link. The B-HA is then determined as an optimal HA for its current PoA after requesting such information from the B-IS.

The MT then contacts the B-HA to establish security credentials and obtain a home address. Typically, the B-HA will request the assistance of B-AAA to authenticate the user and derive the required security credentials. At this point, there are two possible scenarios depending on the user current network. If the MT is located in its home environment, the B-AAA will perform full EAP authentication during which the B-AAA contacts H-AAA for user authentication and obtaining the BSRK. In contrast, if the MT is located in a visited environment, it should have been already authenticated by the broker network for the visited network access. Hence, the BSRK would be available and can be used for the authentication request issued by the B-HA. Upon successful authentication, the MT will be allocated a home address and the security association between the B-HA and the MT is established. Additionally, the MT will be assigned the MIH PoS and MIH IS of the broker network using TLV or AVP. At this point the MT bootstrapping is complete and it can proceed by obtaining an IP address on the local link and send a binding update message to its HA, i.e. B-HA.

IV. VHO Performance Analysis

In this section, we compare the VHO delay for the proposed architecture and the traditional direct loosely-coupled architecture to show the expected delay reduction. Generally, VHO delay (t_{vho}) consists of different components including

- neighborhood discovery delay (t_{nd}) spent in obtaining candidate neighbor network information from the MIH-IS,
- neighborhood scanning delay (t_{ns}) spent in scanning for candidate networks obtained in the previous step,
- application resource query delay (t_{arq}), which corresponds to CC and application negotiation to determine the required resources in candidate networks,
- candidate resource query delay (t_{crq}) during which the resource availability in candidate networks is determined and communicated to the CC,
- VHO decision delay (t_{vd}) required by CC for determining the next network,
- VHO initiation delay (t_{vi}) during which the MT establishes L2 link with the candidate network and configures an L3 address. Simultaneously, the CC communicates its decision to the MIH-PoS to notify it with the target network for any buffered data, and
- VHO completion delay (t_{vc}), which is required for informing the previous PoA with VHO completion to release any

resources allocated for the MT.

Typically, t_{arq} and t_{vd} only involve communication between different modules collocated in the MT. Hence, their contribution to the total t_{vho} can be neglected when compared to other delay components. Additionally, neighborhood discovery can be proactively completed once the MT establishes the link with the PoA. Furthermore, t_{vc} is not an effective part of VHO delay as it just corresponds to a notification for handoff completion. Hence, t_{vho} can be approximated as

$$t_{vho} \simeq t_{ns} + t_{crq} + t_{vi}. \tag{1}$$

The scanning delay depends on the time required for scanning each candidate network ($t_{ns}^{\gamma_i}$), where γ_i corresponds to a candidate network and $i \in \{1, .., n_c\}$ and n_c corresponds to the number of candidate network. Hence, t_{ns} can be expressed as

$$t_{ns} = \max_i t_{ns}^{\gamma_i} \tag{2}$$

Clearly, t_{ns} only depends on technology γ_i specifics and is independent of the adopted architecture. On the contrary, t_{crq} and t_{vi} are architecture dependent as they involve message exchange between different nodes located in various networks. Such communication delay components are estimated as the total of transmission time, propagation time, processing time, and queuing time over both wireless and wired links. In the following analysis, we respectively denote the *PoA*, network gateway, *PoS*, and Home agent with *PoA*, *GW*, *PoS*, and *HA*. Additionally, we use α, γ_s, β, and ϕ to respectively denote the current network, selected candidate network, broker network, and MT home network. We also assume that τ_{ws}^x represents the wireless delay in network x, where x can be α for currently visited network or γ_i for candidate network i. Similarly, τ_{wd} denotes the delay per a wired hop[1]. Hence, the total delay in the wired portion depends on the hop distance between the communicating nodes. Let $a(b)$ denotes node a in network b and $D_{x(n)}^{y(m)}$ represents the hop distance between nodes $x(n)$ and $y(m)$.

Generally, t_{crq} corresponds to the delay of transmitting Candidate_Resource_Query_Request from the MT to the MIH_PoS, the time required by the MIH_PoS to obtain resource availability information from the candidate networks, and the delay required for transmitting Candidate_Resource_Query_Response from the MIH_PoS back to the MT. In the proposed architecture, there is no need to interrogate candidate networks as the resource availability information is obtained by contacting the B-IS, which is assumed to share the same machine with the B-PoS. Hence, t_{crq} for the proposed broker-based solution, denoted as t_{crq}^{β}, can be expressed as

$$t_{crq}^{\beta} = 2t_{ws}^{\alpha} + 2\left(D_{PoA(\alpha)}^{GW(\alpha)} + D_{GW(\alpha)}^{GW(\beta)} + D_{GW(\beta)}^{PoS(\beta)}\right)\tau_{wd} \tag{3}$$

$$= 2t_{ws}^{\alpha} + 2D_{PoA(\alpha)}^{PoS(\beta)}\tau_{wd}. \tag{4}$$

---

[1]The hop delay would be assumed homogeneous for different wired links to simplify the delay analysis.

For traditional direct loosely-coupled architectures, the candidate resource query would be forwarded form the $PoS(\alpha)$ to each $PoS(\gamma_i)$. $PoS(\alpha)$ has to wait for the response for all forwarded messages from every $PoS(\gamma_i)$. Hence, t_{crq} for such loosely-coupled direct architecture, denoted as t_{crq}^d, can be expressed as

$$t_{crq}^d = 2t_{ws}^\alpha + 2\left(\max_i\{D_{PoA(\alpha)}^{GW(\alpha)} + D_{GW(\alpha)}^{GW(\gamma_i)} + D_{GW(\gamma_i)}^{PoS(\gamma_i)}\}\right)\tau_{wd}.$$
(5)

Note that in (5), it is assumed that PoA and PoS are collocated in the same node.

The VHO initiation delay is the sum of two components including L2 delay (t_{L2}) and L3 delay (t_{L3}). Generally, t_{L2} involves a technology-specific link establishment delay ($t_{ts}^{\gamma_s}$) [11] and MT authentication delay (t_{auth}). For example, a MT moving to a WLAN will have to first complete a MT association procedure before going through the authentication phase, while a MT moving to WiMax should perform multiple tasks including synchronization, ranging, and basic capability negotiation before starting the authentication and key-exchange phase [12]. Note that $t_{ts}^{\gamma_s}$ is independent of the system architecture as the signaling only occurs between the MT and the PoA. In the proposed architecture, the authentication can be completed between the MT and the B-AAA server in the broker network due to the introduced level of key hierarchy. Hence, t_{L2} for the proposed broker-based architecture, denoted as t_{L2}^β, is expressed as

$$\begin{aligned}
t_{L2}^\beta &= t_{ts}^{\gamma_s} + 2n_{auth}[\tau_{ws}^{\gamma_s} + (D_{PoA(\gamma_s)}^{GW(\gamma_s)} + D_{GW(\gamma_s)}^{GW(\beta)} + D_{GW(\beta)}^{AAA(\beta)})\tau_{wd}]\\
&= t_{ts}^{\gamma_s} + 2n_{auth}[\tau_{ws}^{\gamma_s} + D_{PoA(\gamma_s)}^{AAA(\beta)}\tau_{wd}],
\end{aligned}$$
(6)

where n_{auth} corresponds to the number of round-trip journeys required for completing authentication and depends on the adopted algorithm for the user authentication and key-exchange. In traditional loosely-coupled direct architecture, the authentication takes place between the MT and its H-AAA server through the visited network authentication server. Hence, t_{L2} for traditional architectures, denoted as t_{L2}^d, can be expressed as

$$\begin{aligned}
t_{L2}^d &= t_{ts}^{\gamma_s} + 2n_{auth}[\tau_{ws}^{\gamma_s} + (D_{PoA(\gamma_s)}^{GW(\gamma_s)} + D_{GW(\gamma_s)}^{GW(\phi)} + D_{GW(\phi)}^{AAA(\phi)})\tau_{wd}]\\
&= t_{ts}^{\gamma_s} + 2n_{auth}[\tau_{ws}^{\gamma_s} + D_{PoA(\gamma_s)}^{AAA(\phi)}\tau_{wd}]
\end{aligned}$$
(7)

Layer 3 VHO delay will vary depending on the adopted mobility management protocol. Assuming that mobile IP is adopted for inter-domain mobility in heterogeneous wireless networks, t_{L3} would involve the delay required to obtain a local IP address in the selected network ($t_{IP}^{\gamma_s}$) and the delay required to bind this address with the MT HA. Typically, the former delay is architecture independent as all the signaling is local to the target network. On the contrary, the latter delay will depend on the location of the MT HA. Hence, t_{L3} for the proposed broker-based architecture, denoted as t_{L3}^β, can be

Table I
PERFORMANCE ANALYSIS PARAMETERS

Parm.		Parm.		Parm.	
$E[t_{ws}^{tot}]$	$150\mu s$	$E[D_\alpha^\beta]$	3	$Var(D_\alpha^\beta)$	1
$CoV(t_{ws}^{tot})$	0.75	$E[D_{\gamma_s}^\phi]$	5	$Var(D_{\gamma_s}^\phi)$	1
n_{auth}	2	$E[D_{max}]$	7	$Var(D_{max})$	1

expressed as

$$\begin{aligned}
t_{L3}^\beta &= t_{IP}^{\gamma_s} + \tau_{ws}^{\gamma_s} + (D_{PoA(\gamma_s)}^{GW(\gamma_s)} + D_{GW(\gamma_s)}^{GW(\beta)} + D_{GW(\beta)}^{HA(\beta)})\tau_{wd}\\
&= t_{IP}^{\gamma_s} + \tau_{ws}^{\gamma_s} + D_{PoA(\gamma_s)}^{HA(\beta)}\tau_{wd}.
\end{aligned}$$
(8)

Similarly, t_{L3} for traditional direct architectures, denoted as t_{L3}^d, can be expressed as

$$\begin{aligned}
t_{L3}^d &= t_{IP}^{\gamma_s} + \tau_{ws}^{\gamma_s} + (D_{PoA(\gamma_s)}^{GW(\gamma_s)} + D_{GW(\gamma_s)}^{GW(\phi)} + D_{GW(\phi)}^{HA(\phi)})\tau_{wd}\\
&= t_{IP}^{\gamma_s} + \tau_{ws}^{\gamma_s} + D_{PoA(\gamma_s)}^{HA(\phi)}\tau_{wd}.
\end{aligned}$$
(9)

After completing the address binding, the traffic is diverted toward the new PoA and the handoff process is completed. Based on the above analysis and substituting from (2), (4), (6), and (8) in (1), t_{VHO} for the broker-based architecture, t_{VHO}^β, can be expressed as

$$\begin{aligned}
t_{VHO}^\beta &= \max_i t_{ns}^{\gamma_i} + 2t_{ws}^\alpha + (3 + 2n_{auth})D_\alpha^\beta\tau_{wd} +\\
&\quad t_{ts}^{\gamma_s} + (2n_{auth} + 1)\tau_{ws}^{\gamma_s} + t_{IP}^{\gamma_s}\\
&= t_{ws}^{tot} + (3 + 2n_{auth})D_\alpha^\beta\tau_{wd},
\end{aligned}$$
(10)(11)

where $t_{ws}^{tot} = \max_i t_{ns}^{\gamma_i} + 2t_{ws}^\alpha + t_{ts}^{\gamma_s} + (2n_{auth} + 1)\tau_{ws}^{\gamma_s} + t_{IP}^{\gamma_s}$ represents the total wireless delay, and D_α^β represents $D_{PoA(\alpha)}^{PoS(\beta)}$, $D_{PoA(\gamma_s)}^{AAA(\beta)}$, and $D_{PoA(\gamma_s)}^{HA(\beta)}$, which are independent identically distributed (i.i.d) random variables. Similarly, substituting from (2), (5), (7), and (9) in (1), t_{VHO} for the traditional direct architectures, t_{VHO}^d, can be expressed as

$$t_{VHO}^d = t_{ws}^{tot} + 2D_{max}\tau_{wd} + (2n_{auth} + 1)[\tau_{ws}^{\gamma_s} + D_{\gamma_s}^\phi\tau_{wd}](12)$$

where we assume that $D_{PoA(\gamma_s)}^{AAA(\phi)}$ and $D_{PoA(\gamma_s)}^{HA(\phi)}$ are i.i.d and are further denoted as $D_{\gamma_s}^\phi$ and D_{max} represents $\max_i\{D_{PoA(\alpha)}^{PoS(\gamma_i)}\}$. By subtracting (10) from (12), one can estimate the expected VHO delay reduction as

$$\begin{aligned}
E[\Delta_{VHO}] &= E[t_{VHO}^d - t_{VHO}^B]\\
&= 2(E[D_{max}] - E[D_\alpha^\beta])E[\tau_{wd}] +\\
&\quad + (2n_{auth} + 1)(E[D_{\gamma_s}^\phi] - E[D_\alpha^\beta])E[\tau_{wd}].
\end{aligned}$$
(13)

Clearly, the VHO delay reduction is independent of any wireless delays. Typically, τ_{wd} is two to three order of magnitude larger than wireless delay components. More importantly, the accumulation of wired delay over multiple hops contributes to the most significant portion of VHO delay. Using the analysis parameters shown in Table I , Fig. 3 shows that the proposed approach results in a significant delay reduction in comparison to traditional direct approach. Several components contribute to this delay drop. First, the proposed IEEE 802.21 resource
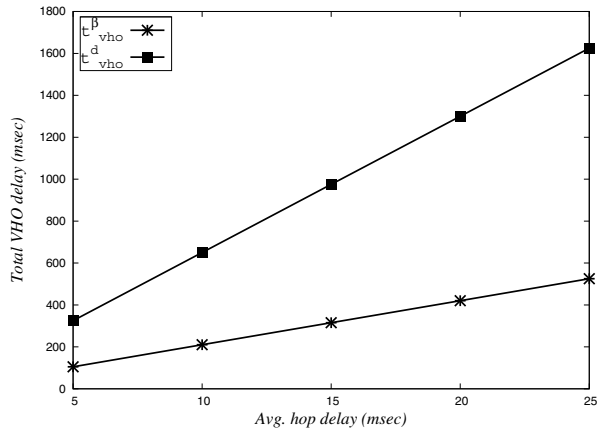
Figure 3. VHO delay versus average wired hop delay

Figure 4. Probability of VHO interruption versus coefficient of variation of wired hop delay τ_{wd}

query mechanism results in a reduction of $2(E[D_{max}] - E[D_\alpha^\beta])E[\tau_{wd}]$. Second, the proposed key hierarchy leads to another delay saving of $2n_{auth}(E[D_{\gamma_s}^\phi] - E[D_\alpha^\beta])E[\tau_{wd}]$. Third, locating the HA at the broker network further reduces the total VHO delay by $(E[D_{\gamma_s}^\phi] - E[D_\alpha^\beta])E[\tau_{wd}]$. Note that for many networks, especially cellular networks, network nodes are not directly accessible. Hence, any messages forwarded to a cellular *PoS* have to go through the network core leading to a significant increase in the VHO delay. Such delay is avoided through the directly accessible broker sub-network in the presented architecture.

Since the architecture design affects the VHO delay, it also impacts the design and performance of proactive VHO algorithms. Typically, proactive VHO algorithms start the handoff procedure before leaving the current network coverage with a sufficient duration D_{VHO}. To ensure seamless VHO, $D_{vho} \geqslant t_{vho}$. Hence, we define the probability of interrupted VHO as

$$P(interrupted VHO) = P(t_{vho} \geqslant D_{vho}). \qquad (14)$$

In order to estimate this probability, we derive the distribution of t_{vho}^β and t_{vho}^d. For the proposed architecture, we first estimate the expected value and the variance of the total wired delay component (τ_{wd}^{tot}); i.e. $\tau_{wd}^{tot} = (3 + 2n_{auth})D_\alpha^\beta \tau_{wd} = \overline{D}\tau_{wd}$, where $\overline{D} = (3 + 2n_{auth})D_\alpha^\beta$. Typically, τ_{wd}^{tot} is a random sum of i.i.d. random variables τ_{wd}. Hence, $E[\tau_{wd}^{tot}] = E[\tau_{wd}]E[\overline{D}]$ and $Var(\tau_{wd}^{tot}) = Var(\tau_{wd})E[\overline{D}] + E[\tau_{wd}]^2 Var(\overline{D})$ [13], where $Var(X)$ represents the variance of random variable X. Note that $\overline{D}$ is a linear function of D_α^β. Hence, $E[\overline{D}] = (3 + 2n_{auth})E[D_\alpha^\beta]$ and $Var(\overline{D}) = (3 + 2n_{auth})^2 Var(D_\alpha^\beta)$. Using the mean and variance of τ_{wd}^{tot} and τ_{ws}^{tot}, each random variable is fitted to a Phase-type (PH) distribution [14] as described in [15]. The distribution of t_{vho}^β is then derived as a sum of two PH distributions. The resultant distribution of the sum can be obtained using [14, Theorem 2.6.1]. Consequently, one can then estimate (14). Using a similar procedure, the VHO interruption probability for direct loosely-coupled architecture is estimated.
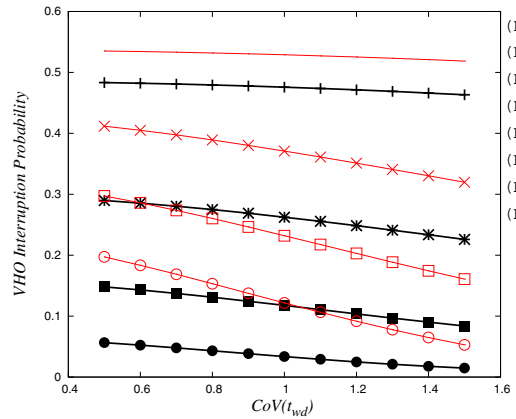
Let the VHO algorithm chooses $D_{vho} = E[\tau_{vho}^{tot}] + a\, std(\tau_{vho}^{tot})$, where a represents a tunable scaling factor and $std(\tau_{vho}^{tot})$ represents the standard deviation of τ_{vho}^{tot}. Fig. 4 plots the probability of VHO interruption versus coefficient of variation of τ_{wd} for (architecture, a) in which B and D respectively refer to our broker-based and direct coupling architectures. Note that the unity coefficient of variation (CoV) is a typical value for M/M/1 queues. Clearly, the figure shows a noticeable reduction in the VHO interruption probability for the proposed architecture in comparison to the direct loosely-coupled approach. It is also worth mentioning that larger the scaling factor implies an increase in the MT battery usage because the MT will operate more than one interface for longer durations. Hence, the resultant smaller VHO delay of our proposed architecture implies both higher VHO reliability and reduced power consumption.

V. CONCLUSION

VHO involves several procedures that contributes to its total delay. Reducing such delay is a crucial requirement to improve VHO performance and ensure seamless transition. In this paper, we propose an indirect scalable architecture that employs state of the art protocols. This architecture employs IEEE 802.21 to enable information exchange that facilitates proactive VHO, MT power saving, and resource management functions. Additionally, a novel resource query mechanism for IEEE 802.21 is introduced to reduce signaling load and delay during VHOs. Furthermore, a novel authentication mechanism is proposed to reduce the authentication delay in visited domains. The proposed architecture and mechanisms lead to significant reduction in VHO delay, VHO interruption probability, and power consumption.

ACKNOWLEDGMENT

REFERENCES

[1] M. Kassar, B. Kervella, and G. Pujolle, "An overview of vertical handover decision strategies in heterogeneous wireless networks," *Computer Communications*, vol. 31, no. 10, pp. 2607 – 2620, 2008.

[2] I. . W. Group, *IEEE Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services*, Draft Standard, IEEE Std., Jan 2009.

[3] V. Narayanan and L. Dondeti, *EAP Extensions for EAP Re-authentication Protocol (ERP)*, IETF RFC 5296 Std., Aug. 2008.

[4] A. Salkintzis, C. Fors, and R. Pazhyannur, "WLAN-GPRS integration for next-generation mobile data networks," *IEEE Wireless Commun.*, vol. 9, no. 5, pp. 112–124, Oct. 2002.

[5] M. Buddhikot, G. Chandranmenon, S. Han, Y.-W. Lee, S. Miller, and L. Salgarelli, "Design and implementation of a WLAN/cdma2000 interworking architecture," *IEEE Commun. Mag.*, vol. 41, no. 11, pp. 90–100, Nov. 2003.

[6] I. Akyildiz, S. Mohanty, and J. Xie, "A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems," *IEEE Commun. Mag.*, vol. 43, no. 6, pp. S29–S36, June 2005.

[7] C. Makaya and S. Pierre, "Reliable Integrated Architecture for Heterogeneous Mobile and Wireless Networks," *J. of Networks*, vol. 2, no. 6, pp. 24–32, 2007.

[8] P. Chiron, E. Njedjou, P. Seite, K. Gosse, E. Melin, and P. Roux, "Architectures for IP-based network-assisted mobility management across heterogeneous networks," *IEEE Wireless Commun.*, vol. 15, no. 2, pp. 18–25, April 2008.

[9] L. Salgarelli, M. Buddhikot, J. Garay, S. Patel, and S. Miller, "Efficient authentication and key distribution in wireless IP networks," *IEEE Wireless Commun.*, vol. 10, no. 6, pp. 52–61, Dec. 2003.

[10] G. D. et al, *Fast Handovers for Mobile IPv6*, Internet Draft, IETF Std., March 2002.

[11] S.-J. Yoo, D. Cypher, and N. Golmie, "Timely effective handover mechanism in heterogeneous wireless networks," *Wireless Personal Communications*, vol. 52, pp. 449–475, 2010.

[12] S.Ahmadi, "An overview of next-generation mobile wimax technology," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 84 –98, june 2009.

[13] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, 2nd ed., 2002.

[14] G. Latouche and V. Ramaswami, *Introduction to Matrix analytic Methods in Stochastic Modeling*, ser. ASA-SIAM series on Statistics and Applied Probability. SIAM, 1999.

[15] A. H. Zahran and C. J. Sreenan, "Threshold-based media streaming optimization for heterogeneous wireless networks," *IEEE Trans. on Mobile Comput.*, vol. 9, pp. 753–764, 2010.