IEEE P802.11  
Wireless LANs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| IEEE 802.11 Real Time Applications TIG Report | | | | |
| Date: 2018-11-11 | | | | |
| Author: | | | | |
| Name | Affiliation | Address | Phone | email |
| Kate Meng | Tencent Technology (Shenzhen) Company Limited | C1, Hi-Tech Park, NanShan, Shenzhen, China | +86 166-7516-1765 | Katemeng@tencent.com |
| Allan Jones | Activision | 3100 Ocean Park Blvd, Santa Monica, CA 90405 | (310) 255-2000 | Allan.jones@activision.com |
| Dave Cavalcanti | Intel | 2111 NE 25th Ave, Hillsboro OR 97124, USA | (503)712-5566 | dave.cavalcanti@intel.com |
| Karthik Iyer | Activision | 3100 Ocean Park Blvd, Santa Monica, CA 90405 | (310) 255-2000 | Karthik.Iyer@activision.com |
| Chenhe Ji | Huawei Technologies Co.,Ltd. | 101 Software Avenue, Yuhua District, Nanjing, China | +86-13813901984 | jichenhe@huawei.com |
| Kazuyuki Sakoda | Sony | 1730 N. First Street, San Jose CA 95112, USA |  | kazuyuki.sakoda@sony.com |
| Akira Kishida | NTT | 1-1, Hikarinooka Yokosuka-Shi, Kanagawa 239-0847, Japan | +81-46-859-2093 | akira.kishida.fs@hco.ntt.co.jp |
| Frank Hsu | MediaTek Inc. |  |  | frank.hsu@mediatek.com |
| James Yee | MediaTek Inc. |  |  | james.yee@mediatek.com |
| Lander Li | Tencent Technology (Shenzhen) Company Limited | C1, Hi-Tech Park, NanShan, Shenzhen, China |  | Landerli@tencent.com |
| Evgeny Khorov | IITP RAS | Bolshoy Karentny per 19, Moscow, Russia |  | khorov@iitp.ru |

Abstract

This document contains report for 802.11 working group RTA TIG, which provides description of real-time application usage model, problem statement, technical requirements and potential solutions.

**Table of contents**

[1 Definitions and abbreviations 7](#_Toc3325488)

[2 Introduction 8](#_Toc3325489)

[3 Purpose 9](#_Toc3325490)

[4 Use Cases 10](#_Toc3325491)

[4.1 Real-time Mobile Gaming 10](#_Toc3325492)

[4.1.1 Challenges and problem highlight 10](#_Toc3325493)

[4.1.2 Network architecture 12](#_Toc3325494)

[4.1.3 Traffic characteristic 12](#_Toc3325495)

[4.1.4 Traffic model 13](#_Toc3325496)

[4.1.5 Requirement metrics 14](#_Toc3325497)

[4.2 Wireless Console Gaming 16](#_Toc3325498)

[4.2.1 Introduction 16](#_Toc3325499)

[4.2.2 Network architecture for an FPS game 17](#_Toc3325500)

[4.2.3 Packet flow of an FPS console game 18](#_Toc3325501)

[4.2.4 Traffic model 19](#_Toc3325502)

[4.2.5 Problem statement 19](#_Toc3325503)

[4.3 Cloud Gaming 20](#_Toc3325504)

[4.3.1 Introduction 20](#_Toc3325505)

[4.3.2 Network architecture 20](#_Toc3325506)

[4.3.3 Traffic characteristic 21](#_Toc3325507)

[4.3.4 Problem statement 22](#_Toc3325508)

[4.4 Industrial Systems 24](#_Toc3325509)

[4.4.1 Real-time industrial use case examples 24](#_Toc3325510)

[4.4.2 Summary of real-time industrial use cases and classes of service 25](#_Toc3325511)

[4.4.3 Traffic profiles 26](#_Toc3325512)

[4.4.4 Problem statement 27](#_Toc3325513)

[4.5 Real-time video 28](#_Toc3325514)

[4.5.1 Introduction 28](#_Toc3325515)

[4.5.2 Interactive video rationale 28](#_Toc3325516)

[4.5.3 Traffic profiles and requirements 30](#_Toc3325517)

[4.5.4 Problem statement 30](#_Toc3325518)

[4.6 Drone Control 31](#_Toc3325519)

[4.6.1 Introduction 31](#_Toc3325520)

[4.6.2 Use cases 31](#_Toc3325521)

[4.6.3 Architecture 32](#_Toc3325522)

[4.6.4 Problem statement 33](#_Toc3325523)

[5 Potential technical features analysis 35](#_Toc3325524)

[5.1 Time sensitive networking extensions to 802.11 35](#_Toc3325525)

[5.2 Priority tagging (Stream identification) 35](#_Toc3325526)

[5.3 Time-aware shaping (802.1Qbv) over 802.11 36](#_Toc3325527)

[5.4 Dual/multiple link 37](#_Toc3325528)

[5.5 Admission Control 39](#_Toc3325529)

[5.6 Channel resource reservation for real-time traffic with OFDMA 39](#_Toc3325530)

[6 Recommendations 41](#_Toc3325531)

[6.1 Real-time applications requirements summary 41](#_Toc3325532)

[6.2 Implementation specific optimizations 42](#_Toc3325533)

[6.3 New capabilities to support real time applications 42](#_Toc3325534)

[7 References 44](#_Toc3325535)

**Tables**

[Table 4‑1 Comparison of lagging rate between 4G and Wi-Fi 10](#_Toc3325536)

[Table 4‑2 Lagging analysis per 2.4GHz and 5GHz Wi-Fi 11](#_Toc3325537)

[Table 4‑3 Traffic model for status sync 13](#_Toc3325538)

[Table 4‑4 Traffic model for frame lockstep sync 14](#_Toc3325539)

[Table 4‑5 Relationship between game experience and latency 15](#_Toc3325540)

[Table 4‑6 Specifications from client to AP 15](#_Toc3325541)

[Table 4‑8 Cloud gaming traffic profile 22](#_Toc3325542)

[Table 4‑7 Factory automation use cases summary 26](#_Toc3325543)

[Table 4‑8 Real-time video traffic profile 30](#_Toc3325544)

[Table 4‑9 Use case and requirements for drones[13] 32](#_Toc3325545)

[Table 6‑1 Requirements metrics of RTA use cases 41](#_Toc3325546)

[Table 6‑2 802.1Q and WMM UP & AC mapping 42](#_Toc3325547)

**Figures**

[Figure 4‑1 Quadrant of latency and jitter variance 10](#_Toc3311818)

[Figure 4‑2 Real-time mobile gaming network architecture 11](#_Toc3311819)

[Figure 4‑7 The latency breakdown from client to game server 14](#_Toc3311820)

[Figure 4‑8 FPS gaming 15](#_Toc3311821)

[Figure 4‑9 FPS dedicated server architecture 16](#_Toc3311822)

[Figure 4‑10 FPS dedicated server architecture 17](#_Toc3311823)

[Figure 4‑11 FPS traffic model 18](#_Toc3311824)

[Figure 4‑8 Cloud gaming network architecture 20](#_Toc3311825)

[Figure 4‑12 Warehouse automation network architecture 24](#_Toc3311826)

[Figure 4‑13 Communication model of a closed loop control system 25](#_Toc3311827)

[Figure 4‑14 Difference between buffered video and live video 27](#_Toc3311828)

[Figure 4‑15 Video content render process 28](#_Toc3311829)

[Figure 4‑16 Example of use case for digital signage. [12] 30](#_Toc3311830)

[Figure 4‑17 Standalone (single drone) 32](#_Toc3311831)

[Figure 4‑18 Standalone (multiple drones) 32](#_Toc3311832)

[Figure 4‑19 Network control 32](#_Toc3311833)

[Figure 4‑20Example of network delay estimation 33](#_Toc3311834)

[Figure 5‑1 STAs share the medium and queues are managed according to a 802.1Qbv schedule 36](#_Toc3311835)

[Figure 5‑2 Using TSN to resolve contention 36](#_Toc3311836)

[Figure 5‑3 Diagram of Duplicate Mode 37](#_Toc3311837)

[Figure 5‑4 Diagram of Joint Mode 37](#_Toc3311838)

[Figure 5‑5 Random Access RU reservation for real-time STAs 39](#_Toc3311839)

# Definitions and abbreviations

**Disconnection:** when latency is too high even causes packet loss, client will disconnect from game servers after several retries to establish the connection.

**Dual link:** Dual link in this report refers to two independent bands which transmit and receive data simultaneously from one AP to one specific STA.

**End-to-End latency:** System level round trip time (RTT) among devices in a feedback loop system involving 802.11 link to transmit data between devices. End-to-End latency includes all factors that contribute to impact latency, including 802.11 link transmission delay, non-802.11 link transmission delay, signal processing delay, delay caused by synchronization, etc. For instance, in-game latency is one form of the End-End latency.

**Forward link:** Single directional link from any entity to an entity serving application client, i.e., output, user, etc.

**FPS game:** First Person Shooting game, which requires low latency capabilities among game genres.

**Gateway:** Connect WLAN to WAN (Internet). Usually this entity refers to Wi-Fi router.

**Intra-BSS latency:** Time consumed to transmit an MSDU from a STA to another STA within a BSS. It is equal to elapsed time while an MSDU travels from MAC-SAP of a transmitter STA to MAC-SAP of a receiver STA in a BSS.

**Jitter variance:** Jitter variance reflects the fluctuation of latency over time. It can be evaluated by the time difference between two adjacent data arrival at the receiver, i.e., in-game ping values. In this document, the jitter variance is calculated as standard deviation during a period of time.

**Lagging in game:** Lag is a noticeable delay between the action of players and the reaction of the server in a game. Usually, high end to end latency cause lagging and lagging can be reflected as picture freeze or “fast-forward” or frame skipping in game.

**Packet loss:** MSDU delivery failure observed at a receiver STA’s MAC-SAP. Note that packet loss does not include frame error which is recovered by MAC layer retransmissions.

**Real-time mobile game:** The mobile game can host multiple players together in a single game session and transfer data messages between the game server and every connected players. Real-time indicates the feedback should be presented on screen as soon as the users make any actions in game. In order to havegood game experience, the end to end latency from client to game servers plus game servers processing time should be unnoticeable to users as they play.

**Real-time video:** Video streaming application which particularly requires strict real-time delivery. Exact types of the applications include, but not limited to, VR/AR headset relying on main console, video cable replacement by 802.11 link.

**Reverse link:** Single directional link from an entity serving application client, i.e., output, user, to another entity.

**Video frame rate:** The frequency (rate) at which consecutive images/pictures called video frames are displayed on a display.

# Introduction

People have easy access to the Internet via Wi-Fi networks. However, due to the interference and congestion, latency and reliability are not guaranteed, which brings in problems for latency sensitive applications like real-time gaming, robotics and industrial automations. Random high worst-case latency is one of the main problems encountered by real-time applications.

This report consolidates key discussions of RTA TIG about some real-time application scenarios, problems with current Wi-Fi networks and requirements for future 802.11 standards.

# Purpose

The purpose of this report is to provide detailed information on real-time application, issues captured during discussions and submissions that were presented within the group, aiming to help facilitate information sharing within the larger working group. This report also provides feedback that could help to leverage work being done throughout the 802.11 working groups in order to facilitate potential solutions emerging from them.

# Use Cases

## **Real-time Mobile Gaming**

Real-time mobile gaming is a fast-developing application category. Different from traditional games, real time mobile gaming is very sensitive to network latency and stability.

The mobile game can connect multiple players together in a single game session and exchange data messages between game server and connected players. Real-time means the feedback should present on screen as users operate in game. For good game experience, the end to end latency plus game servers processing time should not be noticed by users as they play the game.

The challenges that real-time mobile gaming encounter is the **worst-case latency**. Since the high latency spike is highly likely to cause packet loss and packet disorder, hence impact quality of experience.

### Challenges and problem highlight

Users connect to game under various environment, such as home, restaurants, subway, malls etc.

Around 60% users connect to game via Wi-Fi, with 20-30% possibilities that users encounter lagging in-game.

60-70% lagging because of problems between client and gateway. (around 10-15% are remain unknown which pending a validate process/ test plan to locate problem).

Comparison of lagging rate between LTE and Wi-Fi based on big data is as the following table

|  |  |
| --- | --- |
| Mode | Chance of lagging |
| Wi-Fi | 23% |
| LTE | 15% |

Table 4‑1 Comparison of lagging rate between 4G and Wi-Fi

Below is lagging analysis per 2.4GHz and 5GHz Wi-Fi based on data from over millions pvp (player vs player) games.

|  |  |  |
| --- | --- | --- |
| **Band** | **No lagging** | **Lagging** |
| **2.4GHz** | 74.59% | 25.41% |
| **5GHz** | 86.87% | 13.13% |

Table 4‑2 Lagging analysis per 2.4GHz and 5GHz Wi-Fi



Figure 4‑1 Quadrant of latency and jitter variance

Description of the quadrant:

Latency is from STA to AP RTT. Jitter is standard deviation of latency.

Quadrant A: Network status between STA and AP is good. This is the best scenario.

Quadrant B: Network status between STA and AP barely impact the game experience.

Quadrant C: Network status between STA and AP impact game experience.

Quadrant D: Network status between STA and AP impact game experience negatively.

Game experience of point X and Y shall be enhanced by sacrifice latency to improve the jitter.

### Network architecture

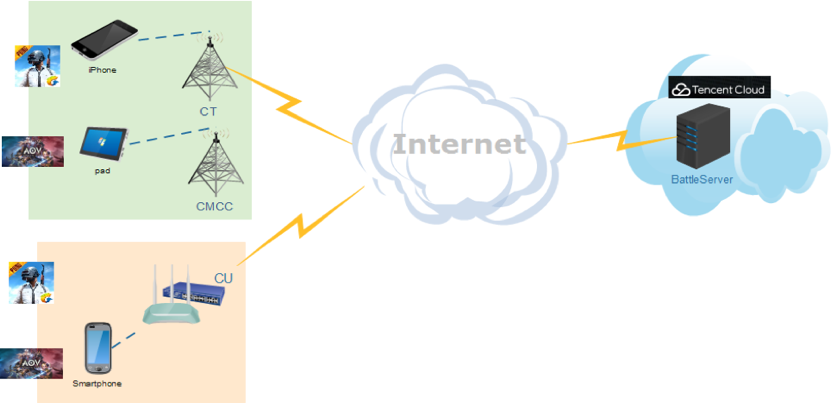
Users connect to game servers with mobile devices like mobile phone and tablets via cellular or wireless network and Internet.

Figure 4‑2 Real-time mobile gaming network architecture

### Traffic characteristic

There are multiple functions running in game applications both chronologically and simultaneously. Following characteristics refers to game traffic only.

**Packet size:** small packet with size of 30-500 Bytes (uplink and downlink), usually downlink packets are bigger than uplink. [[1]](#footnote-1)

**Packet flow rate:** Average every 30-60ms one packet (uplink and downlink), usually downlink packet interval is larger than uplink. [[2]](#footnote-2)

**Bandwidth between client and AP:** 0.1-1Mbps[[3]](#footnote-3)

**Game data:** Usually, the uplink packets carry users’ instructions, while downlink packets carry calculation results or calculation instructions. So, both uplink and downlink packets are important and should be delivered as soon as possible.

**Protocol:** UDP for in-game data.

### Traffic model

There are two popular mechanisms of synchronization between client and game server in real-time mobile games, one is frame lockstep sync, the other one is status sync.

There are three stages took into consideration in the analysis.

Initial stage: Selected users are in same battle room and ready to start gaming. Game map is yet to present to users.

Gaming stage: Users can operate in game.

Ending stage: Game over and present game results.

1. **Traffic model for status sync real-time mobile game**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Component** | **Distribution** | | **Parameters**[[4]](#footnote-4) | | **PDF(probability density function)** |
| L | UL | DL | UL |
| **Initial packet size(Byte)** | Uniform | Uniform | a=0,  b=20 | a=0,  b=20 |  |
| **Packet arrival time (ms)** | Largest Extreme Value | Largest Extreme Value | a=13,  b=3.7 | a=15,  b=5.7 |  |
| **Packet size (Byte)** | Largest Extreme Value | Largest Extreme Value | a=50,  b=11 | a=38,  b=3.7 |  |
| **End packet (Byte)** | Uniform | Uniform | a=500,  b=600 | a=400,  b=550 |  |

Table 4‑3 Traffic model for status sync

1. **Traffic model for frame lockstep sync real-time mobile game**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Component** | **Distribution** | | **Parameters** | | **PDF** |
| DL | UL | DL | UL[[5]](#footnote-5) |
| **Initial packet size**  **(Byte)** | Uniform | Uniform | a=0,  b=80 | a=0,  b=80 |  |
| **Packet arrival time (ms)** | Largest Extreme Value | Largest Extreme Value | a=28,  b=4.2 | a=22,  b=3.4 |  |
| **Packet size (Byte)** | Largest Extreme Value | Largest Extreme Value | a=210,  b=35 | a=92,  b=38 |  |
| **End packet (Byte)** | Uniform | Uniform | a=1400,  b=1500 | a=500,  b=600 |  |

Table 4‑4 Traffic model for frame lockstep sync

### Requirement metrics

Real-time mobile game request instant feedback of in-game operation from multiple players in the same battle room, and lagging will cause picture not fluent and impact user experience negatively.

Moreover, the time difference that users get the response will cause unfairness because the response is not synchronous.

A common observation is, during a period time, although average latency is low, worst case latency (spike) can be several times higher than the average. Jitter will cause spike in latency, when latency is too high which exceeds threshold and then cause packet loss, lagging is expected.

The relationship between game experience and end to end latency is as below table.



Table 4‑5 Relationship between game experience and latency

The average latency from client to AP is 7ms, however user experience needs to be considered over the complete path from client to game server.

The RTT latency breakdown from client to game server is as below.



Figure 4‑7 The latency breakdown from client to game server

Specifications from client to AP in below table (based on experience value)

| Specification | Value |
| --- | --- |
| Intra BSS latency | <5ms |
| Packet loss | <0.1% |
| Jitter variance | *σ* <2ms |
| High latency packet (>30ms) | <1% |

Table 4‑6 Specifications from client to AP

## **Wireless Console Gaming**

### Introduction

Console gaming which is also known as the video game is played on devices made especially for gaming. Starting with the Arcade games, it transformed to console games which began with cartridge storages which has changed to disks and inbuilt hard-drive storages. These advancements have now optimized the consoles to play online multiplayer gaming which involves instant response for your actions in the game and enables interactive gameplay.

Console gaming involves various genres of games, but the main genre we are focusing on is latency sensitive online FPS (First Person Shooter) games. This is an interactive gaming experience with real-time feedback and response. A Synchronized game state is established among players in the same match to get the best performance. FPS gaming is centered around guns and other weapon combats in the first-person point of view with which the player sees the action through the eyes of the player character.

In multiplayer FPS game, more than one person can play in the same game environment at the same time either locally or over the internet. Multiplayer games allow players interact with other individuals in partnership, competition or rivalry, providing them with social communication absent from single-player games. In multiplayer games, players may compete against two or more human contestants, work cooperatively with a human partner to achieve a common goal, supervise other players' activity, co-op. Multiplayer games typically require players to share the resources of a single game system or use networking technology to play together over a greater distance.

Playing online on a console has 2 types of internet connectivity, which is either wired or Wi-Fi. Most of the gaming consoles today support Wi-Fi 5. But Wi-Fi has an especially bad reputation among the gaming community. The main reasons are high latency, lag spikes and jitter. According to a top-selling online console game in the US up to 79% of FPS players are using Wi-Fi connected consoles.



Figure 4‑8 FPS gaming

### Network architecture for an FPS game

**Dedicated Server:**

In this model, many servers are in different geographical locations depending on the game traffic in the region. The main purpose of Dedicated server is to host the game and to perform matchmaking. Fig 4.9 is representation of Dedicated server architecture.

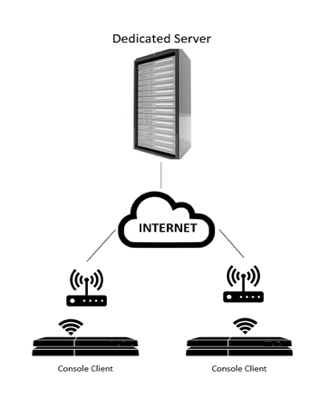


Figure 4‑9 FPS dedicated server architecture

**Local Server:**

This architecture elects a gaming console to host the game which handles all the network traffic to the players connected. The game statistics are still managed by the centralized servers. This method helps to reduce the number of dedicated server infrastructure where there are not many players participating in the game. Fig, 4.9 is a representation if Local Server architecture.

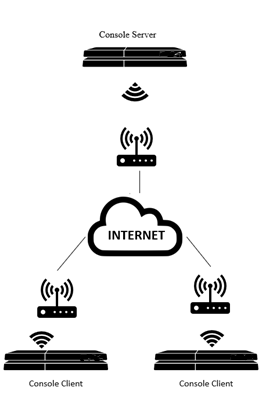


Figure 4‑10 FPS dedicated server architecture

### Packet flow of an FPS console game

Packet protocol: Uses UDP Data packets for gameplay updates to and from the server and TCP for matchmaking.

Packet size: 100-200 bytes depending on the size of data sent.

Packet flow rate: Flow rate will vary from game to game. In our testing we observed approximately 60 packets/sec from a console to server and 15 packets/sec from a server to console depending on the pace of gameplay.

Bandwidth between console and Server: 0.2 - 1 Mbps.

Game data: Carries users’ instructions during uplink and result of those instructions in a downlink.

### Traffic model

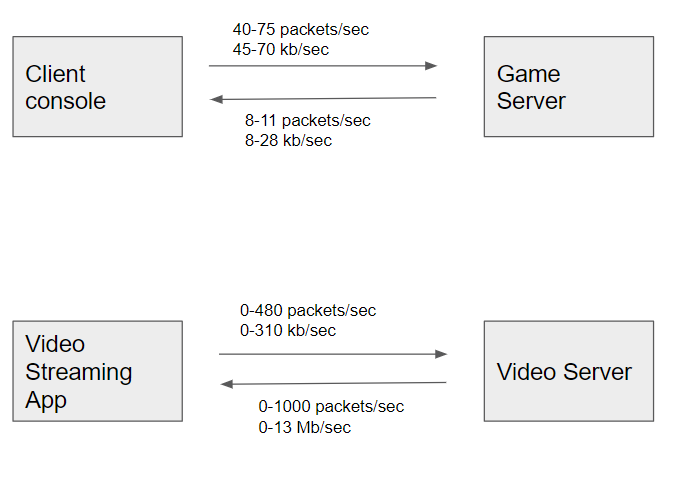


Figure 4‑11 FPS traffic model

Comparing the traffic flow with a video streaming application, FPS games have less packets/sec and bandwidth occupied. Whereas video streaming chokes the whole network with packet blasts. When both are trying to compete, Video packets dominate mostly because of the packet size and bandwidth occupied. This leads many issues with other types of connection in the network.

### Problem statement

Game publishers have reported that up to 79% of their players use Wi-Fi connections. With regards to the Wi-Fi problems encountered in console games, consoles have to contend the wireless channel medium with other devices, who could be labelled with higher priority and intend to occupy large frequency bandwidth. As the number of video streams increasing in a network, the contention becomes more intense. To keep latency low, the total throughput of the Channel is decreased. Voice and Video streams have the priority over data due to Enhanced Distributed Channel Access (EDCA). EDCA/WMM attempts to contain the higher priority streams and offers protection to streams already in progress. These factors impact the smoothness of Best-effort UDP packet blasts during an intense gameplay.

Apart from these prioritization issues, game users also encounter problems like:

* Lag spike during gameplay can last for 1-3 seconds
* There is limited bandwidth available due to high bandwidth streaming devices
* Channel interference could be caused by other networks
* Packet loss is usually compensated by dead reckoning which can cause Rubber Banding
* One might get kicked out of the public match mostly due to high ping to game server.

Regarding game users who face these problems, they are mostly gamers using Wi-Fi, while gamers who have high requirements to network stability or compete on eSports often have to choose wired connections since the current Wi-Fi cannot meet the high stability network requirements for eSports.

As eSports is emerging as a major worldwide online gaming competitions, the random worst case latency become a more and more prominent problem, since players may claim lagging and pause the game frequently. And the interruptions of game impact both players competition moral and spectators’ wathching experience negatively.

## **Cloud Gaming**

### Introduction

Cloud gaming is another type of video game potentially played on light-weight devices at users premise. Unlike other gaming hardware, user devices do not need to render pictures or video. Instead, they are rendered at the cloud server. The picture/video generated at the cloud server are streamed to the user devices, and the user devices just display the received picture/video on its display. The cloud game can accommodate and connect multiple players in a single game session just as mobile gaming scenario.

The cloud gaming requires low latency capability as the user commands in a game session need to be sent back to the cloud server, the cloud server would update game context depending on the received commands, and the cloud server would render the picture/video to be displayed at user devices and stream the picture/video content to the user devices. This cycle needs to be short enough so users do not feel lagging responses.

With cloud gaming experience, users can play large amount of game titles as they will be provided and hosted by the cloud server. Users can pick up game title from the library on the cloud server. Another benefit of the cloud gaming is that the user device could be light-weight in terms of hardware footprint. The user devices only need to decode and display received picture/video content. This way, users can enjoy realistic and immersive game experience without requiring heavy computation at user devices. The light-weight user device leads to lower cost and longer battery life, which could motivate gamers to play on the games more.

### Network architecture

The system configuration is very similar to Mobile Gaming scenario. [15] describes sample configuration and an example of the existing cloud gaming service

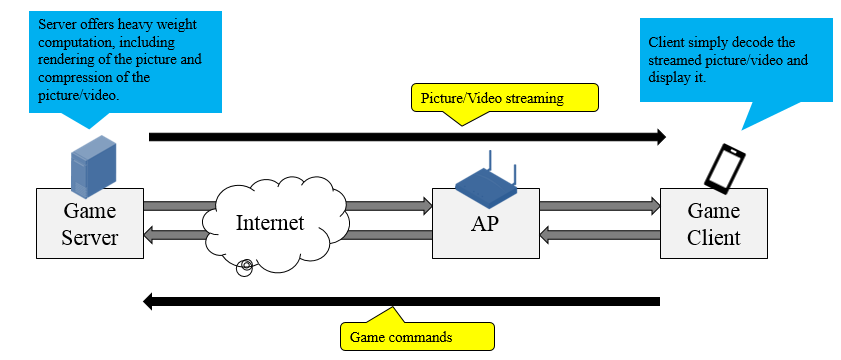


Figure 4‑8 Cloud gaming network architecture

**Below are explanations of the components of cloud gaming network.**

**Game server:** The Game Server is located in the cloud and hosts gaming service. It operates as if it is a remote game console. It runs game session, creates picture/video content, compresses the content, and transmits the content to the Game Client located on the user premise. Also, the Game Server receives game player’s commands from the Game Client, updates game context depending on the received commands, and renders the picture/video to be displayed at user devices.

**Internet:** The data generated by the Game Server will be transferred to user premise through the Internet. Similarly, commands generated by the Game Client will be transferred to the Game Sever through the Internet.

**AP (access point):** The Game Client establish an 802.11 link with AP to communicate with the Game Server. AP manages the 802.11 wireless link providing low latency access.

**Game Client:** The Game Client consists of display device, game controller and commander, picture/video coded, and WLAN modem. The Game Client displays received picture/video data to its display, accepts commands from game controller, and sends the game command to the Game Server. The WLAN modem establishes an 802.11 link with AP and provides the low latency WLAN communication.

### Traffic characteristic

The required bandwidth of the cloud gaming is asymmetric for forward link and reverse link. The reverse link bandwidth is up to 100kbps, whereas the forward link bandwidth varies depending on the picture/video quality. An existing cloud gaming service requires 5Mbps at minimal with suggested bandwidth of 12Mbps at the application layer. However, display generation is migrating from HD to UHD (4K) and user’s expectation to the picture quality continues to get higher, it is suggested that 802.11 community would consider near future requirements rather than minimal and existing requirements as of today.

**Summary of the cloud gaming traffic profiles and requirements[[6]](#footnote-6)**

|  |  |  |
| --- | --- | --- |
| Latency | | |
| End-to-End target latency | <80msec | |
| Intra-BSS target latency | <10msec | |
| Jitter | | |
| End-to-End target jitter variance | *σ <* 6msec | |
| Intra-BSS target jitter variance | *σ <* 2msec | |
| Forward link bandwidth | | |
|  | @ MAC-SAP | @ Application |
| 4K H.265 60FPS for PC or TV | 128Mbps | 128Mbps |
| 4K H.264 60FPS for PC or TV | 160Mbps | 160Mbps |
| 1080P H.265 60FPS for Mobile Device | 32Mbps | 32Mbps |
| 1080P H.264 60FPS for Mobile Device | 40Mbps | 40Mbps |
| Reverse link bandwidth | | |
|  | @ MAC-SAP | @ Application |
| Game commands | ~112 Kbps | ~100 Kbps |
| Reliability (Error rate measured at the MAC-SAP) | | |
| Forward link | BLER < 10e-9 | |
| Reverse link | BLER < 10e-3 | |

Table 4‑8 Cloud gaming traffic profile

### Problem statement

Today, we see many software applications are migrating from package distribution to cloud based subscription model. Game application is not an exception. Cloud gaming could bring benefits to game users and game industry ecosystem. However, the current communication infrastructure and the wireless interface to the user client device are often not designed to meet the cloud gaming requirements. Particularly, 802.11 does not provide a tool to control worst case latency and jitter. It is desirable that the 802.11 network or devices would support features to assure bandwidth and robustness together with latency and jitter required from application, even when the 802.11 network is offering services to other devices. To ensure end to end latency and provide satisfying gaming experience, some cloud games are only accessible to domestic users.

If 802.11 network could provide tools that can control worst case latency and jitter while providing high bandwidth video streaming services, it would become a strong driving force to move game system architecture to a next stage.

## **Industrial Systems**

Industrial systems include a wide range of applications: process monitoring, automation, control systems, human-machine-interfaces (HMI), Automated Guided Vehicles (AGVs), robotics and AR/VR. Recently, several standard developing organizations have published detailed description of industrial application and their requirements, such as:

* **IEEE 802.1 NENDICA Report Wired/Wireless Use Cases and Communication Requirements for Flexible Factories IoT Bridged Network (**[802.1-18-0025-06-ICne](https://mentor.ieee.org/802.1/dcn/18/1-18-0025-06-ICne.pdf));
* **IEC/IEEE 60802** [**Use Cases for Industrial Automation**](http://grouper.ieee.org/groups/802/1/files/public/docs2018/60802-industrial-use-cases-0818-v11.pdf) **(TSN-IA Profile for Industrial Automation);**
* **3GPP** [**TR 22.804**](http://www.tech-invite.com/3m22/tinv-3gpp-22-804.html) **Technical Specification Group Services and System Aspects; Study on Communication for Automation in Vertical Domains.**

The purpose of this document is not to repeat the detailed application descriptions, which can be found in above references. Instead, the focus is to summarize the challenges and requirements of real-time and time-sensitive applications that are most relevant to 802.11.

Many industrial applications can be considered delay-tolerant (e.g. process monitoring, industrial sensor networks, etc.) with latency requirements in the order of 100msec or more. Such applications may be served by existing wireless standards and are not considered in this report. This report focuses only on time-sensitive and real-time applications.

### Real-time industrial use case examples

**Warehousing Logistics**

* **Introduction**

The warehouse often forms the heart of an organization’s operations. It's the place through which all materials, products and merchandise flow to their final destinations. And among the many industries that rely on warehousing, retailers are leading the growing trend to embrace wireless technologies to transform their warehouse operations and gain ground on the competition.

Many retailers, including giants such as Amazon, Home Depot, and Wal-Mart, stock inventory and process orders from large warehouses, some occupying more than one million square feet. From receiving to inventory to outbound distribution, these warehouses increasingly rely on wireless technologies to streamline the logistics of their operations and enhance overall productivity by fulfilling orders more accurately and quickly than before.

Reliance on Wi-Fi for these critical warehouse functions makes seamless, uninterrupted connectivity vital. While there are many challenges to providing a high-performing wireless infrastructure, tools exist to help ensure networks are capable of managing these demands and responding with quick troubleshooting when problems arise.

* **Architecture of warehouse topology**

Real-time Info. Exchange

Warehousing server

Central controller

Warehousing mobile device 2

Warehousing mobile device N

Warehousing mobile device 1

**…**

Figure 4‑12 Warehouse automation network architecture

1. Real-time information: Location info. / Fault feedback / Collaborative strategy etc.
2. Warehousing mobile device: Freight car / Positioning tag / Asset monitoring equipment

* **Problem statement**

With huge areas filled with metal racks, a myriad handheld devices, moving vehicles, dynamic environmental conditions and no pervasive Ethernet network, connectivity is an enormous challenge.

### Summary of real-time industrial use cases and classes of service

The table below summarizes use cases and requirements grouped in three classes of services in industrial systems. The latency bound is defined as the worst-case one-way latency measured at the application layer. The reliability is defined as the percentage of packets expected to be received within the latency bound.

|  |  |  |  |
| --- | --- | --- | --- |
| Applications and Requirements | **Class A** | **Class B** | **Class C** |
| Applications | Interactive video, soft-real-time control, mobile robotics, Automated Guided Vehicles (AGV) | AR/VR, remote HMI, hard-real-time cyclic control, machine tools, production lines | Hard-real-time isochronous control, motion control, printing, packaging |
| Time synchronization | 10-1µs | ~1 µs | ~1 µs |
| Latency bound | 50 -10 ms | 10 – 1 ms | 1ms – 250 µs |
| Reliability | 99% - 99.9% | 99.9% - 99.99% | >99.999% |
| Throughput | High (video)  Low (control, robotics, AGVs) | High (VR)  Moderate-Low (controls/automation/AR) | Moderate-Low |

Table 4‑7 Factory automation use cases summary

### Traffic profiles

The communication model of a closed loop control system, described in the figure below, is applicable to many industrial applications. Therefore, its traffic profile can be considered a representative model for many industrial systems.

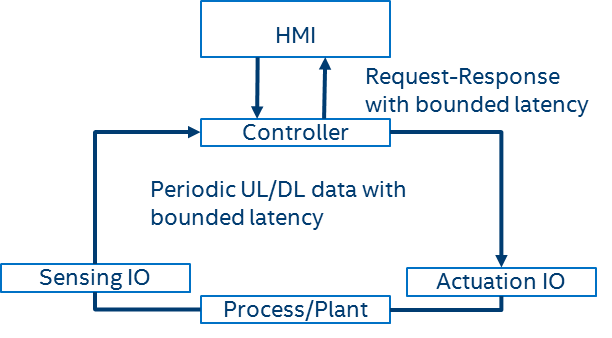


Figure 4‑13 Communication model of a closed loop control system

The main traffic patterns generated by a control system include:

* **Periodic data**: this is the typical pattern generated by remote IO (sensing and actuation) within a closed loop control system. Remote HMI devices can also operate as a remote IO device and generate periodic data. The data sizes are usually fixed and relatively small (e.g. typical from 30 to 300 bytes). The periodic data, which may be exchanged in uplink (sensor to controller) or downlink (controller to actuator) direction, is expected to be delivery within a given latency bound.
* **Event**-based: In contrast to periodic data, a second pattern in closed-loop application relates to event-based traffic patterns, for instance when a certain threshold is hit at a sensor, triggering a reporting. Event-based traffic patterns arise less often in comparison to periodic data.
* **Request/Response**: these are asynchronous messages. The request is typically generated by the HMI device and response is generated by the controller device. This can also be seen as a control loop, but it is not a cyclic one as in the remote I/O (sensor, controller, and actuation) case. The bounded latency in this case applies to the round-trip latency between the times when the request is generated, and the response is received.

More details about industrial traffic profiles and requirements can be found in a recent white paper published by the Industrial Internet Consortium (IIC) [14].

Real-time applications such as video streaming, AR, and VR are also used in industrial scenarios, but they have no industrial-specific requirement. Therefore, they are not described in this section.

### Problem statement

The bounded latency and reliability for periodic and request/response data used in the control cycle is the most challenging requirement for the communication links. The required values depend on the type of control system, but the target performance can be grouped in the classes of service described in the previous section.

It has been shown that average latency in 802.11 networks can be very low, but worst-case latency and jitter can be very high, depending on the traffic and network congestion. Solutions to address congestion and reduce randomness in accessing the medium should be considered to enable more predictable/control of latency. It is also important to note that for real-time industrial applications, providing a predictable worst-case latency is the main problem. As described in the Table 4‑7 above, many applications (e.g. classes A and B) do not necessarily need extreme low latency and but they all need more predictable worst-case (or bounded) latency. There are applications (e.g. Class C) that may require such bounded latency to be very low, but they are unlikely to be the first class of applications that will be enabled over wireless, mainly due to safety concerns. But as wireless capabilities improve, such extreme low latency application may also be supported.

Providing high reliability has not been a focus for 802.11 networks. Operation in unlicensed bands and lack of coordination between BSSs are some of the factors that impacts reliability.

One important aspect to consider is that wireless networks within industrial environments are usually managed, which can help reduce interference from other (un-managed) networks. Furthermore, the traffic configurations and requirements are either pre-set or negotiated during the device association/network entry, and do not change during the system operation. Therefore, it is feasible to provision/configure the network to improve reliability and provide a more predictable performance.

## **Real-time video**

### Introduction

Today, many devices handle video streaming via 802.11 wireless LAN. Most of them are not latency sensitive. However, some video applications require low latency capability, when the application provides interactive play. Example of such applications includes VR/AR, and video cable replacement [3].

In many of these cases, the latency requirements are derived from the video frame rate. As of today, 60Hz framerate is commonly used, i.e., 16.7msec per frame. However, it is possible that the video rendering system would migrate to high frame rate solution, i.e., 120Hz which resulting in 8.33 msec per frame, etc., in the future.

To accommodate end-end signal processing in a video frame, the signal processing delay plus transmission latency need to be less than 16.7 msec. For these applications, ideally, 10[msec] one-way or roundtrip delay should be considered as a targeted specification for the radio link transmission, allowing 6.7msec for other signal processing including, but not limited to, video signal encoding (compression), in-device frame forwarding, video signal decoding (decompression), etc.

When the video frame rate of 120 Hz (8.33msec per frame) is used, ideally, 3 msec delay should be considered as a target for the radio link transmission, allowing 5.33 msec for other signal processing.

The following figure depicts the difference between a video application which does not require low latency capability and a video application which requires low latency capability. In general, low latency requirements arise when there is a control loop in the system.



Figure 4‑14 Difference between buffered video and live video

### Interactive video rationale

The interactive video traffic tends to be high bandwidth while requiring low latency capability. These requirements are derived from video compression techniques. There are two types of video compression approach: intra-frame and inter-frame. Intra-frame compression is an approach to compress a video frame only referring to this particular frame. In other words, with this approach, a video picture can be decoded without relying on other video frame that are sent in other time. Contrarily, inter-frame compression leverages video frame information sent in other time. Typically, this approach excels in terms of compression efficiency as we can leverage video picture correlation among frames to reduce entropy of the video information. However, inter-frame compression approach is applicable only when the video contents are bufferable. In case the video contents are changing per external information in real time fashion, we cannot buffer the video contents and the content changes need to happen in real time manner. As a result, for these applications, intra-frame compression approach is taken, and required bandwidth tends to be higher while requiring low latency capability.

The following data flow explains how a VR application works assuming that 1) the video content is generated at a game console, 2) the content is transferred over wireless link (802.11 technology) to a renderer, VR headset, and the video is displayed at the VR headset.

1. Game video contents are generated at a game console. The video picture is updated per 16.7 msec at video frame rate of 60Hz. When a video picture is determined, the video console compress the picture using intra-frame compression.
2. The game console transmits the compressed video picture to a VR headset over the 802.11 link. The video picture is packetized to fit in 802.11 frames. The received frame might be corrupted by error in physical layer, but MAC layer retransmission will salvage such an error.
3. The entire video picture arrives at the VR headset. Then it starts to decompress the video frame to display the content.
4. User of the VR headset may react to the displayed content and move his/her head. This movement is captured by the head tracker mounted on the VR headset. The VR headset sends the head tracker information back to the game console over the 802.11 link.
5. Once the game console received the head tracker information, it starts to determine the video contents on the next video frame. The next video frame needs to be adjusted based on the orientation of the user, so the user can enjoy immersive experience. After the content on the next video frame is determined, the game console starts to compress the picture, and then send it to the VR headset.

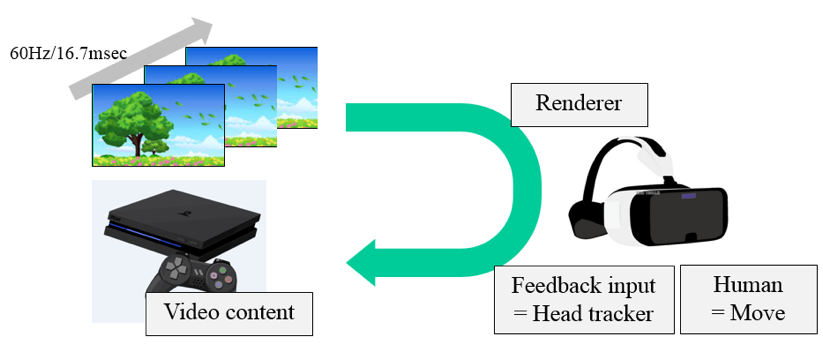


Figure 4‑15 Video content render process

If the video frame rate of 60Hz is used in the system, above transmission cycle shall be completed within 16.7msec ideally. According to a research result reported in [4], overall round trip delay of the control loop should be completed in 11.1msec, video update rate of 90Hz, to maximize user experience with VR headset, i.e., to prevent VR nausea.

### Traffic profiles and requirements

The required bandwidth of the real-time video varies depending on the application. As the real-time video application market is expected to be ramping up and the technology would be evolving rapidly, it is suggested that 802.11 community would consider near future requirements rather than minimal and/or existing requirements as of today.

**Summary of the real-time video traffic profiles and requirements**

|  |  |
| --- | --- |
| Intra-BSS target latency[[7]](#footnote-7) | 3~10msec |
| Forward link bandwidth | 0.1~20 Gbps |
| Reverse link bandwidth | 10~100 Kbps |
| Reliability | BLER < 10e-9 |

Table 4‑8 Real-time video traffic profile

### Problem statement

Together with video application traffic study, low latency requirements have been recognized in previous 802.11 WG projects defining high bandwidth MAC/PHY technologies. However, as pointed out in [3], latency requirements are not treated as a primary goal of these projects and there is no standardized solution or guideline how we can design and deploy an 802.11 network satisfying real-time video latency requirements, as of today.

Thanks to high bandwidth wireless transmission capability, the video streaming application became almost ubiquitous. Now is a time to migrate to the new era of video application. Immersive & interactive video is attracting people these days. It would be desirable that 802.11 network would be able to provide a transmission capability meeting the requirements.

## **Drone Control**

### Introduction

Drone is an aircraft without a human pilot aboard. Drones are rapidly popularized and utilized for a wide array of uses. Gartner mentions that worldwide production of drones neared 3 million units in 2017 [8]. Wi-Fi has an important role to control drones by providing following functions.

* **Tele control**

Controlling motions and functions of the drone. A few Kbps of data rate is required.

* **Data transmission**

Monitoring information from sensors in a drone or information of the status of the drone itself. A few Kbps~Mbps of data rate is required.

* **Picture / video transfer**

Transferring recorded pictures or videos by the drone. More than tens of Mbps of data rate is required.

### Use cases

Drones are widely used in various businesses and use cases. For example, rapid inventory management in a warehouse will be realized by travelling of drones with RFID reader or barcode scanner [9]. Drones are used for not only industrial use but also entertainment. Nowadays, drones are the important device for e-Sports like racing [10] or digital signage [11]. A picture shown below indicates that multiple drones perform public viewing as digital signage by running on ground in one instance.



Figure 4‑16 Example of use case for digital signage. [12]

Table 4-9 indicates drone use cases and their range and latency requirements [13]. Though there are a lot of different use cases for drone control, target communication range will depend on the wireless connectivity technology and regulatory rules for each country. Given the Wi-Fi capabilities, only use cases with target communication ranges in the order of 300 - 100 meter should be considered in the scope of 802.11 and the RTA TIG.

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Use cases | Required distance of communication | Required maximum delay |
| Industry | Inspection from the air | ~ 300 m | ~ 100 ms |
| Route guiding | ~ 1 km | ~ 100 ms |
| Logistics (outdoors) | ~ 5 km | ~ 100 ms |
| Logistics (indoor) | ~ 100 m | ~ 100 ms |
| Security patrol | ~ 1 km | ~ 100 ms |
| Entertainment | Gaming device | TBD | TBD |
| Digital signage | ~ 100 m | TBD |
| Real-time distribution of images | ~ 100 m | TBD |

Table 4‑9 Use case and requirements for drones[13]

### Architecture

There are several architectures for drone control described below.

* **Standalone (single drone)**

Most of commercial drones adopt this architecture. In this architecture, a drone plays a roll of the AP and a controller (e.g. a smartphone or a console) becomes the STA. Although this architecture enables easy control of the drone, only one drone can be managed per a single controller. Fig.4-17 illustrates this architecture.

* **Standalone (multiple drones)**

To control multiple drones at once, the drones are desirable to be the STAs and a controller should be the AP as indicated in Fig.4-18. This architecture will be deployed for digital signage such as public viewing and the number of STAs might be more than ten to one hundred. Due to sharing a same channel with the multiple drones and the AP, functions that ensure reliability will be needed in this architecture.

* **Network control**

In this case, the drone has a possibility to become the AP or the STA. However, it is desirable that the drone plays a roll of the AP. Fig. 4-19 shows this architecture. To control drones over the network remotely, network delay should be considered. The example of network delay estimation is shown in Fig. 4-20.

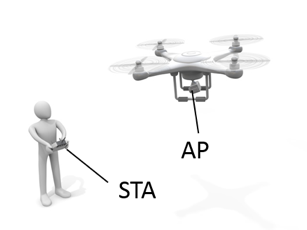


Figure 4‑17 Standalone (single drone)

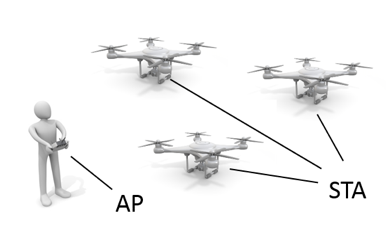


Figure 4‑18 Standalone (multiple drones)

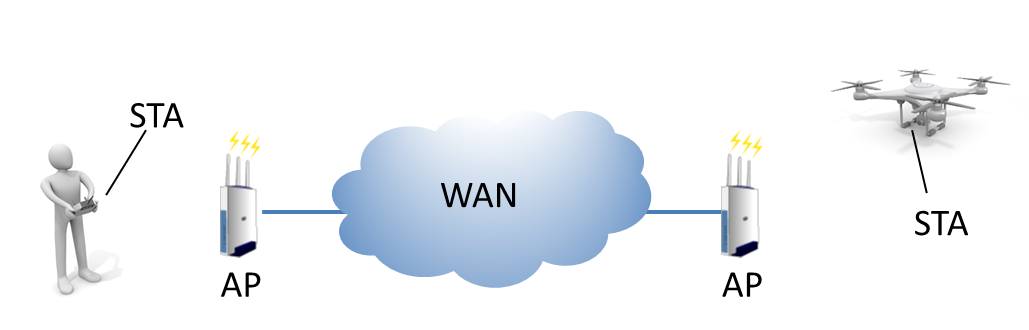


Figure 4‑19 Network control

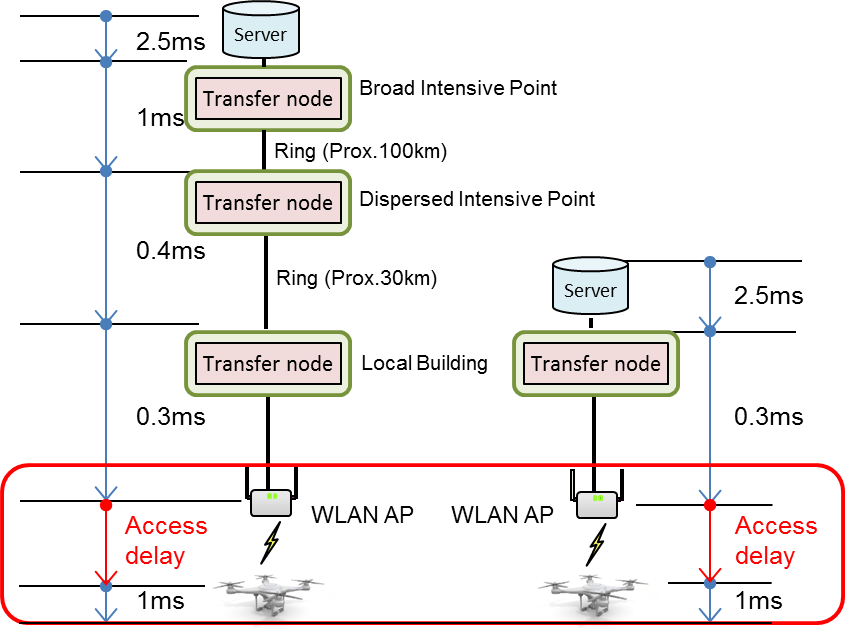


Figure 4‑20Example of network delay estimation

### Problem statement

Almost commercial drones employ Wi-Fi for their communication method and Wi-Fi controls the three types of communication functions described in 4.5.1. Especially, tele control is important for stable operation for drone control. Although tele control does not require high data rate and strict latency requirements shown in Table 5-2 compared to past IEEE 802.11 standards' CSD, loss of tele control might cause unintended motion, crash of drones or other bad incidents. To prevent such loss of tele control, increased reliability is required and the following issues should be considered and verified to support the drone RTA use case:

* **Mixture of tele control and video transfer**

If simultaneous transmission of tele control and video transfer that requires high data rate is executed, quality of tele control might be degraded due to lack of wireless resource.

* **Movement of drones**

Drones are moving STAs in partially broad area. Besides, rapid change of propagation, interference of other BSS and dynamic change of signal strength might be a factor for degrading quality of tele control.

To solve these issues, further investigations should be required such as priority control, separation of control messages and data messages or protection mechanism for control messages.

# Potential technical features analysis

Although current 802.11/Wi-Fi solutions can provide low average latency in uncongested environments, the worst-case latency can vary significantly, impacting the performance or limiting the usage of many real-time and time-sensitive applications over Wi-Fi.

There is a need to explore solutions to better control worst-case latency and jitter and provide more stable/reliable performance. The next sections discuss potential capabilities that can be used to control congestion, which is a major cause of high/variable latency, and also improve reliability.

## **Time sensitive networking extensions to 802.11**

The IEEE 802.1 Time Sensitive Networking (TSN) Task Group has been developing standards for creating distributed, synchronized, real-time systems. Some of the TSN capabilities and standards include: stream reservation (802.1Qat), time synchronization (802.1AS), time-aware shaping (802.1Qbv), frame replication and elimination (FRE defined in P802.1CB), and many others. Although many of the TSN capabilities have been designed assuming Ethernet as transport media, some TSN capabilities have already been extended to 802.11, such as time synchronization (timing measurement capability in 802.11-2012), stream reservation (SRP over 802.11 for AV defined in 802.11aa), 802.11 links in an 802.1Q network (802.11ak). The 802.1 features defined so far over 802.11 do not address the worst-latency and reliability requirements. Additional extensions, such as time-aware shaping and redundancy (FRE) can address the congestion and reliability problems identified earlier. Some of the existing capabilities and potential extensions are described in the following sections.

## **Priority tagging (Stream identification)**

One basic requirement to provide better QoS for time-sensitive traffic is the capability to identify and differentiate time-sensitive packets from other (e.g. best-effort) traffic.

The IEEE 802.1Q standard defines mechanisms used to identify and differentiate time-sensitive streams from other types of traffic. 802.1Q specifies a *Virtual Local Area Network (VLAN) tag* field, which is added to the header of the layer-2 Ethernet frames. Furthermore, 802.1Q defines traffic classes, and allows up to either traffic classes per Ethernet port, with each traffic class associated with a dedicated queue. The *Priority Code Point (PCP)* field of the VLAN Tag determines the traffic class for a given Ethernet frame. Specifically, bridges are configured to map PCP values to traffic classes, and as frames arrive, the bridge directs them to the appropriate queue based on the value of the frame’s PCP field and the configured PCP-to-traffic class mapping. The 802.1Q stream identification mechanisms have already been defined over Ethernet and 802.11.

Given that practical 802.11 MAC implementations are based on the Enhanced Distributed Channel Access (EDCA), time-sensitive traffic must be mapped to one of the four EDCA access categories. It is possible to tag time-sensitive traffic as AC\_VO or AC\_VI and leverage to EDCA parameters to reduce channel access latency. Although EDCA priorities can help reduce latency, the worst-case latency can still vary significantly, especially in case of network congestion or competition multiple applications mapped to the same access category (e.g. voice and gaming, gaming and video).

Moreover, mechanisms that not only identify time-sensitive frames at the 802.11 MAC, but also prioritize such frames in the channel access process are needed to address the worst-case latency issues.

## **Time-aware shaping (802.1Qbv) over 802.11**

The IEEE 802.1Qbv standard defines a Time-Aware Shaping capability consisting of gates to control the opening/closing of queues that share a common egress port within an Ethernet switch. A scheduler defines the times when each queue opens or close, therefore eliminating congestion and ensuring that frames are delivered within the expected latency bounds. This solution can be extended to 802.11 to avoid congestion in a BSS and/or across managed BSSs.

In an 802.11 BSS, the queues at different contending STAs need to be controlled to avoid congestion as illustrated in the figure below. Once a time-sensitive queue is open, a transmission selection module selects a data frame to send and delivers it to the MAC layer. One potential implementation is to use the time-division to separate time-sensitive from other traffic, as shown in Figure 5‑2. The time-aware shaping capability is independent of the channel access mode and could be used with different options (e.g. EDCA, trigger-based access with 802.11ax, etc.). It is important to note that the more efficient and predictable the access, the lower the latency bounds that could be supported by the network.

The extension of the Qbv traffic shaping concept over 802.11 would require modification in the 802.11 MAC layer to enable distribution of the Qbv schedule between managed stations (STAs) and also to ensure the 802.11 EDCA queues are controlled to meet the Qbv schedule.

**Example of time-aware shaping capability to resolve contention between managed STAs is shown in** Figure 5‑1 **below.**

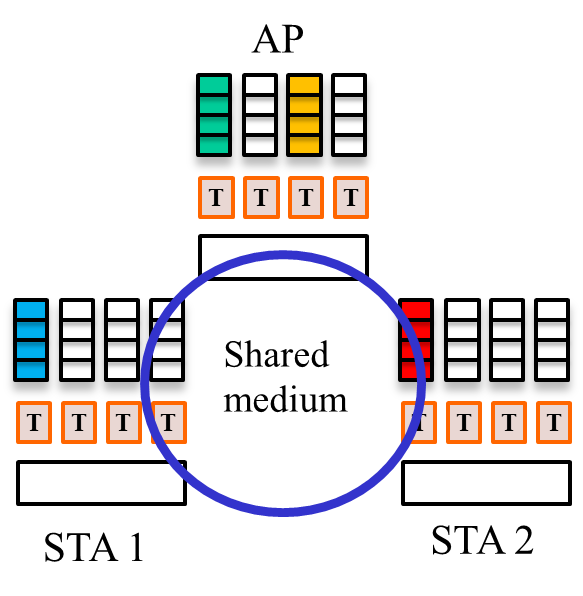


Figure 5‑1 STAs share the medium and queues are managed according to a 802.1Qbv schedule

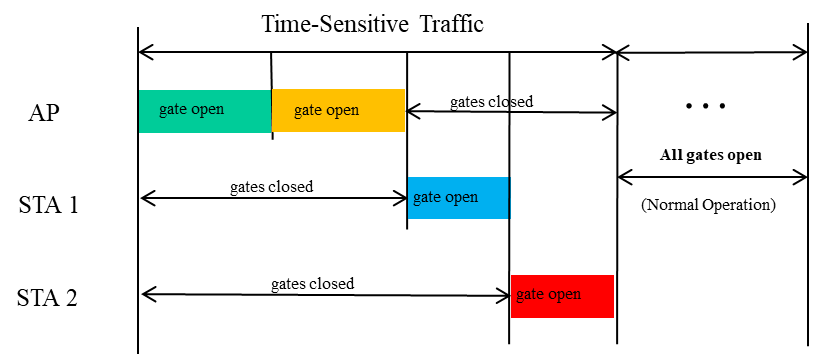


Figure 5‑2 Using TSN to resolve contention

## **Dual/multiple link**

Due to the competitions and interference are common and hardly in control under wireless network, in order to improve the latency stability, dual/multiple link proposal is brought up to address this issue.

Dual/multiple link proposal applies to STA and AP which can work on at least two bands,

There are two implementation modes for dual/multiple link applications.

* + Duplicate Mode：When dual link feature is enabled, each packet in one certain traffic stream will be duplicated at one end and de-duplicated on the other end. The transmission side will tag duplicated packet with a unique sequence number which will be identified at the other end for de-duplication purposes. At the receiver side, de-duplication will be done before the packet is processed. Only the first arrived packet will be accepted and trigger ACK and the second arrived packet will be dropped.

Frame Replication and Elimination (FRE) capability has also been defined for Ethernet by the 802.1CB standard and could be extended to 802.11. The dual link capability can be used in combination with other capabilities to control latency to improve the overall reliability of the system.

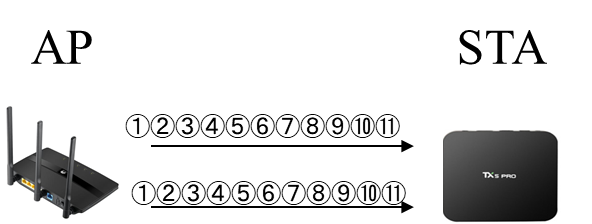


Figure 5‑3 Diagram of Duplicate Mode

* + Joint Mode：Different part of the packets are transmitted in different streams on different links. The transmission side will tag part of packet with a unique sequence number which will be identified at the other end for combining purposes. If the combine processing time is short enough this mode will effectively reduce the latency. Monitoring the physical available bandwidth in the Wi-Fi link can be used to determine which mode to be used for delivering the data.

Since dual/multiple bands work independently at the same time, this feature also can be used for enlarging the throughput.

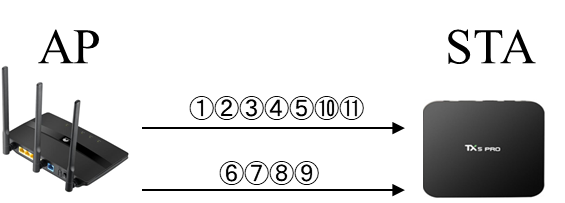


Figure 5‑4 Diagram of Joint Mode

Base on above modes, there are still some problems to be considered, like:

* Traffic load balance: Since the actual bandwidth of the Wi-Fi links are not stable enough, a proper mechanism should be considered to reasonably assign the proper amount of traffic dynamically according to the actual bandwidth status in the two bands: 5 GHz and 2.4 GHz.
* Packet alignment: Joint mode will cause packets out of order. Out of order packets should be differentiated from losses in the MAC layer (losses will cause extra delay). Dealing with out of order packets requires extra buffer and capabilities.

## **Admission Control**

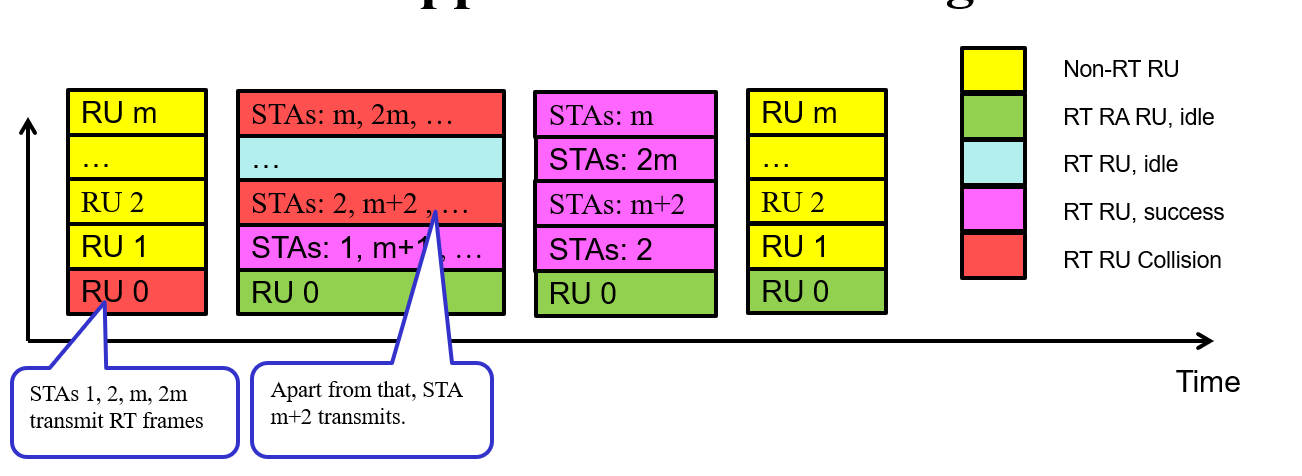
Congestion is a major cause of high/variable latency and it is well known that if the traffic load exceeds the capability of the link, QoS will be degraded. QoS degradation maybe acceptable for many applications today, however emerging time-sensitive applications are highly susceptible to increased latency and jitter. In order to better control QoS, it is important to control access to the network resources. Admission control capabilities have already been defined in the 802.11 MAC. Admission control is also important to ensure that solution to control latency and reliability can be effective.

Admission control can be used within a BSS and across overlapping BSSs in a managed network. Obviously, there is always the possibility of non-managed overlapping BSSs due the unlicensed spectrum used by 802.11. However, there are many scenarios where the network can be managed and solutions to control latency/reliability can be very effective in combination with admission control policies. Furthermore, as new spectrum become available (e.g. 6 to 7 GHz), there will be opportunity to apply more efficient medium access techniques, such as 802.11ax and beyond, and avoid contention with legacy 802.11 systems.

## **Channel resource reservation for real-time traffic with OFDMA**

Delay requirements can be violated because of the long duration of ongoing transmissions and delayed channel access. However, the real-time packets need an opportunity to be transmitted within their delay budget even in deployment with a huge amount of non-real-time traffic. For downlink, this can be done if the duration of frames/TXOP is limited. For uplink, the AP can adaptively reserve a part of the available spectrum for real-time transmissions. This can be done with OFDMA enabled by 802.11ax as discussed in [16].

Specifically, in case of intensive uplink traffic, the AP could reserve OFDMA Random Access (RA) RUs for real-time traffic only, as illuatrated in Figure 5‑5. Currently, 802.11ax allows only allocating RA RUs separately for associated and for non-associated STAs. If more than one STAs transmit in an RA RU dedicated for the real-time traffic, a collision occurs. Although the exact STAs which participate in the collision are unknown, the AP can resolve the collision if it has knowledge of the traffic profile of the contending STAs, which could be done as part of an admission control procedure.. The AP can allocate a dedicated RU for each associated STA that potentially has real-time traffic. If the number of such STAs is too high, but only a few of them have traffic, it makes sense to allocate one RU for two or more STA with a condition that only STAs which have data for transmission transmit in these RU. Currently, this capability is not supported in 11ax. If a collision happens in such an RU, it can be easily resolved next time, since only a small set of STAs could transmit in this RU. It is important to note that latency performance may be degraded if the level of collisions is high.

Figure 5‑5 Random Access RU reservation for real-time STAs

# Recommendations

## Real-time applications requirements summary

The RTA TIG discussed multiple real-time applications in several domains (gaming, industrial automation, drone control, etc.) and their requirements are summarized in Table 7‑1. Real-time applications have been evolving, so do their communication requirements. While voice and video accounted for most of the real-time traffic in the past, new and emerging applications such as real-time gaming, AR/VR, robotics and industrial automation are expected to become more prevalent in the future. Some of these applications also impose new worst-case latency and reliability requirements for Wi-Fi systems. Therefore, one of the recommendations of the RTA TIG to the 802.11 working group is to consider a broader range of real-time application requirements as summarized in Table 6.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Use cases** | | **Intra BSS latency/ms** | **Jitter variance/ms [4]** | **Packet loss** | **Data rate/**  **Mbps** |
| Real-time gaming [2] | | < 5 | < 2 | < 0.1 % | < 1 |
| Cloud gaming [15] | | < 10 | < 2 | Near-lossless | < 0.1 (Reverse link)  > 5Mbps (Forward link) |
| Real-time video [3] | | < 3 ~ 10 | < 1~ 2.5 | Near-lossless | 100 ~ 28,000 |
| Robotics and  industrial automation [1][[8]](#footnote-8) | Equipment control | < 1 ~ 10 | < 0.2~2 | Near-lossless | < 1 |
| Human safety | < 1~ 10 | < 0.2 ~ 2 | Near-lossless | < 1 |
| Haptic technology | <1~5 | <0.2~2 | Lossless | <1 |
| Drone control | <100 | <10 | Lossless | <1  >100 with video |

Table 6‑1 Requirements metrics of RTA use cases

## Implementation specific optimizations

Existing 802.11 QoS mechanisms, such as traffic prioritization through EDCA access categories can be used to reduce latency and jitter. For example, time-sensitive traffic may be tagged as AC\_VO or AC\_VI and EDCA parameters may be adjusted to reduce medium access latency.

Traffic classification and mapping between higher layer services and the 802.11 MAC has been defined and shown in the table below. In this way, STAs and APs can identify the traffic accordingly and do end to end optimization. However, such approaches are not necessarily supported across multiple vendors.

| Priority | 802.1Q PCP Value | WMM UP | 802.11 EDCA Access Category | Designation |
| --- | --- | --- | --- | --- |
| Lowest  Highest | 1 | 1 | BK | Background |
| 2 | 2 | BK | Background |
| 0 | 0 | BE | Best Effort |
| 3 | 3 | BE | Best Effort |
| 4 | 4 | VI | Video (alternate) |
| 5 | 5 | VI | Video (primary) |
| 6 | 6 | VO | Voice (primary) |
| 7 | 7 | VO | Voice (alternate) |

Table 6‑2 802.1Q and WMM UP & AC mapping

Some simulations regarding packet prioritizing are done by tagging traffic from STA to AP as VO [6]. Although such approaches can help reduce average latency, the worst-case latency can still vary significantly due to contention within devices and the overall traffic load on the BSS. Therefore, it is important to consider further enhancements to the 802.11 specification that can address worst case latency issues need by a broad range of real-time applications.

## New capabilities to support real time applications

Potential enhancements and new capabilities to address requirements of emerging real-time applications can be grouped in the following categories:

**Extensions of TSN capabilities to 802.11**: As described earlier, 802.1 TSN standards are addressing real-time applications over Ethernet and extensions of TSN over 802.11 can help better support such applications over wireless medium. TSN features have already been enabled in 802.11, including traffic/stream identification, time synchronization, and integration with Ethernet bridging. But new extensions are required to address the worst-case latency problems in current Wi-Fi deployments. Time-Aware shaping and redundancy through dual links (FRE capability) are examples discussed in this report, which exist in Ethernet TSN, but need support from 802.11 in other to be adapted to wireless medium as discussed in [7]. Other TSN features may also be considered, such as alignment with the TSN management model defined by the 802.1Qcc standard.

Multiband operation simultaneously: Due to the diversity demands for Wi-Fi networks, dual-band even tri-band AP and STA products have been brought up to market and more features are expected, since nowadays one end user tend to utilize multiple media thus multiple traffic streams. So, requests for high con-currency, reducing impact of interference and traffic differentiation are becoming universal demands.

Multiband operations simultaneously can benefit not only real-time applications but also those applications request high throughput and traffic separation.

**New MAC/PHY capabilities that reduce latency and improve reliability**: There is also need for improvements in the 802.11 MAC and PHY layers to enable more predictable worst case latency, which is a fundamental requirement for most real-time application, as discussed previously in the report. It should be noted that for many real-time applications, predicable worst cast latency does not necessarily mean extremely low latency, but the ability to provide more predictable performance is the main requirement. However, in some use cases, the worst case latency requirement may also need to be low. Another related are for improved identified is reliability. Enabling features that can be used to improve overall reliability of 802.11 links are also needed to support emerging real-time applications. Although operation is unlicensed spectrum makes it difficult to provide hard performance guarantees, many Wi-Fi deployments can be managed. Therefore, it is important to enable capabilities that can be leveraged in managed environments to provide more predictable performance.

Potential areas for further enhancements include: reduced PHY overhead, predictable and efficient medium access, better support for time-sensitive small packet transmissions, improving management and time-sensitive data coexistence, coordination between APs, more flexible OFDMA resource allocation scheme, etc.

# References

[1] Zein Nader*, et al.*, “Wired/Wireless Use Cases and Communication Requirements for Flexible Factories IoT Bridged Network,” Nendica Draft Report 802.1-18-0025-06-Icne

[2] Kate Meng, “RTA report discussion,” IEEE 802.11-18-1690r5

[3] Kazuyuki Sakoda, et.al., “Old and new latency requirements,” IEEE 802.11-18/1973r3

[4] Dong-Il Dillon Seo, et.al., “Introduction to Concept of the ‘Network Enablers for seamless HMD based VR Content Service’ IG,” IEEE 802.21-18/0059r1

[5] 11-13/0657r6, “Usage models for IEEE 802.11 High Efficiency WLAN study group (HEW SG) – Liaison with WFA,” Laurent Cariou

[6] Karthik Iyer, “Packet Prioritization Issues follow-up” IEEE 802.11/2098r1

[7] Dave Cavalcanti et al, “Time-Aware shaping (802.1Qbv) support in the 802.11 MAC,” IEEE 802.11/001892r0

[8] Gartner, “Almost 3 Million Personal and Commercial Drones Will Be Shipped in 2017,” Feburuary 2017.

<https://www.gartner.com/en/newsroom/press-releases/2017-02-09-gartner-says-almost-3-million-personal-and-commercial-drones-will-be-shipped-in-2017>

[9] Drone Scan, http://www.dronescan.co/

[10] DRONE CHAMPIONS LEAGUE, https://www.dcl.aero/

[11] SWARM ARENA, <https://ars.electronica.art/futurelab/project/swarm-arena/>

[12] Swarm arena / Ars Electronica Futurelab, https://flic.kr/p/28tPf9k

[13] Association of Radio Industries and Businesses (Japan), “Requirements of Radio systems Utilization for Robots and Technical Requirements of Wireless Communication Systems,” April 2017.

http://www.soumu.go.jp/main\_sosiki/joho\_tsusin/policyreports/joho\_tsusin/idou/robot/94911.html (Japanese)

[14] Time Sensitive Networks for Flexible Manufacturing Testbed – Description of Converged Traffic Types, An Industrial Internet Consortium Results White Paper, IIC:WHT:IS3:V1.0:PB:20180417. <https://www.iiconsortium.org/pdf/IIC_TSN_Testbed_Traffic_Whitepaper_20180418.pdf>

[15] Kazuyuki Sakoda, et.al., “Additional game use case over WLAN,” IEEE 802.11-19/111r0

[16] Evgeny Khorov, et al., “**Performance evaluation of Real Time Communications over Wi-Fi**” IEEE 802.11-18/1946r4

1. The packet size range estimation is based on popular real-time mobile game data capture. Please refer to Traffic Model part. [↑](#footnote-ref-1)
2. The packet characteristics of UL and DL are different since these are using different transmission mechanisms. [↑](#footnote-ref-2)
3. The bandwidth is based on popular real-time game data capture. The main idea is that real-time mobile games are low throughput and the actual data usage is up to game duration. [↑](#footnote-ref-3)
4. In largest extreme value, a is location parameter and b is scale parameter. [↑](#footnote-ref-4)
5. The actual UL packets are fixed discrete value due to game application logic. [↑](#footnote-ref-5)
6. Numbers shown in this table are refrence numbers and should be treated as “soft” bound, obtained from engineering experiences. [↑](#footnote-ref-6)
7. The number includes time fluctuation caused by jitter. If jitter is expected in the network, the target latency should be smaller than this number. For instance, in this table, Intra-BSS target latency = 10msec is almost equivalent to target latency = 7msec with jitter variance *σ <*1msec, assuming 3*σ* is the reasonable boundary to represent the distribution.  [↑](#footnote-ref-7)
8. There may be other wireless applications in industrial automation that are not considered real-time, therefore they are out of the scope of this report. [↑](#footnote-ref-8)