Performance Analysis of Fast Initial Link Setup for IEEE 802.11ai WLANs

Eng Hwee Ong
Nokia Research Center
Itämerenkatu 9, FIN-00180 Helsinki, Finland
Email: eng-hwee.ong@nokia.com

Abstract—The IEEE 802.11ai is an upcoming fast initial link setup (FILS) amendment that could enable a STA to achieve secure link setup in less than 100 ms. A successful link setup process will then allow the STA to send IP traffic with a valid IP address through the AP. In this paper, we first present a performance analysis on how well the legacy 802.11 can support FILS. We then demonstrate that the rate at which the medium saturates is mainly dependent on the offered load which has strong dependencies on the active scan rates (i.e., amount of transmitted probe request frames) and number of responding APs (i.e., amount of transmitted probe response frames) among others. Moreover, the average active scanning duration per channel could be maintained below 5 ms provided the medium is not saturated. Our results also indicate that a combination of the enhanced distributed channel access (EDCA) procedure and the proposed active scanning enhancements provides at least 20% and up to 250% improvements in supporting the functional requirements of FILS as compared to the 802.11 EDCA and legacy distributed coordination function (DCF), respectively.

I. INTRODUCTION

The continuous growth of 802.11-enabled devices such as smartphones, tablets, netbooks, televisions, and portable music players are expected to increase significantly. The Wi-Fi Alliance expects the number of 802.11-enabled devices to surpass 2 billion by 2015 [1]. On the other hand, the high (Gbps) data rate offered by the new 802.11ac WLAN [2] is a key driver for dual mode functionality in mobile devices. In fact, there have been growing trends for network operators to supplement the capacity of their cellular network with 802.11-based connections as the preferred data offloading solution.

In this respect, the notion of FILS becomes important for end-users to fully exploit the high data rate capability of 802.11 WLAN seamlessly so that the benefits of high data rate can be reaped even when offered over a short period of time. For network operators, it reduces the load on the bandwidth-limited cellular network which will lead to increased bandwidth availability. For end-users, it provides them with new services, such as hotspot walk-through Internet access and drive-by information, on their mobile devices [3]. However, the existing IEEE 802.11 standard [4] is not designed to support the notion of FILS. Therefore, the IEEE P802.11 Task Group AI (TGai) has been established in late 2010 to develop amendments to the IEEE 802.11 standard.

In this paper, we provide a performance analysis on how well the legacy 802.11 DCF, EDCA, and TGai active scanning enhancements can support the functional requirements of FILS. To the best of the author’s knowledge, this is the first: (i) attempt to introduce the draft 802.11ai amendment; and (ii) analysis to consider multiple APs in an extended service set (ESS) environment with a large number of STAs in context of the TGai reference use case.

The remainder of this paper is organized as follows. Section II describes the TGai reference use case, key concepts, and requirements to enable FILS. In addition, two active scanning enhancements that are currently discussed in the TGai are introduced. Section III gives a heterogeneous and finite load analytical model to evaluate the functional requirements of FILS under various conditions. Section IV provides the performance analysis and comparison of the legacy 802.11 DCF, EDCA, and TGai active scanning enhancements. Finally, Section V concludes this paper.

II. TGAI REFERENCE USE CASE, KEY CONCEPTS, REQUIREMENTS, AND PROPOSED ENHANCEMENTS

At the time of writing, the reference use case, key technical concepts, functional requirements, and proposed enhancements of TGai have been defined. The following sections present an overview of these various aspects of TGai.

A. Reference Use Case

One of the pertinent problems that the TGai aims to address comes from an environment where a large number of users are constantly entering and leaving the coverage area of an ESS. Such a problem has been generalized as the Tokyo central station use case as shown in Fig. 1 with the necessary requirements in [3]. Every time a STA enters the ESS, the STA has to perform an initial link setup with an AP to establish WLAN connectivity. This works well with the current IEEE 802.11 standard when the number of scanning STAs in a given time period is small. However, when the number of STAs simultaneously entering the ESS with multiple APs becomes large, efficient mechanisms that scale well with such a large number of users are required. This implies the amount of time the STA spends in the initial link setup needs to be minimized while maintaining secure authentication.

B. Key Concepts

Accordingly, the TGai focuses on reducing the duration of time spent in each of the following four phases that constitute the initial link setup procedure [5] as illustrated in Fig. 2. In
other words, the link setup time is the total time spent in all the four phases. Note that link setup may involve more than one AP in the ESS.

The main idea is that the AP and network discovery phases could be optimized for providing robustness against large number of users by amending current rules for scanning and including more context information so that the STA can minimize the amount of time spent in discovery. E.g., time spent by the STA in the AP discovery phase may be significantly reduced when additional context information could be acquired from external sources such as offline Wi-Fi database, location information in mobile devices, etc. However, the current IEEE 802.11 standard does not provide all the necessary means to exploit such context information. Hence, another key aspect is to mitigate the probability of medium congestion by reducing the amount of probe request and response transmissions during active scanning. The reduction of time spent in the network discovery phase is associated with the improvements in generic advertisement service (GAS) protocol [4] that enables the STA to discover the availability of context information in relation to the desired network services. E.g., one way to reduce signaling and consequently the time spent in the network discovery phase may be achieved by using the notion of broadcast instead of unicast addressing in the GAS protocol.

On the other hand, the authentication & association and higher layer (DHCP/IP) phases could be optimized by concurrency of information exchange to reduce the amount of time taken by the initial link setup. E.g., parallelization of higher layer protocol messages, such as IP address assignment, while setting up the link is highly desirable. This paper focuses on the first aspect which concerns the robustness against large number of users. Hence, we consider only the active scanning procedure and its enhancements for AP discovery.

C. Requirements

In order to evaluate the effectiveness of submitted proposals, the TGai has defined the functional requirements in [6], as well as the relevant metrics and evaluation methodology in [7]. The key metrics used by the TGai to characterize different use cases are link attempt rate, medium load, and link setup time. Accordingly, the link setup time and two scalability requirements, viz., minimum user load with at least 100 STAs and robustness in the presence of at least 50% (high) medium load are referred as the functional requirements of FILS. The TGai has further defined in [7] that the minimum link setup time in an artificial (line-of-sight) environment, consisting of a single STA and a single AP in a randomly selected channel, with a mandatory data rate of 6 Mbps with extended rate PHY (ERP)-OFDM PPDU frame format shall be less than 100 ms when no medium load exists and a finite value in presence of high medium load.

In this paper, we focus on deriving the average active scanning duration in the AP discovery phase and the theoretical limits of achievable packet exchanges, which are critical to fulfill the functional requirements of FILS. By assuming that external knowledge is not available (i.e., scanning STA performs active scanning in all supported channels) and each channel has the same medium conditions, the upper bound of the average active scanning duration in the AP discovery phase in order to meet the link setup time requirement can be expressed as

\[ T_{activeFILS_{max}} = \left( T_{linkSetup} - T_{hLayers} \right) / c \]  

where \( T_{activeFILS_{max}} \) is the maximum active scanning duration a FILS-enabled STA may spend per channel in the AP discovery phase, \( T_{linkSetup} \) is the required link setup time for the TGai reference use case, \( T_{hLayers} \) is time required for link setup procedure in both authentication & association and higher layer (DHCP/IP) phases which is about 40 ms according to [5]. \( c \) is the number of channels to be scanned. Thus, \( T_{activeFILS_{max}} \approx 5 \text{ ms} \) when \( T_{linkSetup} = 100 \text{ ms} \), \( T_{hLayers} = 40 \text{ ms} \), and \( c = 11 \) (2.4 GHz).

On the other hand, the theoretical limit of achievable packet exchanges can be expressed as

\[ UB_{pktXchg} = \max_{X \in R_{active} \cap D(X) \leq T_{activeFILS_{max}}} n(X) \]  

where \( UB_{pktXchg} \) is the upper bound of the total probe request frames transmitted in the medium, which corresponds to a MAC delay \( D(X) \) that is bounded by \( T_{activeFILS_{max}} \) as defined in (1) at a given per-STA active scan rate \( R_{active} \), and \( n \) is the number of scanning STAs. Thus, (2) gives the maximum amount of transmitted probe request frames by all scanning STAs without saturating the medium while meeting the delay requirement. Note that the effect of probe response frame transmissions as a result of the received probe request frames and any other medium load are also considered when determining the prevailing medium state. Other evaluation requirements for security and backward compatibility can also be found in [7].

D. Proposed Enhancements

In order to reduce the time spent in the AP discovery phase, the active scanning procedure can be optimized by reducing the amount probe request and/or response frames in the medium, especially, when there is a large number of
scanning STAs and multiple responding APs. Two proposed active scanning enhancements, which are targeted to reduce the time spent in the AP discovery phase, are as follows:

i. Filter list (FIL): The FIL element may be optionally included in the probe request frame so that a scanning STA has the option to state explicitly whether to stop receiving a probe response frame from any particular AP that it has already obtained prior information, e.g., from the previous scan or reception of a beacon frame. The key advantage of FIL is that unnecessary probe response frames may be eliminated.

ii. Comprehensive probe response (CPR): The CPR make use of the existing neighbor report element that may be optionally included in the probe response frame so that a responding AP may identify in a single probe response frame the system information of itself and surrounding APs which are in vicinity of the scanning STA. The key advantage of CPR is that the probability of probe request and/or response pollution and number of channels to be scanned may be reduced.

Interested readers are referred to [8] for further details of the proposed active scanning enhancements within the TGai.

III. ANALYTICAL MODEL

The performance of TGai reference use case, comprising of \( n \) scanning STAs, \( m \) responding APs, and \( k \) data STAs per AP (generating medium load when necessary) in the ESS, illustrated in Fig. 1 can be analyzed with the fusion of Markov chain and queuing models. We consider that each \( m \)th STA performs only active scanning and may perform active scanning for either initial access (in the unassociated state) or discovery (in the associated state). Moreover, each \( m \)th AP transmits a probe response frame to each received probe request frame. Furthermore, each \( k \)th STA generates medium load with a constant bit rate of 512 Kbps by using symmetrical uplink and downlink arrival rates of 32 packets/s, and packet size of 1000 bytes.

In order to capture the transmissions of probe request, probe response, and data frames by the scanning STA, responding AP, and data STA, respectively, we need to consider a heterogeneous and finite load model as in [9]. Further, to examine the performance of TGai reference use case in context of the EDCA procedure, we need to consider a backoff-based prioritization which can be solved numerically by fixed point iteration technique for \( P_{c_{1}}, \ldots, P_{c_{K}} \) and \( \tau_{1}, \ldots, \tau_{KNK} \).

A. Modeling of Transmission and Collision Probabilities

Suppose we denote \( W_{ij} \) as the CW size of an AC with priority \( i \) in the \( j \)th backoff stage. The CW size of \( i \)th AC at different backoff stages \( j \in [0, m] \) can then be written as

\[
W_{ij} = \begin{cases} 
2^j W_{0} & \text{if } 0 \leq j \leq m' \ \\
2^{m'} W_{0} = CW_{\text{max}}, & \text{if } m' < j \leq m
\end{cases}
\]  

(3)

\( W_{ij} \) is the CW size which depends on the number of retransmissions encountered by a packet, \( CW_{\text{min}} \) is the minimum CW size, \( m' \) is the maximum CW increasing factor, and \( CW_{\text{max}} \) is the maximum CW size for \( i \)th AC. \( m \) is the retry limit, which is also the maximum backoff stage, and is the same for all ACs.

Without loss of generality, we compute the transmission probability \( \tau_{ij} \) and collision probability \( P_{c_{i}} \), as shown in (4) overleaf. The derivations of (4) following [9] and [10] are omitted due to space. Note that (4a) gives an expression for the per-STA transmission probability \( \tau_{ij} \) where \( i = 0, \ldots, K \) is the AC label and \( j = 1, \ldots, N_{i} \) is the STA label. Accordingly, (4a) and (4b) form \( 2^{N_{i}} \) coupled non-linear equations which can be solved numerically by fixed point iteration technique for \( P_{c_{1}}, \ldots, P_{c_{K}} \) and \( \tau_{1}, \ldots, \tau_{KNK} \).

B. Modeling of MAC Service Time

First, we model the transmissions of probe request frame by \( n \) scanning STAs, probe response frame by \( m \) responding APs, and data frame by \( k \) data STAs per AP in the ESS of Fig. 1 as Type I, II, and III events, respectively, with different time slot durations. The probabilities of Type I, II, and III events occurring in the medium on the long-term can then be derived as

\[
P_{T_{\text{type}}_i} = \frac{n}{n + m (n + 2k)}, \tag{5a}
\]
\[ \tau_{ij} = \begin{cases} \frac{2(1-2P_{cij})}{W_0(1-(2P_{cij})^{m+1})} \frac{(1-P_{cij})}{(1-P_{cij})} \left(1-2P_{cij}\right)^2, & m \leq m_i \, , \\ \frac{2(1-2P_{cij})}{W_0(1-(2P_{cij})^{m+1})} \frac{(1-P_{cij})}{(1-P_{cij})} + 2m_i \left(W_0 \frac{(1-P_{cij})^{m+1}}{(1-P_{cij})} \right), & m > m_i \end{cases}, \]  

\[ P_{cij} = 1 - \prod_{m_i \neq j} \prod_{m_i = i, m_i \neq \bar{i}} \left[1 - (1 - P_{bmn}) \tau_{mn}\right], \quad i, m = 0, \ldots, K, \quad j, n = 1, \ldots, N_i, \]  

from which we can approximate the medium busy time due to a successful transmission as 

\[ T_s = \sum_{i=\text{TypeI}} T_{OH,i} \times P_i \]  

\[ + T_{SYM} \times \left[ \frac{L_{SER} + 8L_{DATA} + L_{TAIL}}{N_{DBPS}} \right], \]  

where

\[ T_{OH,\text{TypeI}} = T_{PHY} + \delta + T_{SIGEXT} + T_{DIFS}, \]  

\[ T_{OH,\text{TypeII}} = 2T_{PHY} + T_{SIFS} \]  

\[ + T_{SYM} \times \left[ \frac{L_{SER} + 8L_{ACK} + L_{TAIL}}{N_{DBPS}} \right] 

\[ + 2\delta + 2T_{SIGEXT} + T_{DIFS} = T_{OH,\text{TypeII}}, \]  

\[ L_{DATA} = L_{MAChdr} + \sum_{i=\text{TypeI}} \frac{\alpha_i \beta_i L_{PLD,i}}{\sum_{i=\text{TypeI}} \alpha_i \beta_i}, \]  

\[ \beta_i = \frac{\max(0, W_0)}{W_0}. \]  

\[ P_{T_{\text{TypeII}}} = \frac{mn}{n + m(n + 2k)}, \]  

\[ P_{T_{\text{TypeII}}} = \frac{2km}{n + m(n + 2k)}, \]  

\[ T_{DIFS} \]  

during a collision, \( T_c = T_s \) when (6a) can be reasonably approximated on the long-term for an infrastructure BSS [9]. In addition, we note that this approximation improves as long as \( m \) and \( n \) are large as in the TGari reference use case where (5b) will dominate.

Next, the closed-form of the average MAC service time can be expressed as the total amount of time spent by STA \( j \) of \( i \)th AC in both the backoff and transmission states given in [9] by

\[ E[T_{S_{ij}}] = E[T_{BO_{ij}}] + E[T_{TX_{ij}}]. \]  

Upon deriving (7), the key performance metrics of MAC delay, PLR, and throughputs efficiency can also be obtained from a finite queueing model as in [9].

### IV. Numerical Results

The numerical results are obtained based on the analytical model presented in the previous section and the system parameters in Table I. It is evident from Fig. 3 that the MAC service time is dependent on the collision probability. Hence, it exhibits similar trends that correlate very well to the collision probability over the different range of per-STA active scan rates. On the other hand, all the other performance metrics derived from the queueing analysis, viz. queue length, MAC delay, packet loss rate, and throughput are dependent on the MAC service time.

The results also show the effect of per-STA active scan rates and data rates on the link setup time and the scalability requirement of FILS. When \( m = 1 \), \( 6 \) Mbps (single AP with
fixed data rate of 6 Mbps), the average active scanning duration (from the MAC delay plot) is about 4.9 ms and $UB_{pktXchg}$ (cf. (2) for definition) is about 1000 packets/s (from the throughput or departure rate plot). From (1), this implies that DCF can meet the functional requirements of FILS provided that there is only a single responding AP and no medium load, which may not be a typical ESS scenario in practice. When multiple responding APs exist in the ESS, $UB_{pktXchg}$ reduces to 500 packets/s and 400 packets/s for $m = 3$, 6 Mbps and $m = 5$, 6 Mbps, respectively.

On the other hand, when the data rate of DCF is increased from 6 Mbps to 54 Mbps (control rate of 24 Mbps is used for ACK frames), the average active scanning duration is less than 2 ms and $UB_{pktXchg}$ is 1000 packets/s $\forall m$, 54 Mbps (multiple APs with fixed data rate of 54 Mbps). Clearly, a higher data rate can effectively reduce the MAC service time (or increase the MAC service rate $\mu$) which results in a lower collision probability. This has the effect of extending $UB_{pktXchg}$ (or accommodating a higher arrival rate $\lambda$) for a given traffic intensity $\rho = \frac{\lambda}{\mu}$ in the ESS.

Moreover, the analysis presents a number of important characteristics as follows:

- A linear relationship between the per-STA active scan rate and the throughput exists where the throughput increases with the per-STA active scan rate when the medium is not saturated.
- The point where the maximum throughput occurs is dependent on the per-STA active scan rate and the number of responding APs. In other words, the rate at which the medium saturates is dependent on the offered load.
- The transition of medium from the non-saturation to saturation state occurs when there is a marked increase in the collision probability, MAC service time, queue length, MAC delay, and packet loss rate, as well as a corresponding decrease in the throughput.

Next, Fig. 4 shows the effect of packet lengths and data rates on the link setup time and scalability requirements of FILS. We show only the MAC delay and throughput results due to space. In general, the results indicate a longer packet length can be better tolerated with a higher data rate of 54 Mbps as compared to 6 Mbps in the presence of multiple APs. E.g., when the per-STA active scan rate is high (10 packets/s), $UB_{pktXchg}$ occurs at a point where the packet length is short. In addition, the average active scanning delay increases and throughput reduces with increasing packet length when the medium saturates. Moreover, except for the case when the per-STA active scan rate is low (1 packets/s), the throughput is much lower when data rate is 6 Mbps as compared to that of 54 Mbps which is expected.

From both Figs. 3 and 4, it is clear that the average active scanning duration and $UB_{pktXchg}$ are dependent on the offered load which is strongly influenced by the active scan rates, number of responding APs, packet lengths, data rates, medium load, and number of scanning STAs (albeit not shown explicitly in this study). This implies that enhancements targeted at reducing these factors will bring about improvements in supporting the functional requirements of FILS. Moreover, since $UB_{pktXchg}$ considers the delay requirement as defined in (1), the point where it occurs is lower than the maximum throughput. Further, we note that the average active scanning duration per channel can be maintained below 5 ms so long as the medium is in the non-saturation state.

Last, Fig. 5 provides the performance comparison between DCF, EDCA, FIL, and CPR by including high medium load. The notation $EDCA(data, request, response)$ denotes the respective $CW_{min}$ values for different frames. Fig. 5(a) shows that DCF has the worst performance which is expected. A comparison between DCF, $EDCA(31, 15, 15)$, and $EDCA(31, 15, 7)$ suggests that it is effective to prioritize the frames which are exchanged in the medium in the decreasing order of priority: beginning with the probe response frame, probe request frame, and lastly data frame. It is interesting to note that the average active scanning duration of CPR increases rapidly from the onset of medium saturation as compared to FIL. This is because when the number of surrounding APs increases, the probe response frame size increases by a factor of $(m - 1) \times L_{neighborReport}$. This in turn results in a longer time slot duration following a collision which occurs more frequently after the medium saturates. On the other hand, FIL provides the best performance with a simple and effective way to reduce the average active scanning duration.

Fig. 5(b) shows that $UB_{pktXchg}$ with 100 scanning STAs, 3 responding APs, and 6 data STAs (high medium load of approximately 60%) is 200 packets/s for DCF, 400 packets/s for $EDCA(31, 15, 15)$, 500 packets/s for $EDCA(31, 15, 7)$,
active scan rates ($\lambda$) medium load ($n$) responding APs ($m$) = 0 to 5, different per-STA active scan rates ($\lambda_{type} = 1.5, 10$ packets/s), varying per-STA packet lengths.

Fig. 4. Key performance metrics per responding AP with ERP-OFDM PHY (DCF) @ 6 Mbps (Mandatory) and 54 Mbps in an ESS with a fixed number of scanning STAs, multiple responding APs, and high medium load. Note that both FIL and CPR have lower throughput due to lower offered load as $1/3$ of the probe response frames are filtered and $2/3$ of the probe response frames are consolidated to a single probe response frame, respectively, according to § II-D.

V. CONCLUSION

We have presented a performance analysis to evaluate how well the legacy 802.11 DCF, EDCA, and the upcoming 802.11ai amendment can support FILS which involves a large number of users entering and leaving an ESS with multiple APs over a short interval. Through this analysis, we are able to obtain the average active scanning duration and the theoretical limits of achievable packet exchanges from the key performance metrics of MAC delay and throughput or departure rate, respectively. We have shown that the prioritization of frames using CW differentiation is useful to relieve congestion at the APs. We have also demonstrated that a combination of EDCA and the proposed active scanning enhancements provides performance gains of at least 20% and up to 250% in supporting the functional requirements of FILS as compared to the 802.11 EDCA and legacy DCF, respectively.

REFERENCES