



Solving the uncertainty of vertical handovers in multi-radio home networks[☆]

Jing Wang, R. Venkatesha Prasad^{*}, I.G.M.M. Niemegeers

WMC/IRCTR, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands

ARTICLE INFO

Article history:

Received 12 September 2008

Received in revised form 25 December 2009

Accepted 17 February 2010

Available online 20 February 2010

Keywords:

Vertical handover

Decision theory

Markov decision process

60 GHz radio

Heterogeneous network

ABSTRACT

In this paper, we investigate Decision Theory (DT) and Markov Decision Process (MDP) based approaches for vertical handover decision making in in-home networks. Here, we consider a heterogeneous network environment with both legacy WLAN and novel line-of-sight (LOS)-dependent 60 GHz radio systems deployed in home to support high quality multimedia applications. Both the above decision-making approaches take into account multiple subjective and objective factors, such as user preference, network condition, device capability and impact of the environment. We make decisions based on an evaluation of the candidate actions, but with different length of horizon. We show their ability to effectively make decisions, to handover or not, in uncertain situations. The method and results herein are in general applicable to any other situation where such a decision has to be made.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The proliferation of radio technologies is paving the way for supporting a variety of in-home applications with an increasing demand for sustainable high data rates in the near future. To support high quality multimedia streaming, e.g., HDTV broadcasting, the attention of the research community has been on the higher frequency bands. As advocated by Smulders [1], a promising solution is the license-free 60 GHz band where bandwidth is abundant, and the data rates of the order of 1 Gigabit per second are easily feasible. 60 GHz radio is intended to provide connections for very high data rates within a short distance, mainly in the in-door environment. To save transmission power while ensuring satisfactory link quality at high data rates, a directional antenna configuration is recommended for the system to counteract the severe attenuation and to combat multi-path effects [2]. Therefore, line-of-sight (LOS) propagation has been one of the requirements for 60 GHz radio. However, LOS propagation cannot be guaranteed, especially in the in-door environment with many potential obstructions due to human activities and other objects in the surroundings. Experimental results in [16] show that a mean attenuation value of 22 dB is observed when a person obstructs the 60 GHz LOS link.

The trend in the in-home networking is that the network is more heterogeneous and supports multiple radio systems. The problem due to vulnerable 60 GHz LOS transmission can thus be resolved by a vertical handover of a communication session from one radio to another to maintain the continuity of the session. Fig. 1 gives an example, where both a 60 GHz radio cell and a WLAN cell are formed with a single access point, and are overlapping inside the house by using Radio-over-Fibre (RoF) technology [3]. For multimedia content distribution, streaming via a 60 GHz LOS link is always a preferable option as it is able to achieve the high data rate that is enough for the media content with high resolution. Thus it can offer the users a better perceptible service quality. If the LOS link is lost, a strong degradation of the perceived quality is very likely due to the disruption of the session. However, the LOS blocking is often a temporary phenomenon lasting for a short period. Nonetheless on a few occasions it can last for many seconds or even longer. Therefore, it is necessary to perform handover in some cases, whereas we just need to wait for the end of blocking in other cases. Switching back and forth will cause additional cost (power, display pause, etc.). Thus it is critical for the system to make a decision in time to handover. In fact, whenever the LOS link is lost a decision has to be made – whether to switch to the backup WLAN for streaming multimedia content under reduced quality, or to wait without switching hoping that this disturbance is just a transient effect. If the LOS link recovers rapidly with a possible short blocking duration then “waiting” action avoids unnecessary switching. However, if the outage of the link is longer, then “switching” action may avoid session break down. Thus the goal is to minimize perceptible quality degradation and make a well informed decision to switch between the available radios.

[☆] This work was supported by Dutch IOP GenCom “Future Home Networks” project.

^{*} Corresponding author.

E-mail addresses: Jing.Wang@tudelft.nl (J. Wang), R.R.VenkateshaPrasad@tudelft.nl, rvprasad@gmail.com (R. Venkatesha Prasad), I.G.M.M.Niemegeers@tudelft.nl (I.G.M.M. Niemegeers).

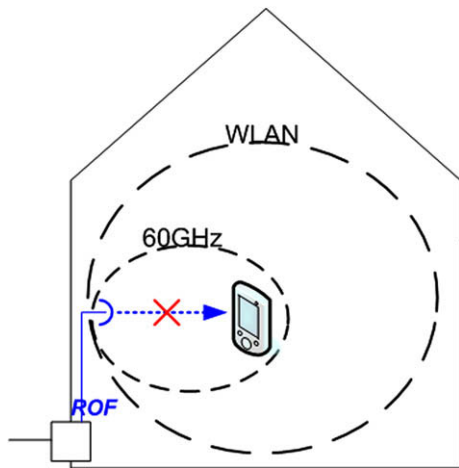


Fig. 1. Indoor radio links.

The main focus of this work is to devise methods to tackle and study this problem.

In this work, we designed and examined two decision algorithms to solve this uncertainty of vertical handovers in 60 GHz radio system as shown in Fig. 1. The first algorithm is based on the decision theory (DT), where the handover decision is modeled as an episodic decision problem and actions are evaluated using the outcome of the current state. The second algorithm uses Markov Decision Process (MDP), where the decision problem is considered to be sequential. It is different from the decision theory approach where the decision depends not only on the outcome of the current state but also on the states that evolve afterwards. Both the algorithms are able to take into account multiple factors to make decisions in uncertain situations based on an evaluation of the candidate actions, but with different time horizon. Without loss of generality, both methods are potentially applicable when a decision has to be made amongst available multiple radio systems when the link quality of the current radio deteriorates significantly. Further, this type of uncertainty can be seen in many other situations such as selecting a route in a network with frequently failing paths. Thus we believe that this study can be altered to suit various such situations with slight modifications.

The rest of the paper is organized as follows. Section 2 provides a brief survey of the earlier studies for vertical handover decision. Section 3 gives an overview of the system model. In Sections 4 and 5, the vertical handover decision-making task is modeled with the Decision Theory and Markov Decision Process, respectively. Section 6 shows the detailed simulation setup, results and discussions. In the end, in Section 7 we conclude this work and discuss the future enhancements.

2. Related studies

Traditionally, vertical handover decisions are simply made based on comparing different radio systems in terms of link and network level performance indicators, such as, the received signal strength and throughput. de Sousa et al. [4], examined four algorithms – load balancing algorithm, coverage threshold algorithm, rate maximizing algorithm and theoretical circuit switched equivalent algorithm. They show through simulations that by including further information such as location of the users, the handover performance could be greatly improved. Recently, more sophisticated decision algorithms largely based on Artificial Intelligence (AI) techniques are getting prominence. AI techniques are used to exploit relevant factors from networks and devices to better the vertical handover strategies. A vertical handover algorithm based on

fuzzy control theory [5] considers multiple criteria such as information about the load, velocity of mobile terminal and a set of rules defined from *a priori* knowledge. Following similar principle, a decision algorithm has been proposed in [6] taking into account power levels of received signals, cost of operation of a particular network and the amount of unused bandwidth. In [8] vertical handover problem is formulated as a Markov decision process, where link reward and signaling cost functions are introduced to evaluate actions. However, the reward is measured based on the bandwidth and connection delay, and thus the decision is insulated from the user's perceived experience of the service. Perceived quality of service, nonetheless, is of great importance for emerging user-centric networking paradigm [9]. In [7] a context-aware decision algorithm based on Analytic Hierarchy Process (AHP) is designed considering both static and dynamic context of the user, characteristics of terminals and the network. The networks are ranked with great consideration of the users' preference before the mobile terminal executes a handover. However, the ranking algorithm lacks sophistication and is formulated in a rather simplistic manner with coarse granularity.

It has always been a challenge for networks and/or mobile terminals to solve the issue as described in our example shown in Fig. 1. On one hand, the dynamic and random nature of occurrences of blocking events bring high uncertainty into the handover decision making; on the other hand, the decision should be made under substantial influence of multiple subjective and objective factors such as user preferences, network conditions, device capabilities and environmental impact. This kind of uncertain and complex situations is thus fuelling the need for sophisticated networking approaches and decision algorithms. This paper attempts to tackle this problem in networks where 60 GHz and WLAN are present; and in general, networks with multiple radios.

3. System model

3.1. Users' input

User preferences describe the users' experience of playing multimedia content which is delivered by the network with different radio systems. It is typically represented as discrete values or scores similar to ITU-T's Mean Opinion Score (MOS) [10]. The scores are usually found according to the rankings given by the users to quantitatively indicate their preferences on one radio system over the other. In this work, we use the so-called *utility* or *reward* to quantify the user experience when media stream is delivered through different radios. The density of this quantity, u , is defined as the subjective mean opinion score in each time unit as,

$$u = \begin{cases} u_{60}, & \text{streaming through 60 GHz radio system,} \\ u_w, & \text{streaming through WLAN system,} \\ u_z, & \text{connecting to neither system,} \\ & \text{ceased streaming and display.} \end{cases}$$

Qualitatively, it is easy to see that $u_{60} > u_w > u_z$.

3.2. The environment

The environmental impact is the occurrence of the blocking events for the sessions. It can be stochastically modeled by taking the duration of blocking and the interval between two blocking events. We model these events using expectation of blocking duration $E(t_{blk})$ and expectation of the duration between two successive blocking events $E(t_{blk})$. With a model it would be easy to acquire some important collective characters of these randomly occurring events. It is also possible to predict the duration of blocking and

interval between blocking events by applying certain learning and prediction techniques.

3.3. Network characteristics

Switching time, t_{swt} , from one radio system is an important aspect. This depends on the network characteristics. The switching time is highly affected by the factors such as, number of users within the radio system, the traffic load, the link quality, etc. It is the sum of two activities: (1) the time to tune a device to the target radio and configure the network, which is mainly the time taken for network association and network address allocation; and (2) the time for session level handshake to switch and continue the on-going session through a different radio system.

3.4. Device capability

Keeping in mind that the multimedia streaming applications practically use jitter buffer to alleviate the transmission delay variance, we consider the buffering capacity of the display device to represent this aspect. For convenience, we measure the buffer size using the buffering time, t_{buf} , to describe how long the buffered multimedia content can be displayed. We use these further in our studies.

4. Model using the decision theory

4.1. The principle

The first proposed decision algorithm is based on the decision theory (DT), which has been largely used in AI for intelligent agents to select an action from several alternatives under uncertain situations. In our case, the uncertainty is due to lack of knowledge of the precise starting time and the duration of each blocking event. When 60 GHz LOS link is lost, the algorithm has to select one of the actions – “handover” or “wait”. The decision is made by evaluating the utility of all possible consequent states led by each action, i.e., the decision is to take the action a^* such that,

$$a^* = \arg \max_{a_i \in A} \sum_{S_{ij} \in S} P_{ij} U(S_{ij}). \quad (1)$$

Here, $U(S_{ij})$ is the utility of state S_{ij} , which is the consequent state of action a_i with probability P_{ij} . Instead of utility per se, we can also make decisions based on the utility difference between each possible consequent state and the ideal state. There is no blocking in the ideal state, where it is possible to use the preferred 60 GHz LOS link always with high bandwidth. This difference reflects the degradation perceived by a user when 60 GHz LOS is blocked. We choose the action which brings the smallest difference, in other words the lowest degradation. Therefore, we rewrite (1) into (2), and use $\Delta U(S_{ij})$ to represent the utility difference.

$$a^* = \arg \min_{a_i \in A} \sum_{S_{ij} \in S} P_{ij} \Delta U(S_{ij}). \quad (2)$$

Based on (2) our proposed decision algorithm consists of the following five basic steps:

- (1) Identify and collect the information about user preference, network, device and environment, and define decision-making situation.
- (2) Define transitions to consequent states for each action.
- (3) Calculate the probability of consequent states for each action, and the utility difference for each state.
- (4) Calculate the expected utility difference for each action by summing up the utility difference of each consequent state scaled with its probability.

- (5) Compare the expected utility difference for all actions. The action with the smallest degradation will be selected.

Before elaborating the above steps further we list some assumptions below:

- All 60 GHz LOS connections have similar channel quality and offers similar perceptible experience. Further, it is assumed to be the same for WLAN connections.
- When a device is using 60 GHz radio interface, WLAN interface is deactivated to conserve energy and it is periodically activated to detect access points for handover information. While using WLAN connection, 60 GHz interface is always active. In this way, the device can immediately detect 60 GHz LOS link as soon as it is available and switch back to it.
- The stochastic model and prediction of the blocking events are obtainable and derived from a knowledge module. The knowledge module is a logical entity capable of perceiving environmental parameters and it can process parameters to derive the required information by using appropriate learning algorithms.

4.2. Decision-making situation¹

We first describe the decision-making situation. To calculate the expected utility of each state, a quadruple is defined [11]: $D = \langle P_c(S), A, proj, U \rangle$.

U is the utility, quantifying the users' perceived experience. Here, we take the user input – u_{60} , u_w and u_z , which have been defined in Section 3.1, as the measure for utility.

A is the set of actions, $\{a_i\}$, that represents a pool of possible actions. In our case the actions can be a_1 : perform handover to WLAN when 60 GHz radio is lost, or a_2 : do not perform handover and wait till 60 GHz link recovers.

$P_c(S)$ is the probability of initial state while making a decision. In our case, the decision is made only when 60 GHz LOS link is temporarily blocked. Therefore, there is only one initial state S_0 , which is “loss of 60 GHz LOS connection” with probability $P_c(S_0) = 1$.

$proj$ stands for projection function, which represents the mapping from initial state to each possible consequent state S_{ij} under certain action, that is, $proj(P_c(S), a_i) = P(S_{ij})$, which is represented by P_{ij} in (2).

4.3. State transitions

The probability of each consequent state is determined by several factors – duration of blocking event, buffering time and switching time. These factors lead to four possible consequent states, i.e., $S_{ij} \in S : \{S_{11}, S_{12}, S_{21}, S_{22}\}$ as a consequence of two actions as shown in Fig. 2.

State- S_{11} : This is a possible resultant state due to “handover” action. In this case, the switching time is shorter than the buffering time. As enough content is buffered which is able to cover the interrupted streaming during handover, no discontinuity in the content play-out is observable. The perceptible degradation of multimedia streaming quality is only due to the use of low speed WLAN link. As shown in Fig. 3(a), at the beginning of the blocking event when 60 GHz radio link is interrupted the system begins the procedure to switch to WLAN system. After a period of t_{swt} , streaming session is successfully handed over to WLAN, but data buffered via 60 GHz radio system is still displayed with good perceptible quality within the duration of t_{buf} . After that, the displayed content

¹ The term is defined in decision theory to specify the agent's knowledge about the environment, the agent's assessment as to its possible courses of an action, possible results of the actions and desirability of these results [11].

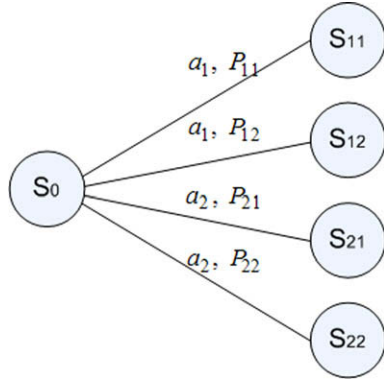


Fig. 2. Consequent states under actions.

is transmitted via WLAN. At the end of the blocking event the system switches back to 60 GHz. After finishing t'_{buf} length of the buffered data by WLAN, multimedia content via 60 GHz link is again displayed.

State- S_{12} : This is another possible resultant state due to the “handover” action. In this case, the switching time is longer than the buffering time. In this case, two infirmities happen: (a) observable degradation due to switching over to a lower speed link, and (b) discontinuity in play-out of the content due to lack of enough buffered data to cover the switching time during handover. In Fig. 3(b), t_z represents the period when the multimedia stops playing after exhausting the buffered packets and before successfully connecting to WLAN system.

State- S_{21} : This is under the action “Wait”. In this case, the device buffers enough multimedia content to cover the relatively short blocking duration, as shown in Fig. 3(c). Therefore, no degradation and no discontinuity will be observed.

State- S_{22} : This is another possible consequent state under the action “Wait”, where the buffered content is not enough to cover the relatively long blocking. Thus the play-out will cease for the duration t_z after playing out all the buffered data before 60 GHz radio recovers (shown in Fig. 3(d)).

4.4. Calculation of the utility degradation

Based on the above description of the state transitions, the utility functions can be derived here. The probability of each consequent state is given by Eq. (3).

$$\begin{cases} P_{11} = \Pr(t_{buf} \geq t_{swt}) \\ P_{12} = \Pr(t_{buf} < t_{swt}) \\ P_{21} = \Pr(t_{buf} \geq t_{blk}) \\ P_{22} = \Pr(t_{buf} < t_{blk}) \end{cases} \quad \text{with } \sum_j P_{ij} = 1. \quad (3)$$

Referring to the relation between different times as shown in Fig. 3, the utility degradation of each possible consequent state is formulated in (4). Here, t_{blk} can be either the expectation $E(t_{blk})$ or the predicted value of the blocking duration,

$$\begin{cases} \Delta U(S_{11}) = t_{WLAN}(u_{60} - u_w) = (t_{blk} + t'_{swt} + t'_{buf} - t_{buf})(u_{60} - u_w), \\ \Delta U(S_{12}) = t_z(u_{60} - u_z) + t_{WLAN}(u_{60} - u_w) \\ \quad = (t_{swt} - t_{buf})(u_{60} - u_z) + (t_{blk} + t'_{swt} + t'_{buf} - t_{swt})(u_{60} - u_w), \\ \Delta U(S_{21}) = 0, \\ \Delta U(S_{22}) = t_z(u_{60} - u_z) = (t_{blk} - t_{buf})(u_{60} - u_z). \end{cases} \quad (4)$$

For a_1 and a_2 the expected utility degradation can thus be formulated as,

$$\begin{cases} \Delta U_{a_1} = P_{11}\Delta U(S_{11}) + P_{12}\Delta U(S_{12}), \\ \Delta U_{a_2} = P_{21}\Delta U(S_{21}) + P_{22}\Delta U(S_{22}). \end{cases} \quad (5)$$

By comparing ΔU_{a_1} and ΔU_{a_2} the action corresponding to a smaller value can be chosen.

5. Modeling using Markov decision process

As an extension of the decision theory approach, we have further modeled the handover decision problem using Markov Decision Process (MDP) taking into account the further influence of the actions. For example, in the decision theory approach, it seems pointless to stay using WLAN radio as soon as 60 GHz radio system is recovered, since 60 GHz radio generates higher immediate utility. However, by

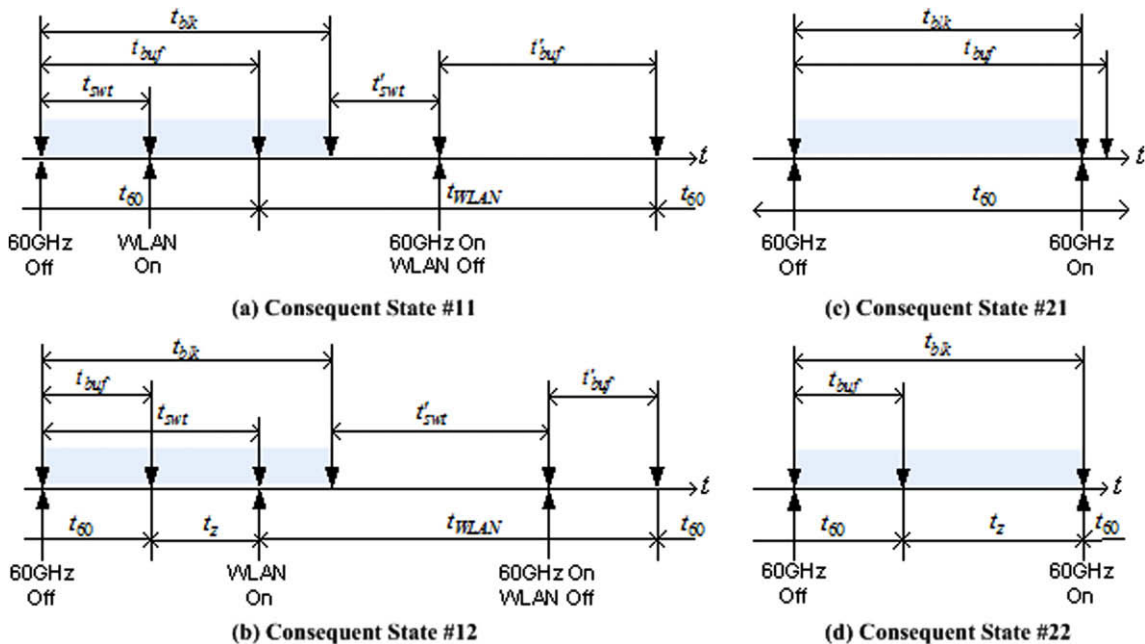


Fig. 3. Time line of the consequent states.

using MDP, we would also take into account whether the recovered 60 GHz link lasts only for a short duration. If so, staying with WLAN could be a wiser choice, as it avoids unnecessary handover back and forth in a short period.

As defined in [12], five key aspects of MDP model are: (1) a set of decision epochs, (2) a set of system states, (3) a set of available actions, (4) a set of states and actions dependent on immediate rewards or costs, and (5) a set of states and action dependent transition probabilities. We explain these in the sequel.

5.1. Decision epochs

The decision epochs are taken every time the network situation changes. Two alternatively occurring events are associated with the decision epochs – the event that the 60 GHz radio LOS link is lost, and the event that the link is available. We denote the events as a vector E of two binary values,

$$E : \{e_i\} = \begin{cases} e_1, & \text{if a 60 GHz LOS link blocking starts,} \\ e_2, & \text{if a 60 GHz LOS link blocking ends.} \end{cases}$$

The interval τ_k between two consecutive epochs say, k th and $(k+1)$ th decision epochs is the duration of an event, either a blocking or an unblocking has happened at the k th decision epoch and it follows,

$$E(\tau_k) = \begin{cases} E(t_{blk}) = \mu_1, & \text{when } E_k = e_1, \\ E(t_{nblk}) = \mu_2, & \text{when } E_k = e_2. \end{cases}$$

5.2. System states

The system states at each decision epoch are represented as a multi-dimensional vector. It includes the observable states of the environment, the network and the mobile device, all of which affect the decision-making at a given decision epoch. The state space S is represented as a three-dimensional vector, $S = E \times N \times T$.

Here, E is the event on the decision epoch as defined in the previous section. N stands for the radio interface that the mobile device uses at the decision epoch.

$$N : \{n_i\} = \begin{cases} n_1, & \text{the device is using the 60 GHz radio interface,} \\ n_2, & \text{the device is using the WLAN radio interface.} \end{cases}$$

T represents collectively switching duration and buffering duration, $T = (t_{swt}, t_{buf})$. As the durations are, in principle, continuous variables, t_{swt} and t_{buf} are discretized by a step size Δ to make the state space countable. At the k th decision epoch, the system state is,

$$X_k = \{E_k; N_k; t_{swt}(k), t_{buf}(k)\} \in S.$$

5.3. Actions

In MDP model, the actions are a_1 : perform handover to the other radio system, or a_2 : do not perform handover but stay using the current radio system.

5.4. Transition probabilities

The transition probabilities from the state at the k th decision epoch to the state at the $(k+1)$ th decision epoch can be considered as the joint probability of the system state transition at each dimension under the action at the k th decision epoch. That is,

$$\Pr(X_{k+1}|X_k, a_k) = \Pr(E_{k+1}, N_{k+1}, t_{swt}(k+1), t_{buf}(k+1)|E_k, N_k, t_{swt}(k), t_{buf}(k), a_k). \quad (6)$$

Given current state, we consider that the random variables of next state are independent of each other, but only depend on current state and the action chosen. Thus we can transform (6) into,

$$\Pr(X_{k+1}|X_k, a_k) = \Pr(E_{k+1}|X_k, a_k)\Pr(N_{k+1}|X_k, a_k)\Pr(t_{swt}(k+1), t_{buf}(k+1)|X_k, a_k). \quad (7)$$

The blocking and the unblocking events always occur alternatively and are independent of the actions. Thus we have,

$$\Pr(E_{k+1}|X_k, a_k) = \begin{cases} 1, & \text{if } E_k \neq E_{k+1}, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Whether or not the mobile device continues to use the same radio interface depends on both action taken and system state with respect to time. Thus we have,

$$\Pr(N_{k+1}|X_k, a_k) = \begin{cases} \Pr(t_{swt}(k) > \tau_k), & \text{for } (N_k = N_{k+1}) \cap (a_k = a_1), \\ \Pr(t_{swt}(k) \leq \tau_k), & \text{for } (N_k \neq N_{k+1}) \cap (a_k = a_1), \\ 1, & \text{for } (N_k = N_{k+1}) \cap (a_k = a_2), \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

If we consider switching and buffering time are mutually independent random variables depending only on network and device and not influenced by actions, then,

$$\Pr(t_{swt}(k+1), t_{buf}(k+1)|X_k, a_k) = \Pr(t_{swt}(k+1)|t_{swt}(k))\Pr(t_{buf}(k+1)|t_{buf}(k)). \quad (10)$$

Further if we assume both $t_{swt}(k)$ and $t_{buf}(k)$ are uniformly distributed within the range of $[t_{swt_min}, t_{swt_max}]$ and $[t_{buf_min}, t_{buf_max}]$, respectively, then we have,

$$\gamma = \Pr(t_{swt}(k+1), t_{buf}(k+1)|t_{swt}(k), t_{buf}(k), a_k) = \frac{1}{(t_{swt_max} - t_{swt_min})(t_{buf_max} - t_{buf_min})}. \quad (11)$$

Based on 9, 10 and 11, 8 can be further written into,

$$\Pr(X_{k+1}|X_k, a_k) = \begin{cases} \gamma\Pr(N_{k+1}|X_k, a_k), & \text{if } E_{k+1} \neq E_k, \\ 0, & \text{if } E_{k+1} = E_k. \end{cases} \quad (12)$$

5.5. Rewards

At the k th decision epoch, given the current state X_k and the chosen action a_k , the reward function is expressed as,

$$R(X_k, a_k) = \sum_{X_{k+1} \in S} \Pr(X_{k+1}|X_k, a_k)r(X_{k+1}|X_k, a_k). \quad (13)$$

Here, $r(X_{k+1}|X_k, a_k)$ is the reward when the system being in state X_k , taking an action a_k while going to the next state X_{k+1} . We still use reward density to quantify the reward as defined in Section 3.1 to take into account the users' perceived quality between two decision epochs. We also introduce the switching cost [17], K_s , which reflects the impact of re-routing operation, signaling load, extra battery consumption, etc. We define,

$$K_s(a_k) = \begin{cases} K, & K > 0, \text{ if } a_k = a_1, \\ 0, & \text{otherwise.} \end{cases}$$

In general the reward is the sum of multiplication of different classes of reward density with the corresponding duration within the interval of two successive decision epochs, formulated as,

$$\begin{cases} r(X_{k+1}|X_k, a_k) = \mu_{60}t_{60}(k) + \mu_w t_w(k) + \mu_z t_z(k) - K_s(a_k), \\ t_{60}(k) + t_w(k) + t_z(k) = \tau_k. \end{cases} \quad (14)$$

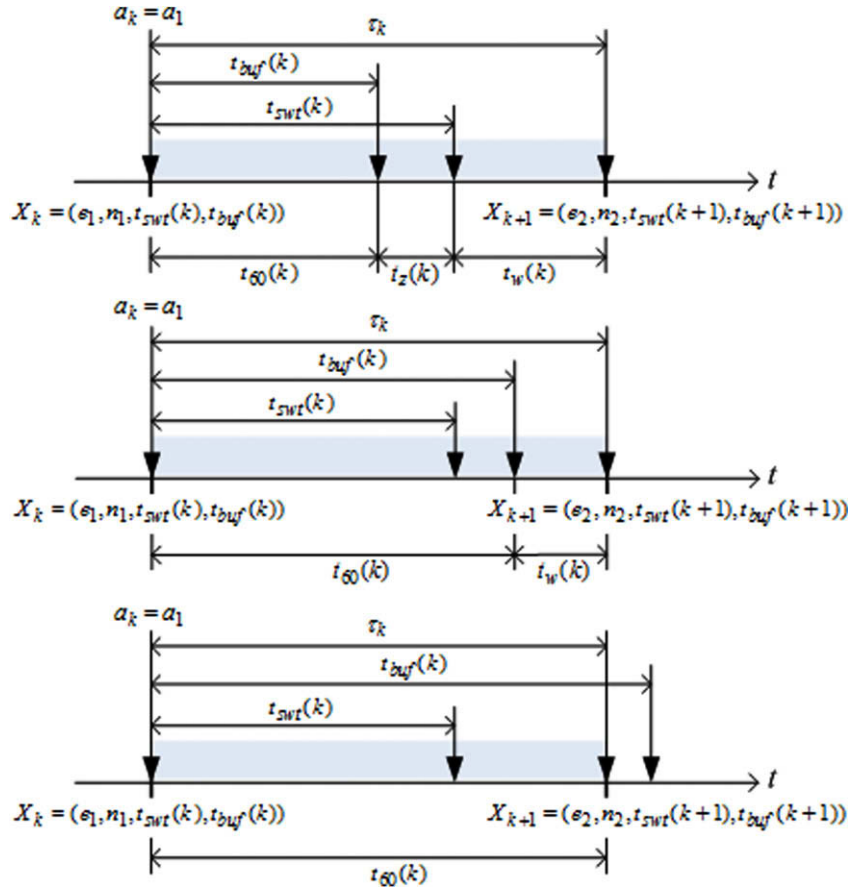


Fig. 4. The reward function example.

$$r(e_2, n_{k+1}, T(k+1)|e_1, n_1, T(k), a_k) = \begin{cases} \mu_{60} \min(\tau_k, t_{buf}(k)) + \mu_z \max(0, \tau_k - t_{buf}(k)) - K, & \text{for } n_{k+1} = n_k, a_k = a_1, \\ \mu_{60} \min(\tau_k, t_{buf}(k)) + \mu_w \min(\max(0, \tau_k - t_{buf}(k)), \tau_k - t_{swt}(k)) \\ + \mu_z \max(0, t_{swt}(k) - t_{buf}(k)) - K, & \text{for } n_{k+1} \neq n_k, a_k = a_1, \\ \mu_{60} \min(\tau_k, t_{buf}(k)) + \mu_z \max(0, \tau_k - t_{buf}(k)), & \text{for } n_{k+1} = n_k, a_k = a_2. \end{cases} \quad (15a)$$

$$r(e_2, n_{k+1}, T(k+1)|e_1, n_2, T(k), a_k) = \begin{cases} \mu_w \tau_k - K, & \text{for } n_{k+1} = n_k, a_k = a_1, \\ \mu_w \min(t_{buf}(k) + t_{swt}(k), \tau_k) + \mu_0 \max(0, \tau_k - t_{swt}(k) - t_{buf}(k)) - K, & \text{for } n_{k+1} \neq n_k, a_k = a_1, \\ \mu_w \tau_k, & \text{for } n_{k+1} = n_k, a_k = a_2. \end{cases} \quad (15b)$$

$$r(e_1, n_{k+1}, T(k+1)|e_2, n_1, T(k), a_k) = \begin{cases} \mu_{60} \tau_k - K, & \text{for } n_{k+1} = n_k, a_k = a_1, \\ \mu_{60} \min(t_{buf}(k) + t_{swt}(k), \tau_k) + \mu_w \max(0, \tau_k - t_{swt}(k) - t_{buf}(k)) - K, & \text{for } n_{k+1} \neq n_k, a_k = a_1, \\ \mu_{60} \tau_k, & \text{for } n_{k+1} = n_k, a_k = a_2. \end{cases} \quad (15c)$$

$$r(e_1, n_{k+1}, T(k+1)|e_2, n_2, T(k), a_k) = \begin{cases} \mu_w \tau_k - K, & \text{for } n_{k+1} = n_k, a_k = a_1, \\ \mu_{60} \max(0, \tau_k - t_{swt}(k) - t_{buf}(k)) + \mu_w \min(t_{buf}(k) + t_{swt}(k), \tau_k) - K, & \text{for } n_{k+1} \neq n_k, a_k = a_1, \\ \mu_w \tau_k, & \text{for } n_{k+1} = n_k, a_k = a_2. \end{cases} \quad (15d)$$

Fig. 4 illustrates an example that, under the “handover” action the current state X_k transits to the next state X_{k+1} . Accordingly reward in all cases can be formulated as (15a)–(15d),

5.6. Finding an optimal policy

A policy of MDP model specifies the action that should be taken under any given state. In the optimal policy π^* , the action corresponding to each state $s \in S$ is the one that maximizes the “long-term” total reward from s onwards,

$$\begin{cases} \pi^*(s) = \arg \max_a \sum_{s' \in S} \Pr(s'|s, a)U(s'), \\ U(s) = R(s) + \gamma \max_a \sum_{s' \in S} \Pr^-(s'|s, a)U(s'). \end{cases} \quad (16)$$

Here, γ is the discount factor representing far future consequent states that are less important for decision-making than the near future ones. Taking (11) as the value of the discount factor, we have in (16) $\Pr^-(s'|s, a) = \Pr(N'|s, a)$.

To find the optimal policy, we applied the value iteration algorithm [15]. The basic idea is to use the iterative approach to update the rewards of all the states until reaching equilibrium.

6. Simulation setup, results and discussions

The proposed handover decision algorithms have been implemented in Matlab. The 60 GHz LOS link blockage is modeled as an independent blocking event with duration being exponentially distributed with rate μ . Similarly LOS link non-blocking duration is modeled using λ , as shown in Fig. 5. In simulations, different $E(t_{blk})$ and $E(t_{nblk})$ are used to generate different types of blocking situations. When $E(t_{blk})$ is small, e.g., around 2 s, the blocking is considered mainly due to people walking through the 60 GHz LOS link with moderate speeds. A higher $E(t_{blk})$ indicates that the blocking is caused by other factors, such as people standing for longer duration obstructing the LOS link or devices being placed behind an obstacle, etc.

Buffering time is dependent on the data rate and the maximum size of the buffer available on the device. As the link quality changes from time to time, we took buffering time, t_{buf} and t'_{buf} randomly between t_{buf_min} and t_{buf_max} . We set t_{buf_min} to zero to represent that no data are buffered, and t_{buf_max} corresponds to maximum play-out duration with full buffer. Switching time consists of both network configuration and session handshake, which are affected by dynamic network situations. We set t_{swt} and t'_{swt} randomly between t_{swt_min} and t_{swt_max} . As 60 GHz radio has been proposed to use beacon based device discovery scheme [14] and so does WLAN, it leads to a short network association time in most of the cases. In [13], the measurement showed that the SIP based session handshake in IPv6 can be below 500 ms with certain modifications. Therefore, it is reasonable to assume t_{swt_min} being around 0.5 s under favorable network conditions. We varied boundaries of buffering time and switching time and examined our decision algorithms under different possible scenarios.

For the user input, we assigned a set of values of 10, 5 and 0 for u_{60} , u_w and u_z , respectively, and the switching cost K_s is fixed at 2

units. The details of the parameters used in the following simulations are listed in Table 1.

6.1. Performance of DT approach

First to get a general view of the performance of the DT approach, we compared it (we use the notation “DT”) with other three naive decision algorithms: (1) “RND” – randomly choose some action, (2) “SWT” – always handover and (3) “WAIT” – do not handover to WLAN when 60 GHz LOS link is blocked. We consider here a benchmark algorithm that knows precisely each blocking duration and is able to choose the action resulting in lowest utility degradation. The performance metric we consider here is the probability of choosing an optimal action, found by the benchmark algorithm, by the three algorithms. In the simulation, the

Table 1
Parameter setting.

Parameter	Value	Notes
u_{60}	10	Utility density using 60 GHz streaming
u_w	5	Utility density using WLAN streaming
u_z	0	Utility density no streaming
K_s	2	Switching cost
$E(t_{nblk})$	30 s	Mean non-blocking duration (DT)
	5:5:60 (s)	Mean non-blocking duration (MDP)
$E(t_{blk})$	1:1:10 (s)	Mean blocking duration (DT)
	0.5:0.5:10 (s)	Mean blocking duration (MDP)
t_{swt_min}	0.5 s	Minimum switching time
t_{swt_max}	2 s	Maximum switching time
t_{buf_min}	0	Minimum buffer time
t_{buf_max}	5 s	Maximum buffer time
$conf$	90%	Prediction confidence level
err	10–90%	Prediction error range
Δ	0.5 s	Step size (MDP only)

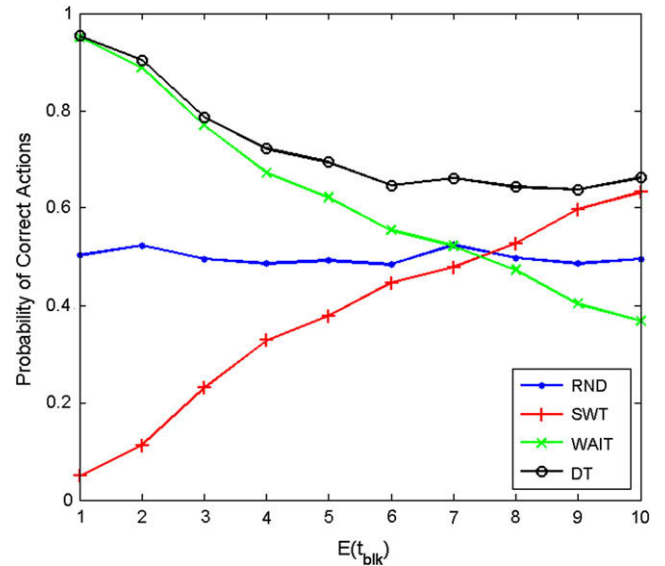


Fig. 6. Performance of DT using mean.

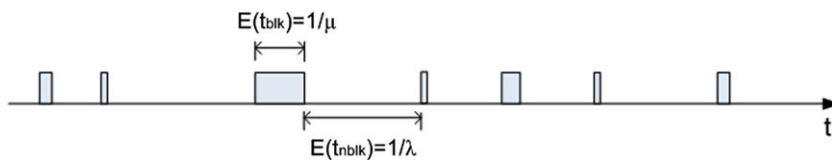


Fig. 5. Blocking event model.

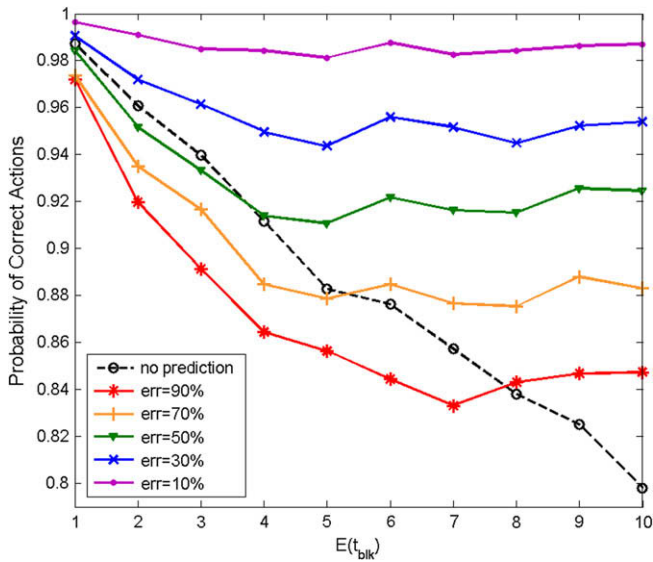


Fig. 7. Performance of DT using predictions.

switching time and the buffering time are chosen randomly between 0.5–2 s and 0–5 s, respectively.

Shown in Fig. 6, the “wait” action is more appropriate than the “handover” when $E(t_{blk})$ is small, whereas it is better to handover when $E(t_{blk})$ is large. Randomly choosing the actions results in around 50% optimal actions for all $E(t_{blk})$ cases as expected. The DT approach yields in each case the best result. However, its performance is worse under some $E(t_{blk})$ values, for example from 5 s to 9 s. The reason is that the decision-making is based on a comparison between ΔU_{a_1} and ΔU_{a_2} referring to (5). Since random $E(t_{blk})$ values are used in place of real durations, it is unable to represent by varying the length of blocking events. For certain switching time and buffering time, the individual blocking duration is specifically critical to determine the outcome of comparison of ΔU_{a_1} and ΔU_{a_2} . Using $E(t_{blk})$ alone is too generic to derive precise decisions in those cases.

Therefore, we further consider predicted length of each blocking event instead of just $E(t_{blk})$ to reflect the dynamic nature of the blocking duration. The prediction is manipulated from the generated blocking durations. Taking into account the errors in prediction, we added errors onto those blocking durations. With a

certain confidence level, *conf*, the error of each predicted value is restricted within the range *err* percent of its corresponding blocking duration. We tested the DT approach by varying *err* from 10% to 90% in steps of 20%, under 90% *conf*. We also compared the performance of using $E(t_{blk})$ at the same time. From Fig. 7 with the increasing prediction error, the flawed information leads to the incorrect decision in all blocking cases. But compared to using only $E(t_{blk})$, using predicated values results in lower performance degradation when $E(t_{blk})$ is large. This is due to the fact that large $E(t_{blk})$ causes higher errors in individual durations. In some cases the difference between the mean and the individual values can sometimes be larger than 90% of the value. The result suggests that, with 90% confidence level the prediction with an error of less than 50% of the true value can be obtained. It is better to use prediction, rather than the mean, to make a decision.

6.2. Performance of MDP approach

To evaluate the MDP approach, we first used the mean blocking and non-blocking duration $E(t_{blk})$ and $E(t_{nblk})$, and compared its performance with two algorithms. They are, (a) “RND” – randomly choose between handover or wait; and (b) “GRD” – a greedy algorithm which always uses 60 GHz radio whenever the LOS link is available and switches to WLAN otherwise. In the simulation, a step size of 0.5 s is set to empirically leverage the state space size and the accuracy of MDP. A larger step size leads to a smaller state space thus needing lesser computational resources, however, with a smaller step size finer resolution can be obtained.

Fig. 8 shows two different views of the results when $E(t_{blk})$ and $E(t_{nblk})$ are varied. For all three algorithms, decreasing $E(t_{blk})$ or increasing $E(t_{nblk})$ raises the average reward density, which is defined as the ratio of total reward to the time duration in a single simulation round. This is reasonable since the shorter and less-frequent blockings create more chances for the mobile device to use the high quality 60 GHz radio system. In general, MDP gives the best result as its performance plane is mostly on top which is clearly shown in Fig. 8a. An exception, where MDP and GRD planes cross happens when $E(t_{nblk})$ is below 5 s (see Fig. 8b) and $E(t_{blk})$ is larger than 3 s (see Fig. 8a). This indicates an extremely turbulent environment, where the blocking events last long and occurs quite often. In these cases, GRD algorithm gives about 10% better result than MDP. When $E(t_{blk})$ takes large values MDP and GRD perform similarly.

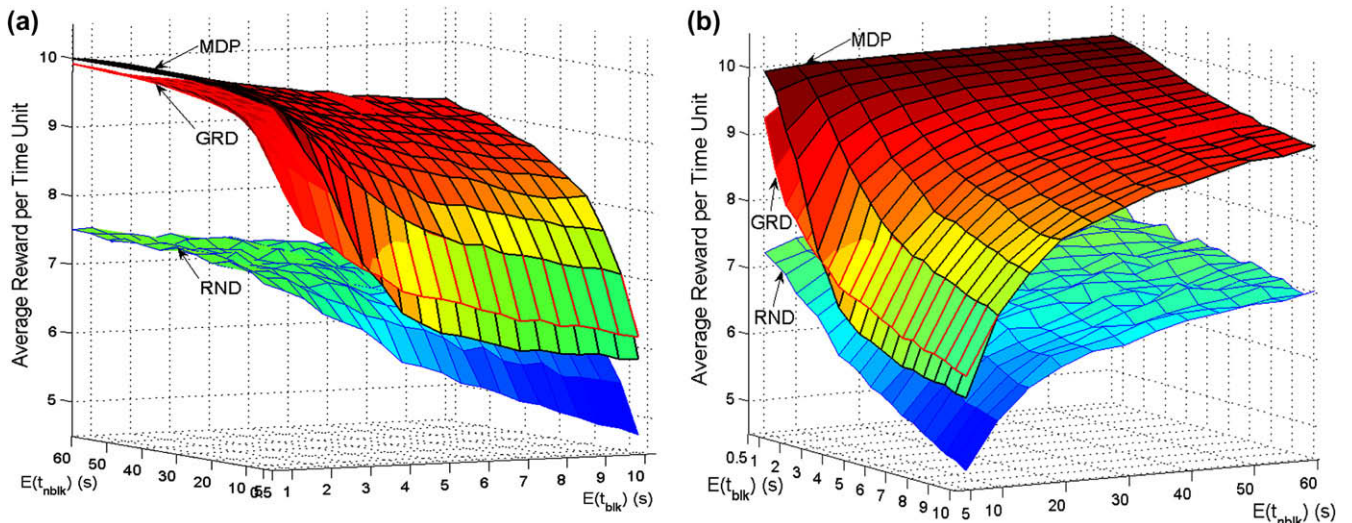


Fig. 8. Performance of MDP using mean blocking durations.

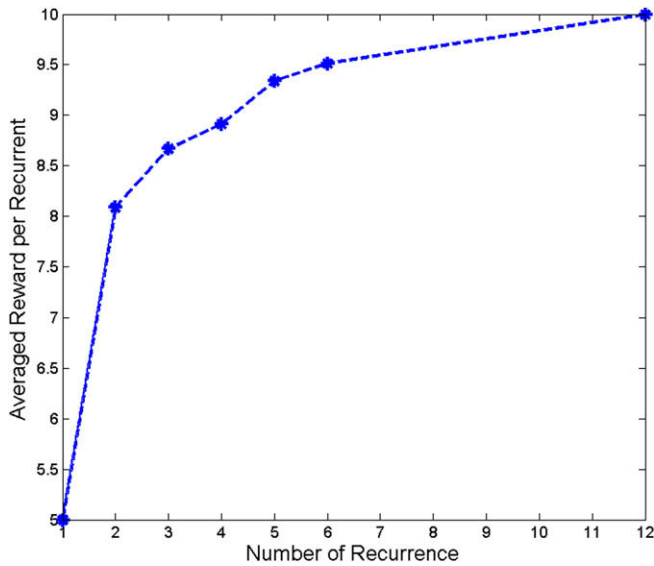


Fig. 9. Reoccurring states in MDP.

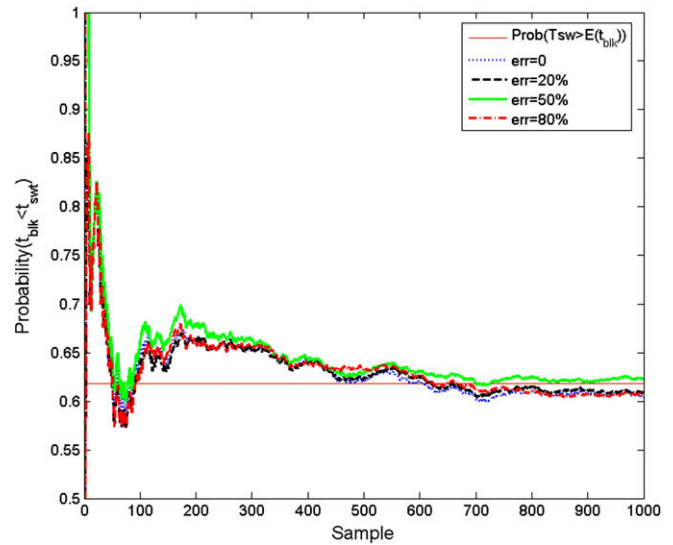


Fig. 11. An example of the estimation process.

Further we tested MDP with predicted $E(t_{blk})$ and $E(t_{nblk})$ instead of using their mean values. We took a simple learning process to derive the expectation and the state transition probability by recording the previously observed blocking events. In (16), the reward of each state consists of two parts, the short-run current state reward $R(s)$ and the long-run averaged state reward $\sum_s \Pr^-(s'|s, a)U(s')$. As the prediction of the next event duration can only contribute to the first part (as DT does), we need to estimate $E(t_{blk})$, $E(t_{nblk})$ and the state transition probability to calculate the long-run part. The learning process used all the previously observed blocking events to perform an estimation, which was updated after every new event and the policy was recalculated accordingly.

First we tracked the reward for the recurrent states. The recurrent state is the state being visited more than once during one simulation round, i.e., the state vector $\{E; N; t_{swt}, t_{buf}\}$ has been observed with the same value more than once. We found over 88% recurrent states have increasing rewards by using more observations of the previous occurrence. We selected those recurrent states with the time averaged reward of 10 under the optimal action, that is, the states where

the user can always observe high quality streaming content under the optimal action, and averaged the reward from all the states with the same recurrence round. When $E(t_{blk})$, $E(t_{nblk})$ and err set to 2 s, 5 s and 20% respectively, Fig. 9 shows the most frequent state occurred 12 times in a simulation round with 2000 states. The average reward of the recurrent states increases with time and reaches the optimal value at the end. This indicates the gain from the learning process by updating estimations with more observations – it is able to give a better result with the experience to deal with the same state occurred before.

We changed the value of err to represent the case with no prediction error, and the cases with prediction error uniformly distributed with 20%, 50% and 80% of the true value. The performance was stable regardless of the difference in the prediction accuracy as shown in Fig. 10. The reason for this immunity to the predicted error can be explained by Fig. 11 with an example that shows the estimation processes for the probability that the switching time is longer than the blocking duration. The error is counteracted with the increasing number of samples, leading to the estimation results with minor difference.

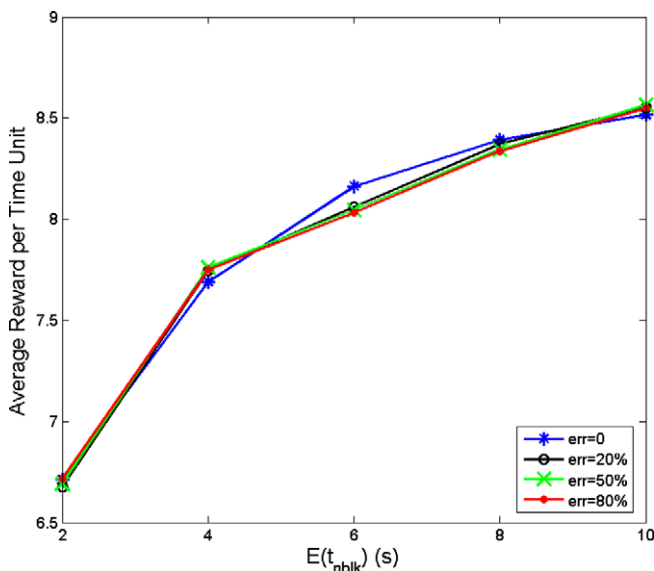


Fig. 10. MDP with different prediction accuracy.

6.3. Comparison of DT and MDP approaches

First we set $E(t_{blk})$ to be 2 s and t_{buf_max} 5 s with varying $E(t_{nblk})$. In Fig. 12, MDP outperforms DT when $E(t_{blk})$ is small, but after 10 s the performance of the two approaches becomes very similar. The reason is, when the blocking interval gets longer the actions chosen by DT and MDP tend to be the same – that is handover to 60 GHz when LOS link is available and this state is likely to remain for a longer period. When $E(t_{nblk})$ is expected to be short, the MDP shows more obvious advantages by making more rational decisions by considering the future consequences. This also reflects that the performance of the approach is influenced by values of the parameters used in considered scenarios.

When we tested both approaches with the equal $E(t_{blk})$ and $E(t_{nblk})$ but with the different buffer capacity, Fig. 13 shows that MDP outperforms DT within a certain range of $E(t_{blk})$ and $E(t_{nblk})$ – when they take relatively small values. The two algorithms have the same performance when $E(t_{blk})$ and $E(t_{nblk})$ are large. With the increasing buffer duration, the range in which MDP outperforms DT becomes larger. This implies that MDP can handle more complicated situations in a better way, where $E(t_{blk})$ and $E(t_{nblk})$ are comparable to

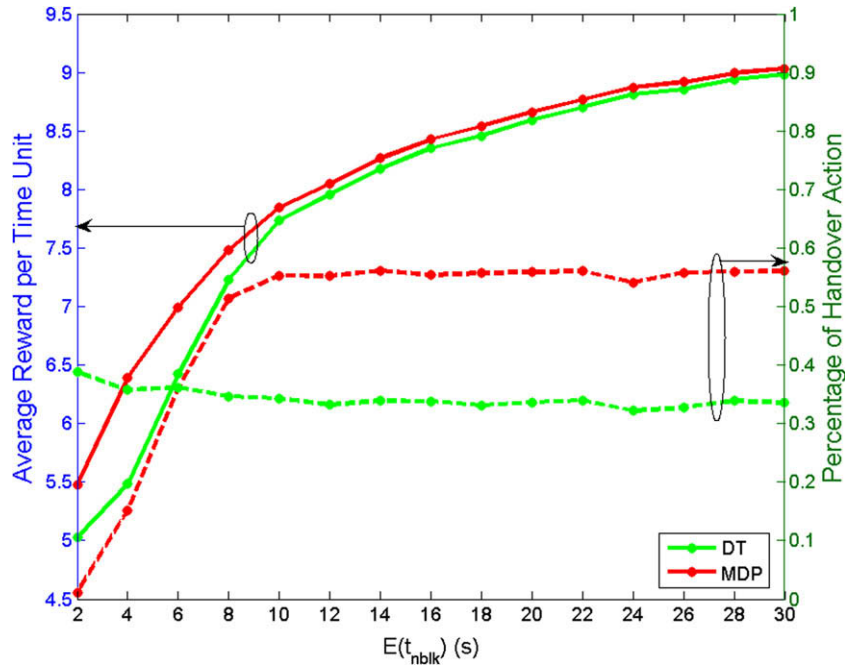


Fig. 12. Performance comparison of DT and MDP (solid line: average reward, dash line: percentage of handover action).

the maximal buffering time, that is, at each decision epoch the state is highly unpredictable and with more uncertain consequences.

The percentage of handover actions chosen by both approaches are shown in Figs. 12 and 14. In the cases where $E(t_{nblk})$ is relatively short and the buffer size is comparatively larger, the percentage of the handover action decided by MDP is lower than that of DT. This is due to the fact that DT switches back to use 60 GHz radio every time the link recovers, whereas MDP does not if the link recovery is likely to be short. It reflects that MDP is capable of taking more considerate decisions by filtering out the impact of trivial environmental changes.

Figs. 15 and 16 show the comparison of using MDP and DT approaches using the same prediction methods. We set $E(t_{blk})$ and $E(t_{nblk})$ of 2 s and 5 s, respectively with err being 20%. First the prediction error is uniformly distributed within the error range. In

Fig. 15, we divided every 2000 simulated events into eight groups according to the order of occurrence. We counted the number of the decisions made by MDP, which gave no-less and higher rewards than those made by DT approach. More than 70% actions under MDP gave no-less rewards than that of DT, and 10% to 20% actions resulted in higher rewards. It also shows that with more events MDP is highly beneficial. Especially after 500 events MDP gave a relatively stable performance with over 80% actions resulting in no lesser reward and over 20% giving higher rewards. This is the result of the estimation process, which has been observed to converge to actual value after 500 samples and thus it leads to a stable performance afterwards. Fig. 16 shows that in general MDP outperforms DT under different cases of blocking intervals. With the same reason to the result using $E(t_{blk})$, the benefit is more obvious when the length of blocking duration and interval are closer.

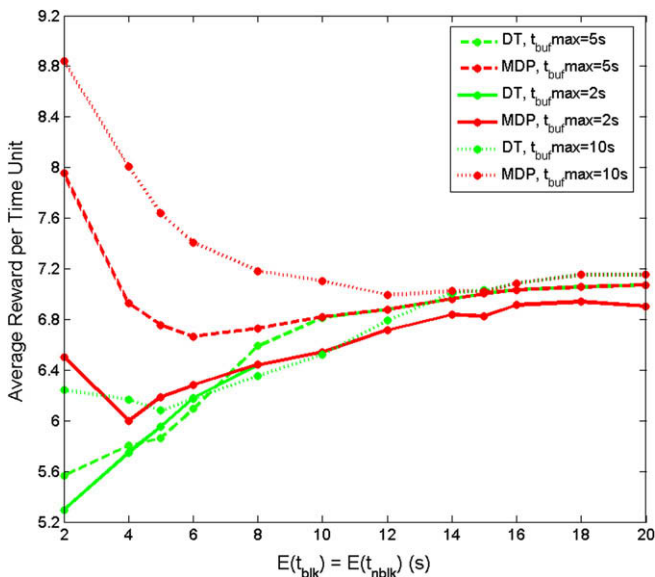


Fig. 13. Averaged reward comparison.

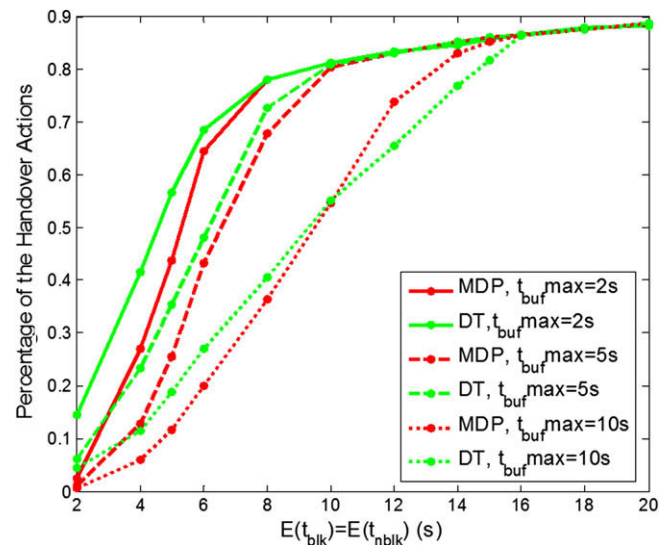


Fig. 14. Action percentage comparison.

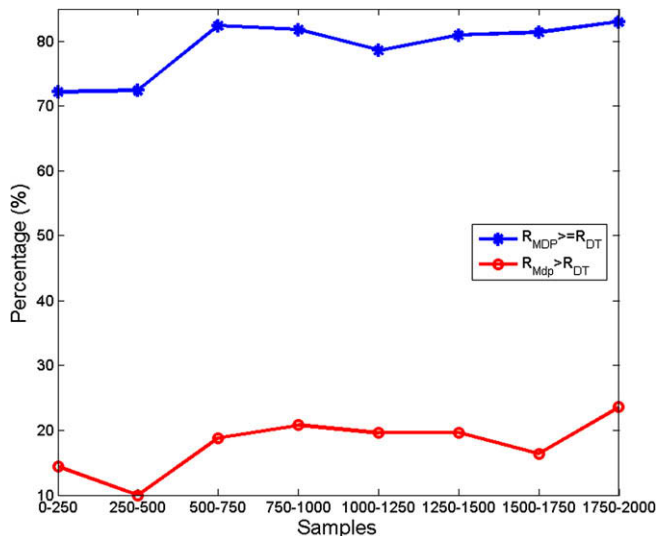


Fig. 15. Percentage of better decisions.

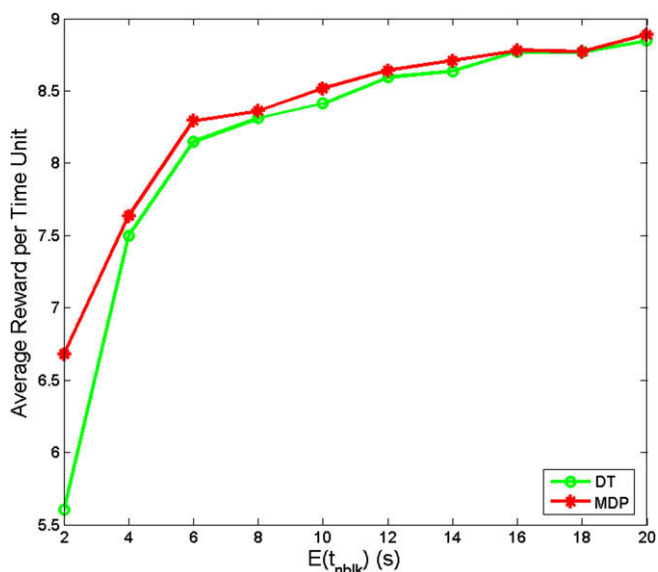


Fig. 16. DT versus MDP under 20% error.

An important note here is the trade-off between performance and complexity of the algorithms. The formulated DT algorithm grows as $O(|S|)$, considering the decision-making only happens when 60 GHz LOS blocked, and for each action it results in two consequent states. Whereas, for MDP as we used value iteration algorithm, it grows as $O(|S|^2|A|)$ for each iteration. Further if we consider including more radio systems, as the number of interfaces linearly increase, the number of states and actions for both the algorithms, it leads to a linear increase in complexity of DT, but a cubic growth rate for MDP.

7. Conclusion

In this paper, we presented two promising approaches to make vertical handover decisions for, but not restricted to, the in-home network with co-existing 60 GHz radio and WLAN. The two decision algorithms we proposed are based on the Decision Theory (DT) and

Markov Decision Process (MDP), both of which take into account the multiple factors from users, environment, network and devices. The simulation results show that these sophisticated algorithms give a better performance on deciding the correct actions in the partially observable and non-deterministic environment compared to other simple decision-making strategies. As expected the MDP approach outperforms the DT approach under many configurations. Furthermore, the results also indicate the specific situations in which the effectiveness of the algorithms is limited. While we claim that the above approaches can result in better decision making, we are aware that we need to derive and understand the limitations of these approaches through analytical means. In real situations, we may not be able to decouple all the parameters used here. For example, the buffering duration is related to the interval of the previous blocking event and the switching duration. If the switching has happened in the very recent past buffer may not be full. The utility can be affected by the screen size of the device. Thus, the next step is to explore the factors affecting the decision-making with more realistic and more inter-connected models.

References

- [1] P.F.M. Smulders, 60 GHz radio: prospects and future directions, in: Proceedings of the IEEE Benelux Symposium Chapter on Communications and Vehicular Technology, Eindhoven, The Netherlands, 2003.
- [2] H. Yang, M.H.A.J. Herben, P.F.M. Smulders, Frequency selectivity of 60-GHz LOS and NLOS indoor radio channels, in: Proceedings of the IEEE Vehicular Technology Conference, Melbourne, Australia, May 2006.
- [3] B.L. Dang, V. Prasad, I. Niemegeers, M.G. Larrode, A.M.J. Koonen, Toward a seamless communication architecture for in-building networks at the 60 GHz band, in: Proceedings of the 31st IEEE Conference on Local Computer Networks (LCN2006), Florida, USA, November, 2006.
- [4] V.A. de Sousa, R.A. de O. Neto, F. de S. Chaves, L.S. Cardoso, F.R.P. Cavalcanti, Access selection with anticipatory vertical handover for multi-access networks, in: Proceedings of the IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, Helsinki, Finland, September 2006.
- [5] S. Horrich, S.B. Jamaa, P. Godlewski, Adaptive vertical mobility decision in heterogeneous networks, in: Proceedings of Third International Conference on Wireless and Mobile Communications (ICWMC'07), Guadeloupe, French Caribbean, 2007.
- [6] H. Liao, L. Tie, Z. Du, A vertical handover decision algorithm based on fuzzy control theory, in: Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCS'06), April 2006.
- [7] T. Ahmed, K. Kyamakya, M. Ludwig, Design and implementation of a context-aware decision algorithm for heterogeneous networks, in: Proceedings of the 21st Annual ACM Symposium on Applied Computing, Dijon, France, April 2006.
- [8] E. Stevens-Navarro, V.W.S. Wong, Y. Lin, A vertical handoff decision algorithm for heterogeneous wireless networks, in: Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC'07), Hong Kong, China, March 2007.
- [9] M. Jacobsson, I. Niemegeers, Privacy and anonymity in personal networks, in: Proceedings of the IEEE PerSec, 2005.
- [10] ITU-T Recommendation P.800, Methods for subjective determination of transmission quality, 1996.
- [11] S. Parsons et al. (Eds.), Game Theory and Decision Theory in Agent-based System, Kluwer Academic Publishers, Dordrecht, 2002.
- [12] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley-Interscience, New York, 1994.
- [13] N. Nakajima, A. Dutta, S. Das, H. Schulzrinne, Handoff delay analysis and measurement for SIP based mobility in IPv6, in: Proceedings of the IEEE International Conference on Communications (ICC'03), Alaska US, May 2003.
- [14] X. An, R. Hekmat, Directional MAC protocol for millimeter wave based wireless personal area networks, in: Proceedings of the IEEE 67th Vehicular Technology Conference: VTC2008-Spring, Singapore, May 2008.
- [15] Dimitri P. Bertsekas, A new value iteration method for the average cost dynamic programming problem, SIAM Journal on Control and Optimization 36 (2) (1998) 742–759.
- [16] N. Moraitis, P. Constantinou, Indoor channel measurements and characterization at 60 GHz for wireless local area network applications, IEEE Transactions on Antennas and Propagation 52 (12) (2004) 3180–3189.
- [17] E. Stevens-Navarro, Y. Lin, V.W.S. Wong, An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks, IEEE Transactions on Vehicular Technology 57 (2) (2008) 1243–1254.