Abstract—Voice over Internet Protocol (VoIP) is a form of voice communication that uses audio data to transmit voice signals to the end user. VoIP is one of the most important technologies in the World of communication. Around, 20 years of research on VoIP, some problems of VoIP are still remaining. During the past decade and with growing of wireless technologies, we have seen that many papers turn their concentration from Wired-LAN to Wireless-LAN. VoIP over Wireless LAN (WLAN) faces many challenges due to the loose nature of wireless network. Issues like providing Quality of Service (QoS) at a good level, dedicating capacity for calls and having secure calls is more difficult rather than wired LAN. Therefore VoIP over WLAN (VoWLAN) remains a challenging research topic. In this paper we consolidate and address major VoWLAN issues. This research is helpful for those researchers wants to do research in Voice over IP technology over WLAN network.  

Keywords—Capacity, QoS, Security, VoIP Issues, WLAN.

I. INTRODUCTION

A. Background of Voice over IP

VoICE over Internet Protocol (VoIP) is one of the most important technologies in the World of communication. VoIP is simply a way to make phone calls through the internet. VoIP transmits packet via packet-switched network in which voice packets may take the most efficient path. On the other hand, the traditional public switched telephone network (PSTN) is a circuit-switched network which requires a dedicated line for telecommunications activity [1]. Furthermore, Internet was initially used for transmit data traffic and it is performing this task really well. However, Internet is best-effort network and therefore it is not sufficient enough for the transmission of real-time traffic such as VoIP [2].

In addition, there are about 1 billion fixed telephone lines and 2 billion cell phones in the World that use PSTN systems. In the near future, they will move to networks that are based on open protocols known as VoIP [3]. That can be seen from the increasing number of VoIP users. For instance there are more than eighty million subscribers of Skype; a very popular VoIP commercial application [4]. VoIP has gained popularity due to the more advantages it can offer than PSTN systems especially that voice is transmitted in digital form which enables VoIP to provide more features [5]. However, VoIP still suffer few drawbacks which user should consider when deploying VoIP system [6].

B. VoIP components

Basically, VoIP system can be configured in these connection modes respectively; PC to PC, Telephony to Telephony and PC to Telephony [7].

VoIP consists of three essential components: CODEC (Coder/Decoder), packetizer and playout buffer [8], [9]. At the sender side, an adequate sample of analogue voice signals are converted to digital signals, compressed and then encoded into a predetermined format using voice codec. There are various voice codecs developed and standardized by International Telecommunication Union-Telecommunication (ITU-T) such as G.711, G.729, etc. Next, packetization process is performed which fragment encoded voice into equal size of packets. Furthermore, in each packet, some protocol headers from different layers are attached to the encoded voice. Protocols headers added to voice packets are of Real-time Transport Protocol (RTP), User Datagram Protocol (UDP), and Internet Protocol (IP) as well as data link layer header. In addition, RTP and Real-Time Control Protocol (RTCP) were designed at the application layer to support real-time applications. Although TCP transport protocol is commonly used in the internet, UDP protocol is preferred in VoIP and other delay-sensitive real-time applications. TCP protocol is suitable for less delay-sensitive data packets and not for delay-sensitive packets due to the acknowledgement (ACK) scheme that TCP applies. This scheme introduces delay as receiver has to notify the sender for each received packet by sending ACK.

The packets are then sent out over IP network to its destination where the reverse process of decoding and depacketizing of the received packets is carried out. During the transmission process, time variations of packets delivery (jitter) may occur. Hence, a playout buffer is used at the receiver end to smoothen the playout by mitigating the incurred jitter. Packets are queued at the playout buffer for a playout time before being played. However, packets arriving later than the playout time are discarded. The principle components of a VoIP system, which covers the end-to-end transmission of voice, are illustrated in Figure 1.
C. Background of Wireless LAN

Wireless LAN is one of the mainly organized wireless technologies all over the World and is likely to play a major function in the next-generation wireless voice call networks. The architecture of this type of network is the same as Local Area Network (LAN)’s except that the transmission happens via radio frequency (RF) or Infrared (IR) and not through physical wires/cables, and at the MAC sub-layer, it uses different standard protocol. The main characteristics of the WLANs are mobility, simplicity, scalability, edibility and cost effectiveness.

Simply, wireless network allows nodes to communicate with each other wirelessly and it can be configured in two ways: infrastructure less mode and infrastructure mode. In the first mode which is called Ad hoc or peer to peer (P2P) network, there is no fixed point and each node can directly communicate with all other nodes. On the other hand, the second method of wireless is where the transmission between two or more nodes goes through a third node called Access Point (AP) [11] as shown in figure 2. Thus when any terminal wants to send packets to other terminal, packets would be sent to the AP first which will forward them to their destination. In each terminal, one of the IEEE 802.11 WLAN standards protocols is deployed. There are several 802.11 protocols operating on different frequency bands such as: 802.11a support 5GHz and 54Mbps data rate, 802.11b supports 2.4GHz and 11Mbps, 802.11g support 2.4GHz and 54 Mbps, and 802.11n supports 2.4 or 5 GHz and 150Mbps data rate [23], [24]. The 802.11 workgroup in the institute of IEEE is responsible for developing and setting these standards and it added alphabetic character in each standard name, such as "a" or "b" to further describe each group that has been assigned for a specific task in developing the WLAN standard [12].

Conversely, WLAN provides connections to the IP networks and VoIP applications are already running over IP networks. Consequently, these two new technologies are merged to incorporate VoIP over WLANs (VoWLAN).

D. IEEE 802.11 MAC

The IEEE 802.11 WLAN networks known as a wireless Ethernet play an important function in future-generation networks. Medium Access Control (MAC) sub-layer categorizes two functions, Distributed Coordination Function (DCF) and Point Coordination Function (PCF).

The IEEE 802.11 WLANs support both contention-based DCF and contention-free PCF functions. PCF uses centralized polling system which requires AP as access coordinator. DCF obligates a station (STA) to listen for the channel status; it is busy or free. DCF uses Carrier Sensing Multiple Access/Collision Avoidance (CSMA/CA) as the access method [13]-[16]. CSMA/CA deploys InterFrame Space (IFS), Contention Window (CW) and Acknowledgment. IFS is waiting period of transmission over IP-based network. In DCF access method, STA senses traffic before send or share packet over IP-based network. If DCF found medium is sensed busy, sender would be waiting until the medium ready for the transmission. This process is called DCF interfram space (DIFS). Then the sender STA will send Request-to-Send (RTS) to get permission from receiver by sending Clear-to-Send (CTS). Meanwhile, other STAs are informed about the time period of transmission to renew the local timer of their Network Allocation Vectors (NAV). CW is a number of slots which ranging from 0 to 1. CW stops its timer while station finds the channel busy due to overflow or bursty traffic over networks and restart timer when the channel is sensed as idle. STA can calculate the random counter interval backoff and choose from CW. The senders then enter the backoff phase, in which every sender chooses a random backoff counter from (0, CWmin). Then CW as: Backoff Phase = rand (0, CW), slot time, where CWmin<CW<CWmax. The backoff counter is reduced by 1 for each idle time slot and stop until busy period. When the backoff counter reaches 0, then sender starts the transmission and the receiver sends an ACK to sender [17]-[22].
II. VoWLAN QoS Issues

Quality of service (QoS) can be defined as the network ability to provide good services that satisfy its customers. In other words, QoS measures the degree of user satisfactions; the higher the QoS, the higher degree of user satisfaction.

QoS is not a big concern, when WLAN is mainly used to transmit data packets, as opposite with the increasing demand on applying VoIP over WLAN and transmitting voice packets, QoS has become a critical issue [25], because real-time applications, unlike data (non real-time) applications, are very sensitive to delay. Therefore, QoS of VoIP is an import concern to ensure that voice packets are not delayed, lost or dropped during the transmission over the network. On the other hand, as VoIP is considered a viable alternative of PSTN and it is expected to replace it in the future. VoIP has to provide customers with high quality of services, the same as the quality of PSTN services or better. Therefore, VoIP applications require stringent QoS in order to satisfy their users. However, IP networks still cannot meet the required QoS of VoIP [26]. VoIP quality of service is measured based on different parameters like delay, jitter, packet loss, throughput and echo. VoIP QoS is improved by controlling the values of these parameters to be within the acceptable range [27]. Factors affecting QoS are briefly described in the following sections.

A. Delay

Delay can be defined as the total time it takes since a person, communicating another person, speaks words and hearing them at the other end. Unlike data applications, VoIP applications are very sensitive to delay although they can tolerate packet loss to some extent. End-to-end or mouth to ear delay is one of the main factors affecting QoS and should be less than 150ms for good network connection as defined by ITU G.114 [26] while delay of less than 100ms is defined by the European Telecommunications Standard Institute (ETSI). Delay is mainly caused by network congestion which leads to a slow delivery of packets [28]. Furthermore, delay is affected by several parameters or algorithms which can be categorized into: delay at the source, delay at the receiver, and network delay. Some of the delay parameters are known while some others are inconsistent [29].

1) Delay at the source

The delay of the whole process performed at the sender side before transmitting the voice packet over the network is caused by several components: codec, packetization and process [30]. Codec functions introduce some delays when processing the analogue-to-digital conversion. The more bits compressed, the less the bandwidth required, and the longer the delay added. For packetization delay, it’s the time taken to place the chunks of frames in packets which would be transmitted across the network. The third component of source delay is when the computer passes the packets into the network for transmission to other side.

2) Delay at the receiver

The reverse process that carried out at the sender is performed at the receiver adding more delay: process delay and decoding delay including decompressing delay. Additionally, Playback delay is incurred when playing out the voice stream which includes the jitter buffer delay as well.

3) Network delay

Network delay in WLAN environment is the total delay of both WLAN and backbone networks as illustrated in figure 2. Queueing, transmission and propagation are other components of network delays. The propagation delay is the delay in the physical media of the network, while transmission delay includes router’s delay and MAC retransmission delay [31].

B. Jitter

IP network does not guarantee of packets delivery time which introduces variation in transmission delay. This variation is known as jitter [25] and it has more negative effects on voice quality [8].

Since voice packets of the same flow are not received at the same time. Therefore, jitter buffer are introduced to diminish the jitter effect and make the conversation smoothly as it holds a number of packets in a queue before playout. The buffer queue size can be fixed or adaptive which varies based on network condition, voice character [31], for better performance. Buffer jitter adaptive techniques perform better as it reduces the possibility of buffer overflow and underflow.

Overflow issue is where number of packets received is getting larger than the buffer size as a result buffer discards packets that cannot hold. On the other hand, underflow buffer occurs when some packets are needed for playout but buffer is empty. Hence, several adaptive buffer algorithms, such as in reference [31], [32] and [33], are introduced to adjust buffer size in order to improve quality of VoIP.

C. Packet loss

Packets transmitted over IP network may be lost in the network or arrived corrupted or late. Packets would be discarded, when they arrive late at the jitter buffer of the receiver or when there is overflow in jitter buffer or router buffer. Therefore packet loss is the total loss occurs due to network congestion and late arrival [33]. In case of packet loss, the sender is informed to retransmit the lost packets and this is cause more delay and thus affecting transmission QoS.

Furthermore, VoIP system can tolerate packet loss to some extend as 1% or less is acceptable for roll quality while for business quality 3% or less is acceptable [34]. Hence, more than 3% of packet loss degrades the speech quality.

Techniques and algorithms have been designed to resolve packet loss problem. MAC sub-layer at the sender side uses the acknowledgement scheme to retransmit lost packets but this technique is bandwidth consuming. In addition, forward-error correction (FEC) is a mathematical technique that helps receiver to reconstruct lost packets from previously sent packets.
On the other hand, codecs at the higher layer employ their own error concealment techniques to recover packets lost. The techniques used by codec called packet-loss concealment (PLC). Consequently reference [35] has proposed a technique which makes use of the available recovery techniques. The proposed scheme dynamically changes the recovery scheme according to the rate of packet loss.

Moreover, Fragmentation technique at the MAC layer also plays a role in packet loss due to the relation between packet size and packet loss rate. The packet with bigger size has higher probability to be discarded than the ones with small size in case of errors in the packet [36]. Thus, fragmentation techniques should be carefully designed considering its effect on packet loss rate.

D. Echo

In VoIP, Echo is experienced when caller at the sender side hears a reflection of his voice after he talked on the phone or the microphone whereas the callee does not notice the echo. Echo is the term of the reflections of the sent voice signals by the far end. Echo could be electrical echo which exists in PSTN networks or acoustic echo which is an issue in VoIP networks [37]. Electrical echo is generated due to the conversion between the 2 wire lines and other circuit of other type in a hybrid environment. While the acoustic coupling between the phone’s microphone and speaker causes the acoustic echo. Echo is affected by delay and it becomes noticeable when it is loud and delayed mostly when the round trip delay is more than 50 ms [25]. Since round trip delay of VoIP can exceed 100ms, echo cancellers are being used.

E. Throughput

This parameter concerns about the maximum number of bits received out of the total number of bits sent during an interval of time. In IEEE 802.11 networks, the bit rate of each standard is specified such as IEEE 802.11b transmits at 1Mbps, 2Mbps, 5.5Mbps and 11Mbps. Furthermore, the standard has the capability of performing dynamic rate switching with the objective of improving performance of wireless link [38].

However, transmission throughput is still less than that specified in the standard due to the overhead introduced by WLAN network protocols. The throughput achieved is between the range 50% to 70% of the transmission rate which is low comparing to Ethernet throughput which achieves 80% to 90% of the transmission rate [39]. There are methods have been introduced to improve the throughput of WLAN such as aggregation technique [40] and adaptive techniques [41]-[43].

III. VoIP CAPACITY OVER WLAN

VoIP is placed on IP-based WLAN IEEE 802.11 standards. Combination of these two technologies and collaboration with Next Generation Networks (NGNs) might be a leading application in the World of communication [44]. VoIP over WLANS main designing issue is to verify the network voice capacity. The voice capacity of WLANs is measured, based on IEEE 802.11 standards, voice codes and packetization intervals [45]. The following table states the capacity of IEEE 802.11b.

<table>
<thead>
<tr>
<th>No.</th>
<th>Packet Interval (ms)</th>
<th>G.711 Connections</th>
<th>G.729 Connections</th>
<th>G.723 Connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>11</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>15</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>19</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>22</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>25</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>

IV. VOIP QUALITY OF SERVICE MEASUREMENTS

Speech quality measurement methods for VoIP are divided into subjective and objective methods [46]. The tow methods are explained briefly in the following paragraphs.

A. Subjective methods

Subjective methods aim to find the average user perception of a system’s speech quality by asking a panel of human listeners and provide them a limited response choice. ITU-T in Recommendation P.800 [47] introduced Mean Opinion Score (MOS) based on user perception that is ranged from 1 (poor) to 5 (excellent) for subjective determination of voice quality.

B. Objective methods

Objective methods are based on mathematics that measure physical quantities of the system (i.e., delay and packet loss) [34]. Objective methods include two techniques:

1) Intrusive: this method generally uses injection of test signal into the network and during this time network cannot carry live traffic. Intrusive methods can be applied during developing system before presence of users [46].

2) Non-intrusive: this method executes without reference signal and developed to apply on real-time traffic. Usually this type of method performs its prediction directly from network impairment parameters such as jitter, delay and packet loss [48].

a) PSQM

In recommendation P.861 they standardized Perceptual Speech Quality Measurement (PSQM) and improved their original model of MOS score. A score range is [0 - 6.5]. 0 represents a perfect alignment between the reference and distorted output signal means excellent quality. According to the scope of P.861, PSQM is suitable for evaluating quality of speech codecs [49].

b) PESQ

ITU-T recommendation P.862 explained Perceptual Evaluation of Speech Quality (PESQ). Since PSQM was not able to take proper account of variable delay, coding
distortions and channel error in real systems, PESQ was designed to address these effects [50]. It has similar result scale to MOS and range is from 1 to 4.5.

c) PAMS

British Telecom developed Perceptual Analysis Measurement System (PAMS). From the measurement point of view PAMS is similar to PSQM as both of them use the comparison method for received signal and source signal. It implements a model based on human perceptual factors [48]. The resulting score is similar to MOS score from 1 to 5.

d) E-model

E-model is presented by ITU-T Recommendation G.107 which incorporates PQSM and PAMS. E-model combines a different impairment factor to predict QoS as R-value.

\[
R = R_0 - Is - Id - Ie + A \quad (1)
\]

Where: \(R\) represents of the resulting voice quality (from 0 to 100), \(R_0\) refers to signal to noise ratio, \(Is\) characterizes the simultaneous impairment factor such as too loud speech level [34], \(Id\) represents mouth-to-ear delay, \(le\) is equipment impairment factor (e.g. codec specific characteristics), and \(A\) is advantage of access.

e) Other measurement methods

Recently ITU-T approved 3 Recommendation P.561 [51], P.562 [52] and P.563 [53] and IETF recommended an extension report to RTCP protocol called RTP Control Protocol Extended Reports (RTCP-XR) [54]. Also authors in [44] have proposed a new "Quality and Quantity" factor, which combines these three metrics; blocking and dropping rate also voice quality (MOS value).

V. VOIP SECURITY CHALLENGES

In comparing to PSTN, VoIP faces many security issues and PSTN provide high level of security due to their loosely characters [55] but some of the issues are similar for both like eavesdropping type of attacks and toll fraud. "Man in the middle" is a type of eavesdropping which listening and changing conversation is possible for attacker. Attacker is able to interrupt the conversations then play back previous conversation instead to change the conversation like "yes" to "no" or take advantage by asking username and password [56]. In toll fraud, the motivation is making free calls especially for high cost long distance calls [55], actually fraudulent calls are billed to a victim, and it is typically possible due to miss configurations.

Since Internet is insecure and it is the underlying network for VoIP, some threats and risks inherited from this network to the VoIP. The most important risk is Denial of Service (DoS) attack which tries to make resources unavailable or impede a service from functioning well which is vary in performing and motivation. VoIP DoS attack aimed to bring a service down. VoIP runs voice over stream of IP packets, for example hackers can keep packet streaming running or they can make system busy with flooding of call requests. Consequently they will make inefficiency in use of resource, hearing long term busy signals and force disconnects of calls due to lack of resources [55].

Another security threat is Spam over Internet Telephony (SPIT) which troubles only VoIP like VoIP spam is related to voice-mail inboxes, in fact it is similar to spam Email in mailbox, spammers will send many unwanted message and audio-commercials to make voice-mail inboxes full to need extra voice storage capacity.

Other issue like Protocol Fuzzing is belongs to IP applications [55]. Fuzzing is generating random or sequential input data to inject them in an application normally fuzzing is utilized for debugging and testing IP applications [58]. In protocol fuzzing (malformed messages) target is a protocol like SIP stack protocol. Attacker use this method to finds vulnerabilities and creates various packet type that aimed target system to process these packets instead of normal packet, and as result system will face with failures or crash, buffer overflow, application delays and Infinite loop.

This paper has focused on VoIP over WLANs and wireless network uses open medium for communication, on such network authenticate VoIP traffic is required.

Available Solutions:

Some recommendations in end-user level and infrastructure level for secure VoIP discussed in [59]. For last issue, WEP (Wired Equivalent Privacy) is the weakest attempt, WPA (Wi-Fi Protected Access) is stronger and WPA2 (enhanced WPA) is advance encryption algorithm suggested to have secure IEEE 802.11 WLANs. Moreover, the isolation of the corporate VoIP network from the public Internet will enhance network security and performance. Due to this fact, incorporating VPN tunneling suggested in [60]. Spam filters can be applied, to scan voice-based SPIT messages for prohibited content. Some security issues can be controlled by utilizing IPSec (a frame work for securing communication by encrypting each packet), firewall and NAT.

In spite of existence of powerful prevention technologies, VoIP still faces many security challenges which is require to consideration in any VoIP deployment. Implementing security mechanism such as encryption will lead to more delay due to encryption and decryption also building new header to packets [34] thus it has effect on QoS and capacity therefore it should be consider before tuning and developing security policies.

VI. CONCLUSION

This paper made a review on VoIP over WLAN, its advantages and challenges. As a major concerned issue, QoS was explained and its contributing factors were stated in details. Although, VoIP can tolerate packet loss to some extent, it is very sensitive to delay factor. Jitter also plays a main role on voice quality therefore jitter buffer is introduced to smooth the playout of packets. Echo and throughput are also other factors that affect the quality of voice. Different techniques for voice QoS assessment were mentioned as well. Furthermore, WLAN is a bandwidth limited network which may limit the number of VoIP calls. Accordingly, this paper discussed capacity and its measurement technique as it is base on IEEE 802.11b standard with voice codes and millisecond
packetization intervals. At the same time as security issues in VoIP poses a challenge, security issues and their available solutions are also investigated in this paper.

VII. FUTURE WORK

In this research paper, we have discussed VoIP issues over WLAN networks. We will extend this survey paper in detail for beginner of VoIP researchers. A possible and useful future work is the evaluation of VoIP. We will put some results to improve VoIP over WLAN networks.

ACKNOWLEDGMENT

We would like to thanks PETRONAS University for their support.

REFERENCES

[29] IEEE Standard for Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements Part 11: "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications"


