Handoff Failure Analysis of Adaptive Keep-Alive Interval (AKI) in 3GPP Generic Access Network (GAN)

Kai-Hsiu Chen, *Student Member, IEEE,* and Jyh-Cheng Chen, *Senior Member, IEEE*

Abstract—The *Generic Access Network (GAN)* developed by 3GPP extends services to unlicensed spectrum. A *mobile station (MS)* with multiple air interfaces can access 3GPP services through GAN and roam seamlessly between different radio access networks. However, mobility management and resource management are critical issues in GAN. To allocate GAN resources efficiently, we propose *Adaptive Keep-alive Interval (AKI)*. By using AKI, an MS can send less signaling messages. In addition, AKI can increase system utilization by releasing unused resources. In this paper, we develop a mathematical model to evaluate the handoff failure probability of the proposed AKI. The analysis is validated by extensive simulations. The results show that the proposed AKI can reduce handoff failure probability and increase GAN utilization significantly.

Index Terms—3GPP Generic Access Network (GAN), internetworking, performance analysis, mobility and resource management, communications systems.

I. INTRODUCTION

THE *Generic Access Network (GAN)* [1], [2] developed in the 3rd *Generation Partnership Project (3GPP)* evolves from the *Unlicensed Mobile Access (UMA)* [3]–[5]. It extends *Global System for Mobile Communications (GSM)/General Packet Radio Service (GPRS)/Universal Mobile Telecommunications System (UMTS)* [6] services to unlicensed spectrum. Basically, a gateway approach is adopted in the GAN architecture to integrate 3GPP core network with other radio technologies. From the viewpoint of a 3GPP core network, GAN is just another *Radio Access Network (RAN)* like *GSM/EDGE RAN (GERAN)*. A *Mobile Station (MS)* with multiple air interfaces can access 3G services through GAN and roam seamlessly between GAN, GERAN, and *UMTS Terrestrial RAN (UTRAN)*. Besides, many different wireless access technologies, such as IEEE 802.11 Wireless Local Area Networks (WLAN) [7] and Bluetooth [8], can be applied to GAN.

Among the different RANs, resource management is an important issue. In order to minimize the handoff failure probability, *Call Admission Control (CAC)* usually reserves some resources for handoff calls [9]–[12]. However, in GAN, an MS may stay in GERAN/UTRAN for a certain period of time without switching to GAN even though the GAN has reserved resources for it. Hence, the reserved resources are wasted and the GAN is under utilized.

As defined in the standards [1], [2], after an MS registers with GAN, resources are reserved to accommodate the MS and a timer called *TU 3906* is triggered. The MS needs to periodically send a *keep-alive message* refreshing the timer in order to avoid its expiration and the resulting release of resources at the GAN. In this paper, we propose *Adaptive Keep-alive Interval (AKI)* algorithm to increase the system utilization. In our proposed technique, before an MS actually switches to GAN, the keep-alive interval is increased to prolong the interval of sending the keep-alive message. Thus, the MS sends fewer and fewer keep-alive messages as time goes by. Furthermore, if an MS does not switch to the *GAN mode* for a long period of time, the keep-alive interval will be long. The reserved GAN resources for the MS are likely to be preempted by another handoff MS. Hence, the GAN resource utilization is improved.

In our previous paper [13], we evaluated the performance of GAN services improved by the proposed AKI by extensive simulations. Among the evaluated metrics, we concluded that the GAN *handoff failure probability* is the most important one for mobile users to access GAN services. When GAN handoff failure probability is reduced, GAN utilization is increased. To further understand the impact of GAN handoff failure probability, a mathematical model is proposed to analyze the handoff failure probability in this paper. In [10], the call blocking probability is analyzed in homogeneous networks by decomposing dependency between networks nodes. Based on [10], bearer reservation and preemption behaviors in Last-Come-First-Preempted are analyzed in [14]. The models presented in [10], [14] cannot apply to heterogeneous networks because the mean network residence time is always the same in [10], [14]. In heterogeneous networks, however, the mean network residence time is different for different type of network. Although [15] analyzes the handoff failure probability with *WLAN-First*[1] strategy for MSs, it cannot model the

---

[1]The *WLAN-First* here means that an MS will switch from a cellular network to a WLAN as long as there is one WLAN available.
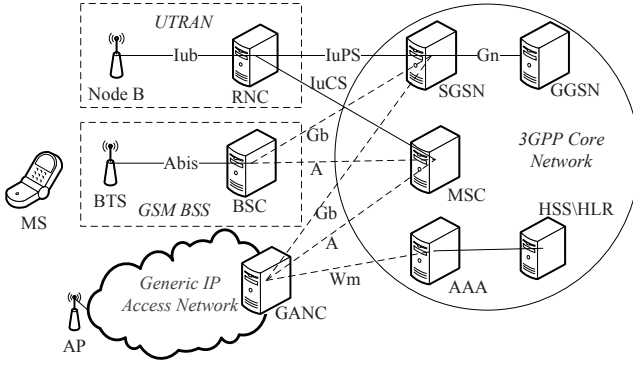
Fig. 1. GAN architecture.

Fig. 2. *"Up"* interface.

pre-registration behavior defined in the GAN standards [1], [2]. In this paper, we propose a mathematical model for the pre-registration and preemption behaviors in the interworking between GERAN/UTRAN and GAN. The proposed analytical model is validated by simulation. The analytical and simulation results show that the handoff failure probability is reduced significantly in the proposed AKI. In addition, the GAN is better utilized.

The contribution of this paper is two-fold. First, we propose a novel adaptive keep-alive strategy to refresh the TU 3906 timer reserving resources at GANs before an MS conducts a handover. It is a critical issue for GAN resource management. We identify the problem which was not solved before and needs to be resolved to improve the GAN performance. Second, we develop an intricate analytic model to evaluate the handoff failure probability. The mathematical model quantifies the performance of the proposed AKI. Also, the analysis shows the performance of a larger scale network, which exams the scalability of the proposed AKI. Extensive simulations have also been conducted to validate the analysis.

The rest of this paper is organized as follows. Section II provides an overview of the GAN system. Section III states the problems to be resolved. Section IV discusses the proposed solution. The proposed analytical model is presented in Section V. Section VI discusses the numerical results. Section VII summarizes this paper.

II. BACKGROUND

The GAN architecture illustrated in Fig. 1 shows a gateway approach to integrate 3GPP core network and GAN. The gateway is called *GAN Controller (GANC)* which also performs the functions of the GSM *Base Station Subsystem (BSS)*. As shown in the figure, both of the GANC and the GSM BSS communicate with the *Mobile Switching Center (MSC)* and *Serving GPRS Support Node (SGSN)* by *"A"* interface and *"Gb"* interface, respectively. Besides, the GANC communicates with the MS by a newly defined interface which is called *"Up"* interface as shown in Fig. 2 [1], [2].

When an MS is powered on, it initially starts in *GERAN/UTRAN mode* by executing the normal GERAN/UTRAN power-on procedure. Using WLAN as an example in GAN, an MS communicates with the GANC through an IEEE 802.11 *Access Point (AP)*. If the MS discovers some appropriate APs, it can switch to *GAN mode*.
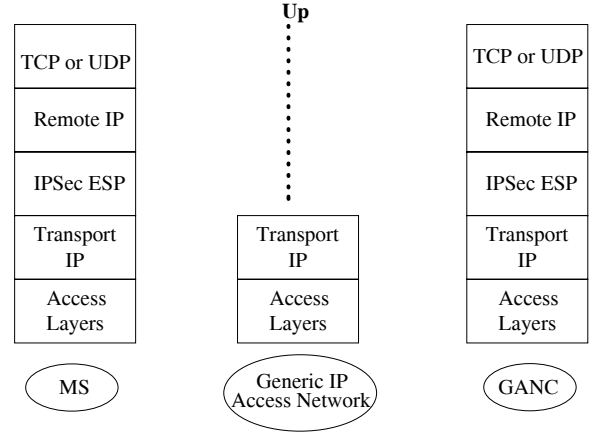
The AP does not have to support any extra GAN functionality. It is transparent between the MS and the GANC. The MS can switch to GAN mode by using the following steps [1], [2]:

1) The MS connects to an appropriate AP.
2) The MS configures its *Transport IP* address or *Local IP* address. The IP address can be obtained by using DHCP [16] or by other ways.
3) With the transport/local IP address, the MS performs the *Discovery Procedure* defined in [1], [2] to find a suitable GANC.
4) The MS establishes an IPSec [17] tunnel with the *Security Gateway (SEGW)* which usually is co-located with the GANC.
5) The MS registers with the GANC through the IPSec tunnel.
6) If the GANC accepts the registration request, it becomes the *serving GANC* of the MS. The MS then can switch to GAN mode.

The *Security Association (SA)* of the IPSec tunnel is constructed by using EAP/SIM [18] or EAP/AKA [19] over Internet Key Exchange (IKEv2) [20]. During the IKEv2 exchanges, the MS and the network are mutually authenticated through the *"Wm"* interface. The IPSec tunnel recognizes the MS by the *Remote IP address*. Therefore, the IPSec tunnel does not need to be re-established when the MS moves to another subnet. Only the Transport IP address will be changed. Hence, the IPSec tunnel and the upper-layer sessions will not break due to the movement of the MS. Besides, the Remote IP address can be configured statically or dynamically by the core network during the registration process. Signaling messages such as registration, deregistration, and keep-alive messages are transmitted over TCP in *"Up"* interface. The voice calls from circuit-switched networks are carried by *Real-time Transport Protocol (RTP)* [21] and transferred as *Voice over IP (VoIP)* sessions in GAN.

If there are several available RANs for an MS, according to user preference, the MS selects one of them to connect to. Four user preferences are defined in [1], [2]:

1) GERAN/UTRAN only: The MS always stays in GERAN/UTRAN cells.
2) GERAN/UTRAN preferred: The MS will switch to GERAN/UTRAN mode as long as there is a suitable

GERAN/UTRAN cell available.

3) GAN preferred: The MS will switch to GAN mode as long as a suitable AP and an appropriate GANC are available.

4) GAN only: The MS always stays in GAN cells.

III. PROBLEM STATEMENT

Based on the standards [1], [2], from the perspective of GAN, the movement of an MS with *active* session(s) from GERAN/UTRAN to GAN is called *handover-in*. The movement from GAN to GERAN/UTRAN is called *handover-out*. For an *idle* MS, the movements are called *rove-in* and *rove-out*, respectively. Note that before an MS can switch to GAN mode, it must successfully register with the GANC first as described in Section II.

In the handover-in case, it is a nontrivial problem to determine when an MS should initiate the registration with the GANC while it is still in GERAN/UTRAN mode. This is especially challenging for GERAN/UTRAN-preferred MSs. Unless GERAN/UTRAN is not available anymore, a GERAN/UTRAN-preferred MS prefers to stay in GERAN/UTRAN instead of switching to GAN mode although it has already registered with GANC. As defined in the standards, after the registration with a GANC is accepted, the MS needs to periodically send keep-alive messages until it is deregistered. If an MS registers with the GANC too early, it will incur signaling overhead and waste energy to send many keep-alive messages. In addition, the resources reserved for the MS in the GAN cannot be used by other MSs. On the other hand, if an MS performs the registration just in the beginning of handover-in, the discovery and registration procedures will incur long handoff delay. Recall that IPSec tunnel needs to be established before performing the registration procedures. Studies have shown that the establishment of IPSec tunnel contributes significantly to handoff delay [22].

Please note only an MS moving from GERAN/UTRAN to GAN will reserve resources in the GAN cell before the MS moves in. For an MS moving from GAN to GERAN/UTRAN, there is no pre-registration. Also, the reservation applies to GERAN/UTRAN-preferred MSs only. For GAN-preferred MSs, they will move into GAN immediately once the pre-registration is accepted. For GERAN/UTRAN-only and GAN-only MSs, they always stay in the same type of cells.

IV. PROPOSED ADAPTIVE KEEP-ALIVE INTERVAL (AKI)

Although the proposed AKI has been presented in [13], to be self-contained, we briefly discuss it again here. The proposed AKI possesses the following key characteristics. The proposed AKI can reduce the amount of keep-alive messages. In addition, it provides a way to decide which reserved GAN resources should be released. In light of the aforementioned discussion, GAN resources may be reserved but not be used immediately by the early registered MSs. Therefore, when the load of GAN is heavy, newly handover-in sessions may be dropped. In the proposed AKI, we re-allocate some reserved GAN resources. Therefore, the probability of unsuccessful handoff to GAN is reduced. The GAN resources, thus, can
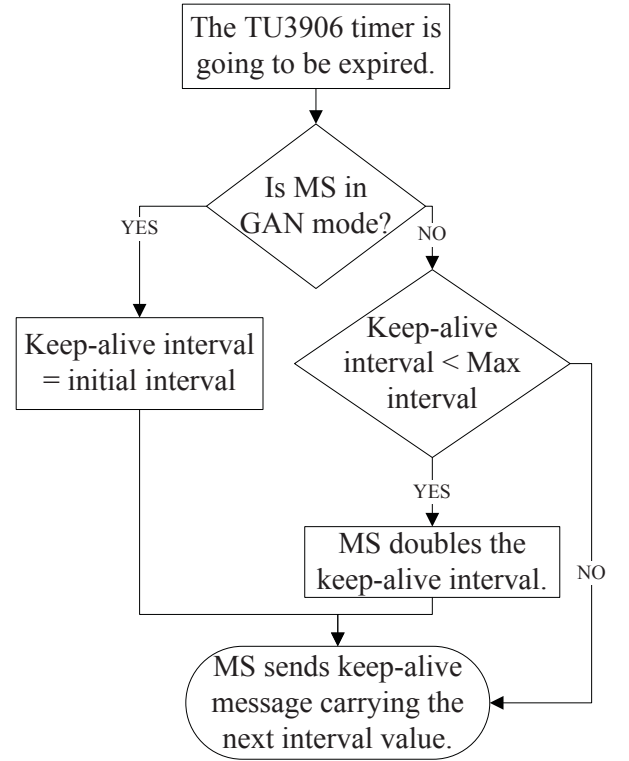

Fig. 3.   MS procedure in AKI.

be used more efficiently. The system utilization is improved as well.

The following sections present how the proposed AKI works in MS and GANC, respectively.

A. MS procedure

Once the GANC accepts the registration, as defined in the standards, it sends the *GA-RC REGISTER ACCEPT* message [2] which carries the initial keep-alive interval to the MS. After the MS receives the message, the keep-alive mechanism is started and controlled by the *TU* 3906 timer [1], [2]. The timer keeps counting down. If there are packets transmitted from the MS through the GANC, the timer is reset to the initial value. Otherwise, the MS needs to send a keep-alive message to the GANC before the timer expires, i.e. counting down to 0.

The AKI, shown in Fig. 3, is proposed to minimize the amount of keep-alive messages for early registered MSs. In AKI, the keep-alive interval doubles exponentially unless it reaches the predefined maximum value or until the MS performs handover-in. The frequency of sending keep-alive messages, thus, is reduced. After the MS performs handover-in, the interval is reset to the initial value as that originally assigned by the serving GANC.

B. GANC procedure

In light of the aforementioned discussion, the GANC will allocate GAN resources for the registered MSs even though the MSs have not really moved into the GAN. When receiving a new registration request from an MS with active session(s), the GANC will reserve the resources in the AP for the session(s). In the proposed AKI, if there is no GAN resource available,

the GANC will decide whether an early registered MS which has not handed over in yet should be preempted. If so, the released resources then are used for the newly arrived MS. However, the MS with the initial value of keep-alive interval will not be chosen to be released. That is, a newly registered MS is guaranteed to hold the GAN resource at least for the period which equals the initial keep-alive interval. Hence, the candidates to be preempted are the MSs with larger keep-alive intervals. That is, they have held the resources for a long time without using the resources. If no candidate is found to release the resources, the new registration fails. Among the candidates to be preempted, we propose three different ways for a GANC to choose:

1) Maximum interval first: The GANC releases the resources of the candidate MS with the largest keep-alive interval.
2) Minimum interval first: The GANC releases the resources of the candidate MS with the smallest keep-alive interval.
3) Random: The GANC randomly releases the resources of a candidate MS.

Intuitively, the candidate MS with the largest value of keep-alive interval may not handover in soon so we may want to release its reserved resources. However, it is hard to predict the movement of an MS accurately. It is also possible that the candidate MS with the largest value of keep-alive interval may be in the boundary already and will handover in very soon. From this perspective, it may be better to choose the candidate MS with the smallest value of keep-alive interval. The first two approaches actually conflict with each other. Therefore, we also propose to pick up a candidate MS randomly. Simulation results shown in Section VI indicate that all of them perform similarly.

To evaluate the complexity of the proposed AKI, we assume the number of reserved sessions in a GAN cell is N. The complexity for memory space of all the three policies is $O(N)$ due to the recording of the keep-alive intervals of the reserved MSs. For time complexity, both of the maximum and minimum interval first policies have the time complexity of $O(N \cdot \log N)$ because of sorting the keep-alive intervals of the reserved MSs. For the random policy, the time complexity is only $O(N)$. Please note N is less than the maximum number of allowable sessions in a GAN cell which is less than the GAN cell capacity.

V. PERFORMANCE ANALYSIS

As mentioned earlier, in our previous paper [13], we have evaluated the proposed AKI with various performance metrics by simulation. Through simulation experiments, we found that the GAN *handoff failure probability* is the most important factor affecting mobile users to access GAN services. To get more insights of the problem, we conduct mathematical analysis for handoff failure probability in this paper. We first describe the system model in Section V-A, followed by the mathematical analysis in Section V-B. The parameters used in the analysis are listed in Table I. In the analysis, *handoff* is defined as the switching of radio connection from a *Base Station (BS)/Access Point (AP)* to another BS/AP.

TABLE I
LIST OF PARAMETERS

	Parameter	Setting
Mean session arrival rate	λ_o	0.05
Mean session holding time	$1/\mu$	600 sec
Density function of session holding time	$f_c(t)$	$\mu e^{-\mu t}$
Mean GAN resource holding time	$1/\eta_g$	300 sec
Density function of GAN resource holding time	$f_{n_g}(t)$	$\eta_g e^{-\eta_g t}$
Mean pre-registration interval during GERAN/UTRAN period	$1/\eta_u$	600 sec
Density function of pre-registration interval during GERAN/UTRAN period	$f_{n_u}(t)$	$\eta_u e^{-\eta_u t}$
GAN cell capacity	C	12
(Pre-registration period)/ (GAN resource holding time)	r_p	0.2
GAN-preferred session ratio	r_g	0.5
GERAN/UTRAN-preferred session ratio	$1 - r_g$	0.5
Neighboring cell number	N_{nb}	4
Neighboring GAN cell number	N_g	2
Neighboring GERAN/UTRAN cell number	$N_{nb} - N_g$	2
Probability that a handoff session moves to/from a GAN cell	$P_g = N_g/N_{nb}$	0.5
Probability that a handoff session moves to/from a GERAN/UTRAN cell	$1 - P_g$	0.5
Initial Keep-alive interval	I_k	20 sec
Session holding time	t_c	
GAN resource holding time	t_g	
GAN pre-registration interval during GERAN/UTRAN period	t_u	
Pre-registration period	t_p	
Residence time in the i-th cell	t_{N_i}	
Density function of t_{N_i}	$f_{n_i}(t)$	
Visited cell numbers during the session holding time	X	
Handoff number from GERAN/UTRAN cells by a session	U	
Handoff number from GAN cells by a session	G	
Handoff session arrival rate	λ_h	
Handoff rate from neighboring GERAN/UTRAN cells	λ_{hu}	
Handoff rate from neighboring GAN cells	λ_{hg}	
GAN resource requesting rate	λ_r	
Pre-registration miss rate	λ_m	
Probability of handoff failure into GAN	P_f	
Probability that a pre-registered session is in high priority	P_h	

Thus, handoff includes the switching from a GAN AP to another GAN AP, from a GAN AP to a GERAN/UTRAN BS, from a GERAN/UTRAN BS to a GAN AP, and from a GERAN/UTRAN BS to another GERAN/UTRAN BS. Recall that as defined in the standards [1], [2], *handover* refers to the inter-system handoff of an MS with active session(s). Thus, handover includes the switching from a GAN AP to a GERAN/UTRAN BS, and from a GERAN/UTRAN BS to a GAN AP.

A. System Model

Fig. 4 illustrates the network topology used in the analysis. To prevent service disruption, generally, GERAN/UTRAN
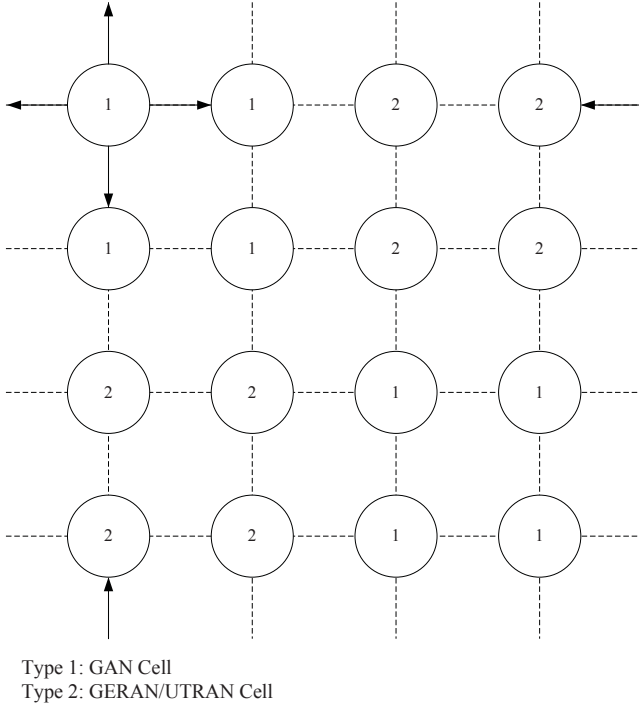
Fig. 4. Network topology for analysis.

Type 1: GAN Cell
Type 2: GERAN/UTRAN Cell

Fig. 5. Events that may occur during the lifetime of a GERAN/UTRAN-preferred session.

cells and GAN cells are overlapped. Therefore, an MS can seamlessly roam between different cells. Please note that Fig. 4 does not show the overlaps of the cells. There are two types of cells in Fig. 4: GERAN/UTRAN cell and GAN cell. For each cell, there are N_{nb} neighboring cells consisting of N_g GAN cells and $(N_{nb} - N_g)$ GERAN/UTRAN cells. In Fig. 4, there are 2 neighboring GAN cells and 2 neighboring GERAN/UTRAN cells around each cell. There are 4 moving directions for each MS. The topology is wrapped-up. That is, an MS moving up outside the boundary will reach to the bottom cell, and same as that in left and right sides. With different user preferences, the handoff decision will be different.

The resource consumption in a GAN cell is modeled by following $M/M/C/C$ *Markov process* [23]. In the Markov process, the arrival rate corresponds to the GAN resource requesting rate. The departure rate is the inverse of the GAN resource holding time. Both of the arrival and departure rates are exponentially distributed. In addition, the GAN cell capacity C is the amount of the total states of the Markov chain. When a handover-in session arrives in state C without available GAN resource, a handoff failure happens and the session is forced to be terminated. By extending the models presented in [10], [14] to heterogeneous networks and the pre-registration behavior in GAN, the GAN resource requesting rate and the GAN resource holding time can be derived. The handoff failure probability then can be evaluated by using *Erlang's loss formula* [23].

As discussed in Sections II and III, only an MS moving from GERAN/UTRAN to GAN will reserve resources in the GAN cell before the MS moves in. For an MS moving from GAN to GERAN/UTRAN, there is no pre-registration as that discussed in Section III. Thus, new sessions originated
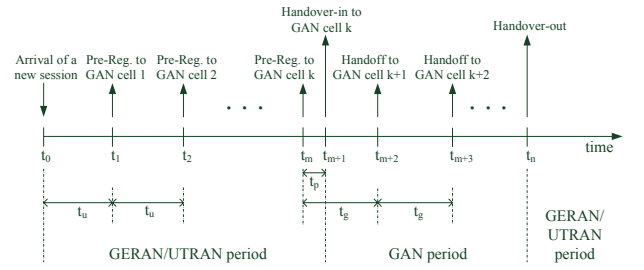
in GAN cells are not related to the problem studied in this paper. Therefore, we assume that new sessions are only originated from GERAN/UTRAN cells. There are both new sessions and handoff sessions in GERAN/UTRAN cells, but only handoff sessions in GAN. In GERAN/UTRAN cells, we assume the arrival of a new session follows Poisson process with mean arrival rate λ_o. The type of the new arrival is binomially distributed as GAN-preferred or GERAN/UTRAN-preferred with probability r_g and $(1 - r_g)$, respectively. We also assume the session holding time of the new session is exponentially distributed with mean $1/\mu$ and the density function $f_c(t) = \mu e^{-\mu t}$.

After staying in one cell for a certain period of time, an MS will move to another cell. The moving direction is uniformly randomized. Before moving, the MS may be in a GAN or GERAN/UTRAN cell. We denote P_g as the probability the MS will move to a GAN cell, where $P_g = N_g/N_{nb}$. Thus, the probability the MS will move to a GERAN/UTRAN cell is $(1 - P_g)$. From the perspective of a cell, the probability that an MS is handed off from a GAN cell is also P_g. The probability that an MS is handed off from a GERAN/UTRAN cell is $(1 - P_g)$. After staying in the GERAN/UTRAN cell for a certain period of time denoted as t_u, the MS pre-registers to the GAN for the next handoff. If the pre-registration is accepted and handover-in is performed, the GAN resource holding time for the MS is denoted as t_g which includes *pre-registration period* and GAN cell residence time. The pre-registration period is defined as the time from the pre-registration is accepted until the MS switches to the GAN cell. We assume t_u and t_g are exponentially distributed with mean $1/\eta_u$ and $1/\eta_g$, respectively. We also define $f_{n_u}(t)$ and $f_{n_g}(t)$ as the density functions, where $f_{n_u}(t) = \eta_u e^{-\eta_u t}$ and $f_{n_g}(t) = \eta_g e^{-\eta_g t}$.

For a GAN-preferred MS, the pre-registration period is 0 because it performs handover-in immediately right after the pre-registration is accepted. If a GAN-preferred MS hands off to a GERAN/UTRAN cell, it implies that the GAN pre-registration is failed. Therefore, for a GAN-preferred MS, its GAN cell residence time equals t_g, and GERAN/UTRAN cell residence time is equal to t_u.

When an MS moves between cells, it is possible that the MS registers with GAN several times but does not handover in. Recall that although a GERAN/UTRAN-preferred MS will register with GAN, it will not handover in unless there is no GERAN/UTRAN available. Fig. 5 shows an example for that. We assume at t_0 a new GERAN/UTRAN-preferred

MS arrives at a GERAN/UTRAN cell. Assuming at t_1, a GERAN/UTRAN-preferred MS pre-registers to GAN but does not handover in. The MS may stay in the same cell or move to another GERAN/UTRAN cell. In Fig. 5, we assume that at t_1, the MS moves to another GERAN/UTRAN cell. The period of time from t_0 to t_1 is exponentially distributed with mean $1/\eta_u$. Because the MS performs pre-registration without handover-in, we call this as *pre-registration miss*. Fig. 5 illustrates another pre-registration miss at t_2. Assuming at t_m, the MS performs another registration with GAN. At this time, we assume no other GERAN/UTRAN cell is available. Therefore, the MS performs handover-in at t_{m+1}. We call the time between t_m and t_{m+1} as *pre-registration period* and denote it as t_p. Let r_p be the ratio of the pre-registration period over GAN resource holding time, i.e., $r_p = t_p/t_g$. Depending on the values of t_p and t_g, the value of r_p will be different. However, the pre-registration period is still exponentially distributed. The mean is r_p/η_g. The GAN cell residence time, from t_{m+1} to t_{m+2}, is set to be $(t_g - t_p)$ which is exponentially distributed with mean $(1-r_p)/\eta_g$. For the handoff between GAN cells, there is no pre-registration period. Therefore, the GAN cell residence time $(t_{m+3}-t_{m+2})$ equals t_g.

From Fig. 5, we can see that pre-registration miss is caused only by a GERAN/UTRAN-preferred MS. Each pre-registration miss also causes an exponential period of time with mean r_p/η_g to occupy extra GAN resource. By modeling each GAN resource as a server to serve handoff sessions, the GAN cell can be modeled as an $M/M/C/C$ system, where C is the GAN cell capacity.

B. Handoff Failure Analysis

There will be a handoff failure when there is no GAN resource available for an MS handing off into GAN. Let t_c be the life time of a session. During the session holding time, suppose that an MS will visit X cells without handoff failure. Let t_{N_i} be the network residence time in the ith cell with density function $f_{n_i}(t)$. The probabilities of $X = k$ and $X > k$, $k \geq 1$, are:

$$
\begin{aligned}
&\Pr[X = k] \\
&= \Pr[t_{N_1} + t_{N_2} + \cdots t_{N_{k-1}} < t_c <= t_{N_1} + t_{N_2} + \cdots t_{N_k}] \\
&= \int_{t_1=0}^{\infty} \int_{t_2=0}^{\infty} \cdots \int_{t_k=0}^{\infty} \int_{t=t_1+t_2+\cdots t_{k-1}}^{t=t_1+t_2+\cdots t_k} f_c(t) \cdot \\
&\quad f_{n_1}(t_1) f_{n_2}(t_2) \cdots f_{n_k}(t_k) dt dt_k \cdots dt_2 dt_1 \\
&= \int_{t_1=0}^{\infty} \cdots \int_{t_k=0}^{\infty} (1 - e^{-\mu t_k}) e^{-\mu(t_1+t_2+\cdots t_{k-1})} \cdot \\
&\quad f_{n_1}(t_1) \cdots f_{n_k}(t_k) dt_k \cdots dt_1 \\
&= [\int_{t_k=0}^{\infty} f_{n_k}(t_k) dt_k - \int_{t_k=0}^{\infty} e^{-\mu t_k} f_{n_k}(t_k) dt_k] \cdot \\
&\quad \prod_{i=1}^{k-1} \int_{t_i=0}^{\infty} e^{-\mu t_i} f_{n_i}(t_i) dt_i \\
&= [1 - \int_{t_k=0}^{\infty} e^{-\mu t_k} f_{n_k}(t_k) dt_k] \prod_{i=1}^{k-1} \int_{t_i=0}^{\infty} e^{-\mu t_i} f_{n_i}(t_i) dt_i
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&\Pr[X > k] \\
&= \int_{t_1=0}^{\infty} \int_{t_2=0}^{\infty} \cdots \int_{t_k=0}^{\infty} \int_{t=t_1+t_2+\cdots t_k}^{\infty} f_c(t) \cdot \\
&\quad f_{n_1}(t_1) f_{n_2}(t_2) \cdots f_{n_k}(t_k) dt dt_k \cdots dt_2 dt_1 \\
&= \int_{t_1=0}^{\infty} \cdots \int_{t_k=0}^{\infty} e^{-\mu(t_1+t_2+\cdots t_k)} \cdot \\
&\quad f_{n_1}(t_1) f_{n_2}(t_2) \cdots f_{n_k}(t_k) dt_k \cdots dt_2 dt_1 \\
&= \prod_{i=1}^{k} \int_{t_i=0}^{\infty} e^{-\mu t_i} f_{n_i}(t_i) dt_i
\end{aligned}
\tag{2}
$$

The Laplace-Stieltjes transform of $f_{n_i}(t_i)$ is:

$$
f_{n_i}^*(s) = \int_{t_i=0}^{\infty} e^{-st} f_{n_i}(t_i) dt_i
\tag{3}
$$

Initially, $f_{n_1}(t_1) = f_{n_u}(t_1)$ because the new MS arrives in the GERAN/UTRAN cell. Hence,

$$
\Pr[X = k] = f_{n_u}^*(\mu)[1 - f_{n_k}^*(\mu)] \prod_{i=2}^{k-1} f_{n_i}^*(\mu)
\tag{4}
$$

$$
\Pr[X > k] = f_{n_u}^*(\mu) \prod_{i=2}^{k} f_{n_i}^*(\mu)
\tag{5}
$$

In the analytic topology, $f_{n_i}^*(\mu)$ is equal to either $f_{n_g}^*(\mu)$ or $f_{n_u}^*(\mu)$, where

$$
f_{n_g}^*(\mu) = \int_{t=0}^{\infty} e^{-\mu t} \eta_g e^{-\eta_g t} dt = \frac{\eta_g}{\mu + \eta_g}
\tag{6}
$$

$$
f_{n_u}^*(\mu) = \int_{t=0}^{\infty} e^{-\mu t} \eta_u e^{-\eta_u t} dt = \frac{\eta_u}{\mu + \eta_u}
\tag{7}
$$

Let P_f be the probability of handoff failure to GAN. If a handoff failure happens, the session is forced to be terminated. In the analytic topology, the numbers of visited GERAN/UTRAN cells and GAN cells by an MS are binomially distributed. Before the session is terminated, suppose that U is the number of moving-outs from GERAN/UTRAN cells. Similarly, G is the number of moving-outs from GAN cells. When handoff failure happens, the probability of $U = u$ with $u \geq 1$ is:

$$
\begin{aligned}
&P_1[U = u] \\
&= f_{n_u}^*(\mu) P_g P_f \cdot \\
&\quad \sum_{g=0}^{\infty} \binom{g+u-1}{g} [P_g(1 - P_f) f_{n_g}^*(\mu)]^g [(1 - P_g) f_{n_u}^*(\mu)]^{u-1}
\end{aligned}
\tag{8}
$$

When the session is terminated at a GAN cell without handoff failure, the probability of $U = u$ with $u \geq 1$ is:

$$
\begin{aligned}
&P_2[U = u] \\
&= f_{n_u}^*(\mu) P_g(1 - P_f)[1 - f_{n_g}^*(\mu)] \cdot \\
&\quad \sum_{g=0}^{\infty} \binom{g+u-1}{g} [P_g(1 - P_f) f_{n_g}^*(\mu)]^g [(1 - P_g) f_{n_u}^*(\mu)]^{u-1}
\end{aligned}
\tag{9}
$$

When the session is terminated at a GERAN/UTRAN cell without handoff failure, the probability of $U = u$ with $u \geq 1$

is:

$$
\begin{aligned}
&\mathrm{P_3}[U = u] \\
&= f_{n_u}^*(\mu)(1 - \mathrm{P}_g)[1 - f_{n_u}^*(\mu)] \cdot \\
&\sum_{g=0}^{\infty} \binom{g+u-1}{g}[\mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu)]^g[(1 - \mathrm{P}_g)f_{n_u}^*(\mu)]^{u-1} \quad (10)
\end{aligned}
$$

By integrating Eq. (8) to (10), we can get:

$$
\begin{aligned}
&Pr[U = u] \\
&= \mathrm{P_1}[U = u] + \mathrm{P_2}[U = u] + \mathrm{P_3}[U = u] \\
&= f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)] \cdot \\
&\sum_{g=0}^{\infty} \binom{g+u-1}{g}[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^g[f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^{u-1} \quad (11)
\end{aligned}
$$

When the session is terminated, the expected number of moving-outs from GERAN/UTRAN cells, $E[U]$, is derived as:

$$
\begin{aligned}
E[U] &= \sum_{u=1}^{\infty} u \Pr(U = u) \\
&= f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)] \cdot \\
&\sum_{u=1}^{\infty}\sum_{g=0}^{\infty} u \binom{g+u-1}{g}[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^g[f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^{u-1} \\
&= f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)] \cdot \\
&\sum_{i=0}^{\infty}\sum_{u=0}^{i} (u + 1) \binom{i}{u}[f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^u[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-u}
\end{aligned}
$$
$$(12)$$

Let

$$
f(u) = \binom{i}{u}[f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^u[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-u} \quad (13)
$$

We then construct a moment generating function $M_u(t)$ as:

$$
\begin{aligned}
M_u(t) &= \sum_{u=0}^{i} e^{t(u+1)}f(u) \\
&= \sum_{u=0}^{i} e^t \binom{i}{u}[e^t f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^u[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-u} \\
&= e^t[e^t f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^i \quad (14)
\end{aligned}
$$

The first differential of $M_u(t)$ is:

$$
\begin{aligned}
M_u{'}(t) &= e^t[e^t f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^i + \\
&e^t i[e^t f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-1} \cdot \\
&e^t f_{n_u}^*(\mu)(1 - \mathrm{P}_g) \quad (15)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{u=0}^{i} (u + 1) \binom{i}{u}[f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^u[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-u} \\
&= M_u{'}(0) \\
&= [f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^i + \\
&i \cdot f_{n_u}^*(\mu)(1 - \mathrm{P}_g)[f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-1}
\end{aligned}
$$
$$(16)$$

By integrating Eq. (12) and Eq. (16), $E[U]$ is derived as:

$$
\begin{aligned}
E[U] &= f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)] \cdot \\
&\{\sum_{i=0}^{\infty}[f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^i + \\
&\sum_{i=0}^{\infty} i \cdot f_{n_u}^*(\mu)(1 - \mathrm{P}_g) \cdot \\
&[f_{n_u}^*(\mu)(1 - \mathrm{P}_g) + f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^{i-1}\} \\
&= f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)] \cdot \\
&\{\frac{1}{1 - f_{n_u}^*(\mu)(1 - \mathrm{P}_g) - f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)} + \\
&\frac{f_{n_u}^*(\mu)(1 - \mathrm{P}_g)}{[1 - f_{n_u}^*(\mu)(1 - \mathrm{P}_g) - f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^2}\} \\
&= \frac{f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu)]}{1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)}
\end{aligned}
$$
$$(17)$$

By using the same technique, we can derive $E[G]$, the expected number of moving-outs from GAN cells before a session is terminated, as:

$$
\begin{aligned}
&Pr[G = g] \\
&= f_{n_u}^*(\mu)[1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)] \cdot \\
&\sum_{i=0}^{\infty} \binom{i+g}{i}[f_{n_u}^*(\mu)(1 - \mathrm{P}_g)]^i[f_{n_g}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)]^g \quad (18)
\end{aligned}
$$

$$
\begin{aligned}
E[G] &= \sum_{g=0}^{\infty} g \Pr(G = g) \\
&= \frac{f_{n_u}^*(\mu)\mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu)}{1 - \mathrm{P}_g(1 - \mathrm{P}_f)f_{n_g}^*(\mu) - (1 - \mathrm{P}_g)f_{n_u}^*(\mu)} \quad (19)
\end{aligned}
$$

Let λ_{hu} and λ_{hg} be the handoff rates to a GAN cell from neighboring GERAN/UTRAN cells and GAN cells, respectively. We also define λ_h as the total handoff rate into a GAN cell. When given the new session arrival rate λ_o, the λ_{hu} and λ_{hg} can be derived as:

$$
\lambda_{hu} = (N_{nb} - N_g)\lambda_o E[U] \quad (20)
$$

$$
\lambda_{hg} = (N_{nb} - N_g)\lambda_o E[G] \quad (21)
$$

Thus, λ_h equals $\lambda_{hu} + \lambda_{hg}$.

Let λ_r be the resource request rate to a GAN cell from neighboring cells. Hence:

$$
\begin{aligned}
\lambda_r &= \frac{\lambda_h}{N_{nb}}r_g + \frac{\lambda_{hg}}{N_{nb}}(1 - r_g) + \\
&(\frac{1}{N_{nb}} + \frac{N_{nb} - N_g}{N_{nb}} \cdot \frac{1}{N_g})\lambda_{hu}(1 - r_g) \quad (22)
\end{aligned}
$$

In Eq. (22), the pre-registration miss comes from GERAN/UTRAN-preferred MSs in neighboring GERAN/UTRAN cells. Let λ_m be the rate of pre-registration miss, which can be derived as:

$$
\lambda_m = \frac{N_{nb} - N_g}{N_{nb}} \cdot \frac{1}{N_g} \cdot \lambda_{hu}(1 - r_g) \quad (23)
$$

The density function of GAN resource holding time $f_{rh}(t)$

is:

$$f_{rh}(t) = \int_{t_c=t}^{\infty} f_c(t_c)f_{n_g}(t)dt_c + \int_{t_n=t}^{\infty} f_c(t)f_{n_g}(t_n)dt_n$$
$$= (\mu + \eta_g)e^{-(\mu+\eta_g)t} \qquad (24)$$

In the case of static keep-alive interval, the busy period of the GAN resource is:

$$\frac{\lambda_m}{\mu + \frac{\eta_g}{r_p}} + \frac{\lambda_r - \lambda_m}{\mu + \eta_g} \qquad (25)$$

By *Erlang's loss formula* [23], the handoff failure probability in the case of static keep-alive interval, denoted as P_{f_static}, is:

$$P_{f\_static} = \frac{1}{C!}(\frac{\lambda_m}{\mu + \frac{\eta_g}{r_p}} + \frac{\lambda_r - \lambda_m}{\mu + \eta_g})^C \cdot$$
$$\{\sum_{i=0}^{C} \frac{1}{i!}(\frac{\lambda_m}{\mu + \frac{\eta_g}{r_p}} + \frac{\lambda_r - \lambda_m}{\mu + \eta_g})^i\}^{-1} \qquad (26)$$

In Eq. (26), λ_r and λ_m are the functions of P_f. By substituting P_{f_static} for P_f, P_{f_static} can be calculated by the iteration method [24] until the value of P_{f_static} is converged.

In the proposed AKI, a pre-registered session will not be preempted during the initial keep-alive interval, denoted as I_k, after a successful pre-registration. That is, a pre-registered session within the I_k period has higher priority. Let P_h be the probability that a pre-registered session is still in higher priority when a new GAN resource request arrives. P_h is derived as:

$$P_h = \int_{t=0}^{t=I_k} \frac{\eta_g}{r_p}e^{-\frac{\eta_g}{r_p}t}dt = 1 - e^{-\frac{\eta_g}{r_p}I_k} \qquad (27)$$

Assuming a session which was preempted can request GAN resource again when it needs to perform handover-in. If the high-priority pre-registration misses in AKI can occupy GAN resource for the whole pre-registration period t_p, the busy period of the GAN resource is at most:

$$\frac{\lambda_m}{\mu + \frac{\eta_g}{r_p}} \cdot P_h + \frac{\lambda_r - \lambda_m}{\mu + \eta_g} \qquad (28)$$

Thus, the upper bound of the handoff failure probability in AKI is:

$$P_{f\_AKI} = \frac{1}{C!}[\frac{\lambda_m(1 - e^{-\frac{\eta_g}{r_p}I_k})}{\mu + \frac{\eta_g}{r_p}} + \frac{\lambda_r - \lambda_m}{\mu + \eta_g}]^C \cdot$$
$$\{\sum_{i=0}^{C} \frac{1}{i!}[\frac{\lambda_m(1 - e^{-\frac{\eta_g}{r_p}I_k})}{\mu + \frac{\eta_g}{r_p}} + \frac{\lambda_r - \lambda_m}{\mu + \eta_g}]^i\}^{-1} \qquad (29)$$

Again, by using the iteration technique, the upper bound of the handoff failure probability in AKI can be derived. Because the busy period in AKI is less than that in the static keep-alive interval case, P_{f_AKI} is also less than P_{f_static}. Therefore, it is proved that the proposed AKI outperforms the static keep-alive interval technique in handoff failure probability.

The effectiveness of GAN resource utilization is also an important performance metric. In this paper, an *effective GAN resource usage* is defined as the actual time the GAN resource is used by an MS, i.e., $t_g - t_p$ in Fig. 5. Furthermore, GAN utilization is defined as the ratio of the effective GAN resource

usages over the total occupancy and free time of all the GAN resources. With the handoff failure probability derived above, GAN resource utilization ρ can be derived as:

$$\rho = \frac{(1-P_f)}{CN_{nb}} \cdot \frac{\lambda_h}{\mu + \eta_g} - \frac{(1-P_f)}{CN_{nb}} \cdot \frac{\lambda_{hu}(1 - r_g)}{\mu + \frac{\eta_g}{r_p}} \qquad (30)$$

For the GERAN/UTRAN-preferred MSs, the resource occupancy time in pre-registration period must be eliminated. With the improvement of handoff failure probability by AKI, GAN utilization is also improved.

VI. SIMULATION AND NUMERIC RESULTS

The analysis presented in Section V is validated by extensive simulations by using ns-2 [25], version 2.34. The parameters used in the simulations are listed in Table I. In order to measure the handoff failure probability, the mean session holding time is set to be greater than or equal to mean cell residence time to generate huge amount of handoff events. IEEE 802.11 WLAN is chosen as the radio for GAN. For VoIP sessions, because G.711 codec [26] is commonly used over RTP, we adopt G.711 codec in the simulations. According to [27], an AP can guarantee the QoS of at most 12 simultaneous G.711 VoIP sessions. In Figs. 6–13, the lines represent the analytical results while the simulation results are presented by points. The simulations were conducted with 95% confidence level. However, the confidence intervals are not drawn here because they are too small and will overlap with other lines and points significantly which will make the figures very difficult to read.

For the mathematical analysis, we analyze the worst case of the proposed AKI, which may be the upper bound or lower bound depending on the performance metric. In Figs. 6–9, for instance, the analysis shows the upper bound of the handoff failure probability. In Figs. 10–13, on the other hand, the analysis shows the lower bound of the GAN utilization. The reason to show the performance bounds of the proposed AKI is that there may be preemption in the proposed AKI. Without *pre-registration miss*, the resource occupancy time is always counted as the pre-registration period plus the GAN cell residence time no matter whether preemption is performed or not. With pre-registration miss, the GAN resource occupancy time in the pre-registration period is over evaluated by counting the time of some candidate MSs. Once a pre-registered MS becomes a preemption candidate, it has no effect on the GAN resource occupancy time. Therefore, the mathematical analysis shows the upper bound of the handoff failure probability and the lower bound of the GAN utilization. Without preemption in the *static keep-alive interval* in which the keep-alive interval remains static as that defined in the standards, the resource occupancy time can be derived accurately in mathematical analysis. Nevertheless, Figs. 6–13 show that even in the worst case, the performance of AKI is still better than that of the *static keep-alive interval*.

When traffic load is heavy, in order to allocate resource for an MS which will use GAN service soon, the serving GANC releases some reserved resources as that described in Section IV-B. Figs. 6–13 also illustrate the simulation results of the three different policies (maximum interval first,

minimum interval first, random). We also compare them with the results of the *static keep-alive interval*. An interesting point is that the three different policies in the proposed AKI perform similarly. The reason for that will be discussed later. Again, they all outperform the *static keep-alive interval*.

Figs. 6–9 show the results of the handoff failure probability into GAN. In Fig. 6, AKI reduces the handoff failure probability when the *GERAN/UTRAN-preferred session ratio* increases. This is because the GERAN/UTRAN-preferred users who do not prefer to switch into GAN are likely to become candidates to be preempted. Therefore, the handoff sessions in the proposed AKI have more chances to get the requested GAN resources. In Fig. 7, when r_p is increased (recall that $r_p = t_p/t_g$), it is earlier to pre-register. When every MS pre-registers earlier, the GAN resources will be exhausted very soon. Therefore, the handoff failure probability is increased. However, by using AKI, Fig. 7 shows that the handoff failure probability can be reduced significantly. Fig. 7 also indicates that without AKI, the handoff failure probability increases linearly. With AKI, the upper bound of the handoff failure probability remains constant. Fig. 8 shows that the handoff failure probability is also affected by new session arrival rate. Again, AKI can reduce the handoff failure probability when session arrival rate is increased. By re-allocating GAN resource to those users who need resources immediately, AKI can alleviate the congestion in GAN. Therefore, AKI makes GAN more tolerant to heavy user traffic. The performance of different GAN capacities is also investigated. When a GAN has more capacity, it can accommodate more handoff sessions. Therefore, as shown in Fig. 9, when the GAN capacity is increased, the handoff failure probability is decreased. Fig. 9 shows that the proposed AKI is particularly useful when the load of the GAN cell is heavy.

Figs. 10–13 depict the results of GAN utilization. The GAN utilization is defined as the proportion of the GAN resources allocated to the move-in sessions. The discussion for GAN utilization is similar to that above. Because the reserved resource in GAN can be preempted in AKI, the GAN resources are used more efficiently. In Fig. 10, because the GERAN/UTRAN-preferred sessions may occupy the GAN resource but not use it immediately, GAN utilization decreases when the GERAN/UTRAN-preferred session ratio increases. In Fig. 11, when r_p is increased, it is earlier to pre-register. Thus, the GAN utilization is decreased. Fig. 12 shows that when session arrival rate increases, the GAN utilization increased too. In Fig. 13, when the capacity of a GAN cell is increased, there are more resources which are not used. By looking at Eq. (30), we can see that reducing handoff failure probability is also beneficial to GAN utilization, ρ. When more sessions move into GAN successfully, the GAN resources are allocated more efficiently to the sessions which are really in-used. Figs. 10–13 indicate that with AKI, the GAN utilization is always increased when comparing with the *static keep-alive interval*.

More simulation results are presented in Table II and Table III which show the improvement of the proposed AKI in percentage when comparing with the *static keep-alive interval*. In Table II and Table III, the GERAN/UTRAN-preferred ratios of 0.2, 0.5, and 0.8 represent the low den-
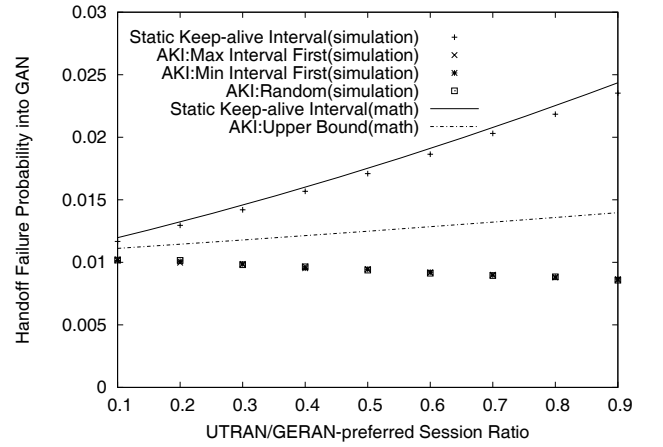

Fig. 6.   Handoff failure probability versus various GERAN/UTRAN-preferred session ratios.
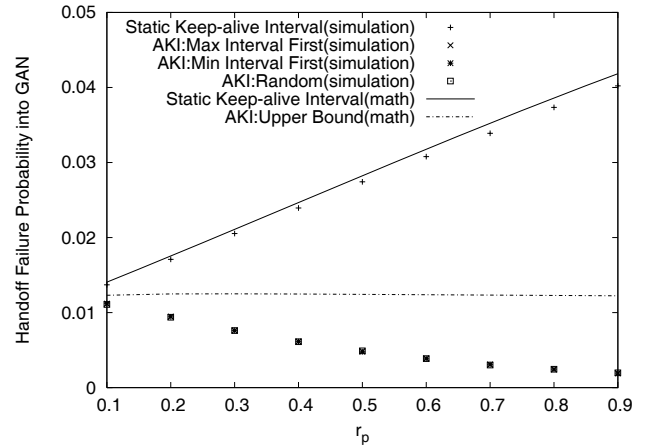

Fig. 7.   Handoff failure probability versus various pre-registration period ratios.

sity, median density, and high density, respectively, of the GERAN/UTRAN-preferred sessions in the systems. Based on the simulation and mathematical results, one can see that the proposed AKI is effective to reduce the handoff failure probability and increase the GAN utilization. The results also show that the three different policies to preempt an MS have similar performance results. This is because the movement behavior of an MS is *memoryless* to the GANC. That is, the movement of next MS is independent of current MS. Without accurate movement prediction, it is hard to say whether the earliest or latest pre-registered MS will move into GAN first. Sometimes it is better to preempt the maximum keep-alive interval session, but sometimes it is better to preempt the one with minimum keep-alive interval. Overall, the performance will be similar because the movement of MSs is a random process. Therefore, a simple way to preempt a reserved session randomly can improve the system performance. We have also conducted more comprehensive experiments with many different parameter settings [28]. They all support the same conclusions we obtain here.

The proposed AKI is simple. It does not need to predict the movement of MSs. Section IV also shows that the proposed AKI has very low complexity in terms of memory space and computational time. The simulation results also show that AKI can outperform the *static keep-alive interval*, the one defined
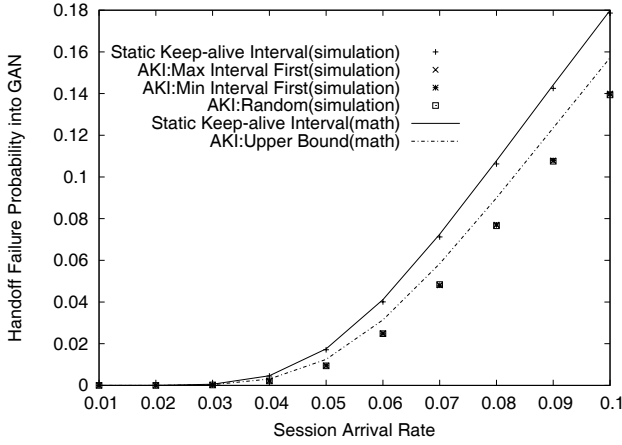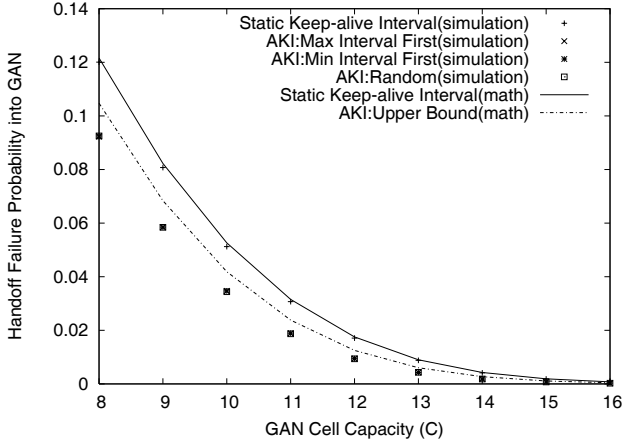
Fig. 8.　Handoff failure probability versus various session arrival rates.

Fig. 10.　GAN utilization versus various GERAN/UTRAN-preferred session ratios.

Fig. 9.　Handoff failure probability versus GAN cell capacities ($r_p = 0.2$, $1 - r_g = 0.5$).

Fig. 11.　GAN utilization versus various pre-registration period ratios.

in the standards. By scrutinizing Eq. (26)(29)(30), however, we found that the performance of AKI can be further improved if we can bound r_p value to a minimum value. This can be achieved if we can predict the movement of MSs *accurately*. By movement prediction, we can also preempt the MS which is going to handoff first. However, accurate prediction of MS movement incurs many extra costs, which may include both software and hardware costs. From the viewpoint of implementation, the proposed AKI just uses the keep-alive behavior which is well-defined in the standards [1], [2]. Only by controlling the keep-alive timer, TU 3906 timer [1], [2], AKI can result in much better performance. It is also compatible with the standards. It does not need to depend on movement prediction techniques. Because AKI is independent of movement prediction techniques, it can incorporate any movement prediction technique to further improve the performance. Many movement prediction techniques have been proposed. They are out of the scope of this paper.

VII. SUMMARY

GAN was proposed in 3GPP to extend mobile services to heterogeneous wireless networks. According to different user preferences, an MS has different strategies to choose the most appropriate RAN. However, it is difficult to effectively assign resources to MSs by the GANC. In this paper, we propose
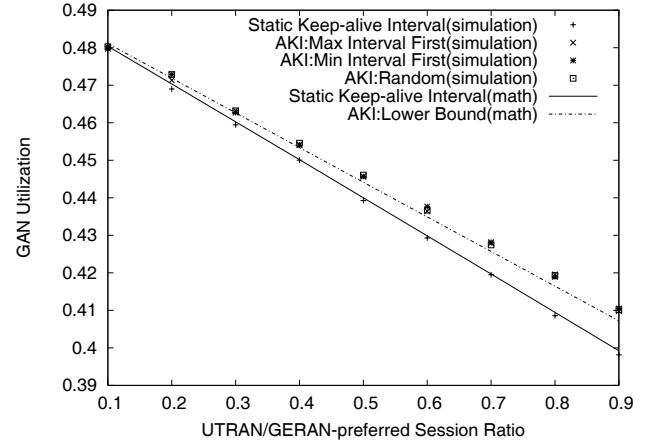
AKI to manage the GAN resources. It can reduce unnecessary GAN signaling messages before actual handover-in happens. Without handoff prediction, a heuristic way to preempt reserved resources is to relate keep-alive interval to the length of pre-registration time. Two extreme (maximum and minimum interval first) policies are evaluated regarding their impact on the AKI performance. Furthermore, random choice as a third alternative is also intuitively included. An interesting result we obtained is that the system performs similarly no matter which policy is used. This is because the movement behavior of an MS is *memoryless* to the GANC. Nevertheless, all of the three policies in AKI outperform the *static keep-alive interval* defined in the standards. The proposed AKI is simple and effective. The innovation of this paper is in that we identify a critical issue which was not solved before and needs to be resolved to improve the GAN performance. We propose a simple yet practical solution which can improve the performance significantly. In addition, in this paper, we develop an intricate analytic model to evaluate the handoff failure probability and GAN utilization. The mathematical model quantifies the performance of the proposed AKI. Also, the analysis shows the performance of a larger scale network, which exams the scalability of the proposed AKI.
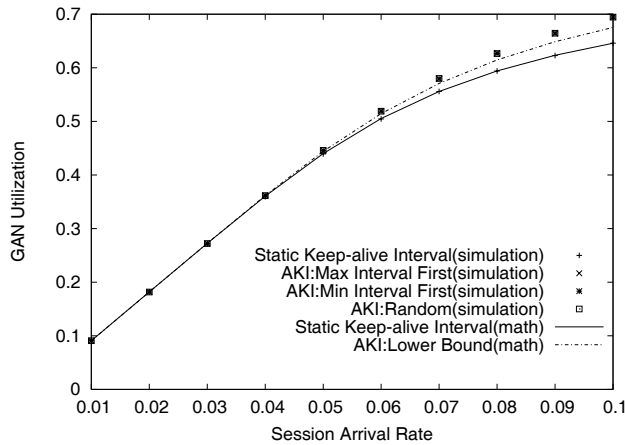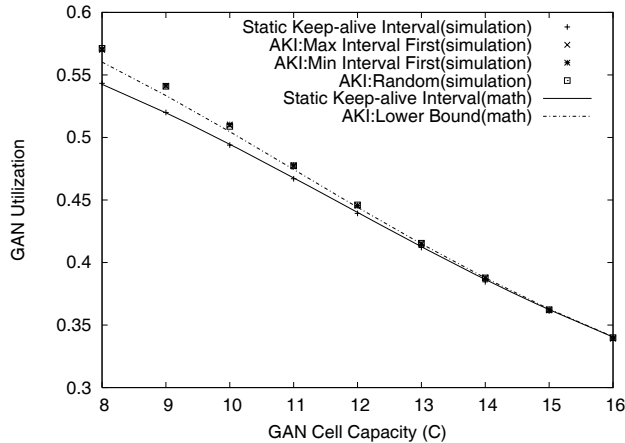
Fig. 12.   GAN utilization versus various session arrival rates.

Fig. 13.   GAN utilization versus various GAN cell capacities ($r_p = 0.2$, $1 - r_g = 0.5$).

TABLE II
HANDOFF FAILURE IMPROVEMENT (SIMULATION λ_o=0.05, C=12)

r_p	$1 - r_g$	Max interval first	Min interval first	Random
0.2	0.2	23.07%	22.22%	21.60%
	0.5	44.65%	44.76%	45.11%
	0.8	59.75%	59.57%	59.52%
0.5	0.2	50.68%	50.93%	50.31%
	0.5	82.21%	82.57%	82.14%
	0.8	93.52%	93.52%	93.64%
0.8	0.2	66.45%	66.61%	66.29%
	0.5	93.52%	93.41%	93.52%
	0.8	98.95%	98.95%	98.95%

TABLE III
UTILIZATION IMPROVEMENT (SIMULATION λ_o=0.05, C=12)

r_p	$1 - r_g$	Max interval first	Min interval first	Random
0.2	0.2	0.50%	0.82%	0.82%
	0.5	1.46%	1.46%	1.55%
	0.8	2.69%	2.53%	2.63%
0.5	0.2	1.70%	1.74%	1.87%
	0.5	4.39%	4.35%	4.47%
	0.8	8.37%	8.50%	8.23%
0.8	0.2	2.62%	2.57%	2.63%
	0.5	7.27%	7.31%	7.32%
	0.8	14.34%	14.37%	14.25%

REFERENCES

[1] "GSM/EDGE radio access network; generic access network (GAN); stage 2," 3GPP Technical Specification Release 9 - TS 43.318 V9.0.0, Dec. 2009.

[2] "GSM/EDGE radio access radio access network; generic access network (GAN); mobile GAN interface layer 3 specification," 3GPP Technical Specification Release 9 - TS 44.318 V9.2.0, Mar. 2010.

[3] "User perspective (stage 1)," UMA Technical Specification R1.0.0, Sep. 2004.

[4] "Architecture (stage 2)," UMA Technical Specification R1.0.4, May 2005.

[5] "Protocols (stage 3)," UMA Technical Specification R1.0.4, May 2005.

[6] J.-C. Chen and T. Zhang, *IP-Based Next-Generation Wireless Networks*. John Wily & Sons, Inc., Jan. 2004. Available: http://www.cs.nctu.edu.tw/~jcc/book.html.

[7] IEEE Std 802.11-2007, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," http://standards.ieee.org/about/get/802/802.11.html, June 2007.

[8] "The Official Bluetooth Technology Info Site." Available: http://www.bluetooth.com/.

[9] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. 35, no. 3, pp. 77–92, Aug. 1986.

[10] Y.-B. Lin, S. Mohan, and A. Noerpel, "Queueing priority channel assignment strategies for PCS hand-off and initial access," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 704–712, Mar. 1994.

[11] B. Li, S. T. Chanson, and C. Lin, "Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks," *Wireless Netw.*, vol. 4, no. 4, pp. 279–290, 1998.

[12] Y. Fang and Y. Zhang, "Call admission control schemes and performance analysis in wireless mobile networks," *IEEE Trans. Veh. Technol.*, vol. 51, no. 2, pp. 371–382, Mar. 2002.

[13] K.-H. Chen and J.-C. Chen, "Improve service availability in 3GPP generic access network (GAN) by adaptive keep-alive interval (AKI)," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2010.

[14] Y.-B. Lin, M.-H. Tsai, and H.-W. Dai, "Bearer reservation with preemption for voice call continuity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2716–2725, May 2009.

[15] W. Song, H. Jiang, and W. Zhuang, "Performance analysis of the WLAN-first scheme in cellular/WLAN interworking," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1932–1943, May 2007.

[16] R. Droms, "Dynamic Host Configuration Protocol," IETF RFC 2131, Mar. 1997.

[17] S. Kent and K. Seo, "Security Architecture for the Internet Protocol," IETF RFC 4301, Dec. 2005.

[18] E. H. Haverinen and E. J. Salowey, "Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP-SIM)," IETF RFC 4186, Jan. 2006.

[19] J. Arkko and H. Haverinen, "Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)," IETF RFC 4187, Jan. 2006.

[20] E. C. Kaufman, "Internet Key Exchange (IKEv2) Protocol," IETF RFC 4306, Dec. 2005.

[21] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 3550, July 2003.

[22] J.-C. Chen, Y.-W. Liu, and L.-W. Lin, "Mobile virtual private networks with dynamic MIP home agent assignment," *Wireless Commun. Mobile Comput.*, vol. 6, no. 5, pp. 601–616, Aug. 2006.

[23] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd edition. John Wiley & Sons, 1998.

[24] G. W. Collins, II, *Fundamental Numerical Methods and Data Analysis*. NASA Astrophysics Data System(ADS), 2003.

[25] "The network simulator - ns-2." Available: http://www.isi.edu/nsnam/ns/.

[26] "Pulse Code Modulation (PCM) of Voice Frequencies," ITU-T Recommendation G.711, 1972.

[27] S. Garg and M. Kappes, "Can I add a VoIP call?" in *Proc. IEEE International Conf. Commun.*, May 2003, pp. 779–783.

[28] K.-H. Chen and J.-C. Chen, "Handoff failure analysis of Adaptive Keep-alive Interval (AKI) in 3GPP Generic Access Network (GAN)," Technical Report, Department of Computer Science, National Chiao Tung University, June 2011.

Kai-Hsiu Chen received his M.S. degree from the Department of Computer Science, National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2005. He is now a Ph.D. candidate in the same department. He was a summer intern at Telcordia Technologies, Piscataway, NJ, USA, in 2009. His research interests include mobility management, admission control, resource management, and performance analysis of wireless networks.

Jyh-Cheng Chen (S'96-M'99-SM'03) received the Ph.D. degree from the State University of New York at Buffalo in 1998. Dr. Chen is Director, Institute of Network Engineering, and Professor, Department of Computer Science, National Chiao Tung University (NCTU), Hsinchu, Taiwan. He was with Bellcore/Telcordia Technologies, Morristown, NJ, USA, 1998-2001, and Telcordia Technologies, Piscataway, NJ, USA, 2008-2010. He has also been with the Department of Computer Science, National Tsing Hua University (NTHU), Hsinchu, Taiwan since 2001 as assistant/associate/full/adjunct professor. He is a coauthor of the book *IP-Based Next-Generation Wireless Networks* (Wiley, 2004). He has published over 80 papers. He is the holder of 19 U.S. patents, six R.O.C. patents, and four P.R.C. patents. Dr. Chen is a Senior Member of the IEEE, and a Senior Member of the ACM. He received the 2000 Telcordia CEO Award, the 2001 SAIC ESTC (Executive Science and Technology Council) Publication Award, the 2004 NTHU New Faculty Research Award, the 2006 NTHU Outstanding Teaching Award, and the 2007 Best Paper Award for Young Scholars, IEEE Communications Society Taipei and Tainan Chapters & IEEE Information Theory Society Taipei Chapter. He is a Technical Editor of the *IEEE Wireless Communications*. He was a guest editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, special issue on "All-IP Wireless Networks," May 2004. He was the Technical Program Co-Chair of the Ninth IFIP/IEEE International Conference on Mobile and Wireless Communications Networks (MWCN '07) held in Ireland, September 2007. He was the Technical Program Co-Chair of the Third IEEE International Conference on Information Technology: Research and Education (ITRE '05). He has been on the technical program committee of numerous international conferences, including IEEE INFOCOM 2005-2006, IEEE GLOBECOM 2005-2009, 2011 and IEEE ICC 2007-2010. He was a Tutorial Speaker at IEEE GLOBECOM 2002, 2003, and 2006 in the subject of next-generation wireless networks. He leads the development of WIRE1x, which is one of the most important implementations of IEEE 802.1x supplicant.