**IEEE P802.24**

**Vertical Applications Technical Advisory Group**

|  |  |  |
| --- | --- | --- |
| Project | IEEE P802.24 Vertical Applications Technical Advisory Group | |
| Title | **Low Latency Communication White Paper** | |
| Date Submitted | 2019-09-18 | |
| Source | Oliver Holland  London, UK | Voice: +1 407 773 6211; +44 7916 311973 E-mail: [oliver.holland@awtg.co.uk](mailto:oliver.holland@awtg.co.uk) |
| Re: | N/A | |
| Abstract | This contribution provides a first version of the Table of Contents of the Low Latency Communication White Paper. It will be updated (along with this Abstract) as the content materializes and is included. | |
| Purpose | Assist in the development of the Low Latency Communication White Paper | |
| Notice | This document has been prepared to assist the IEEE P802.24. It is offered as a basis for discussion and is not binding on the contributing individual(s) or organization(s). The material in this document is subject to change in form and content after further study. The contributor(s) reserve(s) the right to add, amend or withdraw material contained herein. | |
| Release | The contributor acknowledges and accepts that this contribution becomes the property of IEEE and may be made publicly available by P802.24. | |

# Background and Introduction

*General information on low latency communication, background and the drive for it (e.g., linked to “5G” and IMT-2020 requirements), possibilities that are created (in a general sense), challenges that are encountered, etc.*

*The purpose of the white paper is to inform users and IEEE 802 working groups on the applications and requirements for low latency communications.*

# Low Latency Communications Applications

*Some detail about the possible applications for low latency communications. E.g., Haptic Communication expanding the human senses and interactions that can be conveyed over communication links, expanding machine “senses” and interactions that can be conveyed over communication links, the “Tactile Internet” and Industry 4.0, investment/trading(?), etc. It will be noted that many of the applications also link to high reliability, profoundly affecting the low latency solutions that might be used.*

*Different application categories / markets*

*Unique low latency requirements specific to applications and markets*

*Consider dividing applications and use cases into human interaction and machine-only.*

*Haptic applications include a human – tactile and kinesthetic. (Oliver Holland)*

*Haptic and VR environments are the most challenging for low latency since the user is engaged with more senses, and more sensitive? Latency has a bigger effect.*

Remote Surgery and Tele-medicine

*Many critical infrastructure applications are machine only, but still latency-critical.*

*E.G. electric utility grid protection*

*Add content from TSN white paper on grid protection*

*Implications of low latency networks on cyber-security. How is the threat surface changed in a low latency network?*

*Crisis management – ensuring that applications that require low response times are not affected by crisis events. More related to ultra-reliable than low-latency communications.*

## Use Case Applications from 802.11 RTA TIG (11-18-2009r6

## 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. [4]

## Wireless Console Gaming[4]

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. [4]

## Cloud Gaming

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. [4]

## 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. [4]

## Real-time video

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. [4]



Figure 4‑14 Difference between buffered video and live video

## Drone Control

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. [4]

AR/VR

Use Cases: There are a number of AR/VR use cases that are expanded upon in 802.21 report on AR/VR enablers. We won’t replicate these here in this whitepaper, but we can refer to the appropriate document found in the reference section. [5]



















Consider 802.11be and 802.15.3c

# Performance Requirements for Low Latency Communication

Derived from the discussion on applications in Section 2 and also using other sources such as the ITU definition of URLLC, will list the performance requirements of low latency communication such as:

* End-to-end data transfer latency (Edge to Edge)
* Session establishment latency(?)
* Perhaps radio access latency (noting that in some fora, this distinction is made) E.G. use cases with edge intelligence where the device to edge computing service is the critical path.
* Reliability, noting that many applications also have this requirement
* Data capacity (identify trade-offs between achieving low latency and most efficient use of bandwidth)
* Synchronization among flows (e.g., with audio/video for haptic+AV applications…?)
* Etc.
* What is the opportunity for networks to retry lost packets? How does this vary for different applications and use cases?
* Describe the relationship between reliability requirements and data rate. Not all low latency applications require high bandwidth, but the application demands very high reliability (in terms of meeting the latency requirement)
* Some applications have a requirement for precision in the haptic feedback (precision is related to low latency – delay results in error)

## Requirements for AR/VR

## Network Requirements

The network requirements for AR/VR can be summarized in the table below. For more detail the report on AR/VR Use Cases and Enablers can be found in the reference section. [5]





# Key Technologies/Solutions Supporting Low Latency Communication

Summarizing those technologies that have to be considered/utilized in order to achieve low latency, often in conjunction with high reliability. For example:

* Changes to framing to minimize wait time to receive a frame before processing the frame
* Rendering of video can be optimized based on the importance of the image, and whether the user’s eye is looking in that direction. This can allow lower latency overall.
* Video interpolation can potentially compensate for bandwidth limits that would otherwise limit frame rate.
* Prioritization of data within an application can ensure that the most user-perceptible aspects are provided the lowest latency handling in the overall system.
* Softwarization to optimize communication path through invoking elements in software at better locations?
* Network sharing to optimize communication path; neutral hosting, etc., etc.
* Multi-connectivity (as a means to still achieve reliability while reducing latency—noting that many low latency applications also require a vast *increase* in reliability compared with what is currently achieved (at least wirelessly))
* New coding approaches to achieve latency and high reliability
* New protocols
* Others (e.g., security implications and solutions)?
* Using adaptive links, multi path, and multi-band links. Multi-connectivity.
* Etc., etc. (to be added to a refined)

# IEEE 802 Standards Supporting Low Latency Communications

List of IEEE 802 standards/amendments, etc., that can already assist or realize in low latency (some in tandem with high reliability) communication, and reasoning as to how. Also listing target standards for enhancements towards low latency communication, and reasoning as to why.

## Current Standards (published)

802.1 TSN (family of standard – pull from 802.24 TSN white paper)

802.3br Interspersing Express Traffic

802.11ad (60 GHz) defines a scheduled MAC layer

802.11ai Fast Initial Link Setup, 802.11r Fast Handover (“Fast” is a relative term)

802.15.4 TSCH (more predictable, but not extremely low latency – 100 mS range)

802.15.3 support low latency, isochronous streaming. Two way streaming. 802.15.3e specifies fast link setup and teardown.

802.16 and 802.22 provide scheduled MAC with predictable latency (10s of mS)

## Target Standards for Enhancements (amendments being considered or underway)

* + 1. IEEE 802.11ax

IEEE 802.11ax, known as High Efficiency WLAN (HEW) Task Group, has started its standard development with a main goal to reduce the performance degradation in a Wi-Fi dense area. IEEE 802.11ax have accomplished Draft 2.0 in November 2017. The standard is expected to be completed by the end of 2019.

IEEE 802.11ax, which achieves four times as high as 802.11ac, is designed to operate in 2.4 GHz and 5G Hz spectrums. Through increased link efficiency in frequency domain, time domain, and modulation scheme, the 802.11ax can achieve as high as 12.01 Gbps in an ideal condition [6].

At the current development state, this technology does not satisfy the VR network considerations.

* + 1. IEEE 802.11ay

To develop the follow-up of IEEE 802.11ad, IEEE 802.11ay is formed in May 2015 to achieve a maximum throughput of at least 20 Gbps using the unlicensed mm-Wave (60 GHz) band, while maintaining or improving the power efficiency per STA. They have completed Draft 1.0 in January 2018. The standard is planned to be completed in December 2019.

IEEE 802.11ay can provide a high throughput utilizing various technologies, such as channel bonding/aggregation, MIMO (multiple-input and multiple output), and multiple channel access, etc. [6].

At the current development state, the maximum throughput is satisfied but it needs to consider the device mobility due to the directional propagation of electromagnetic wave in 60 GHz band.

* + 1. 802.11be EHT

The Task Group includes Real Time Applications as part of the scope.

* + 1. 802.11bd V2X

Low latency is a requirement for V2V use cases

# Adaptions and Recommendations for IEEE 802 Standards to Enhance Low Latency Communications Support

Suggestions on which technologies (mentioned in Section 4 above) must be introduced, and very high-level suggestions on how it might be done. Both to enhance current standards supporting low latency, as well as the target ones.

Are there common themes that can be applied across 802 technologies to enhance low latency?

We expect the 802.1 TSN TG to continue to provide the overall framework and architecture for low latency across multiple standards.

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. [4]

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

## 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. [4]

**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. [4]

# Conclusion and Future Work/Timeplan

Per usual content in this section. But could also try to project an overall vision/timeplan for implementation of such work—or such content might be extracted to its own Section? Of course, would require careful coordination with the relevant WGs.

Develop a roadmap for all 802 standards relevant to Low Latency

Identify any gaps at the architecture level for consideration by 802.1 TSN

# References

[1] IEEE 802.11 TGay Use Cases: https://mentor.ieee.org/802.11/dcn/15/11- 15-0625-07-00ay-ieee-802-11-tgay-usage-scenarios.pptx.

[2] Quality Requirements for VR, ISO/IEC JTC1/ SC29/WG11 MPEG 116 Std. m39532, 2016.

[3] IMT-Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation ITU-R M.2083-0, Sep. 2015

[4] RTA TIG Report : <https://mentor.ieee.org/802.11/dcn/18/11-18-2009-06-0rta-rta-report-draft.docx>

[5] Goal of the ‘Network Enablers for seamless HMD based VR Content Service’ SG : <https://mentor.ieee.org/802.21/dcn/18/21-18-0065-00-0000-21-18-0065-00-0000-goal-of-the-network-enablers-for-seamless-hmd-based-vr-content-service-sg.pptx>

1. 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-1)