# Design and Implemetaion of a Cloud-Based Metaverse Virtual Performance Platform

Yu Cho Rong
Content Research Section
*Electronics and Telecommunications Research Institute*
Daejeon, South Korea
crryu@etri.re.kr

Choi In Moon
Content Research Section
*Electronics and Telecommunications Research Institute*
Daejeon, South Korea
inmchoi@etri.re.kr

Kang Kyung Kyu
Content Research Section
*Electronics and Telecommunications Research Institute*
Daejeon, South Korea
kangk2@etri.re.kr

Gil Youn Hee
Content Research Section
*Electronics and Telecommunications Research Institute*
Daejeon, South Korea
yhgil@etri.re.kr

***Abstract*— This paper presents a comprehensive design and implementation of a cloud-based metaverse virtual performance platform, using AWS(Amazon Web Services). We proposed a method to effectively support interaction between audiences while supporting real-time virtual performance content data transmission based on a network structure with two tier servers, and utilized cloud services to support more than tens of thousands of audiences.**

***Keywords-Metaverse, Virtual Performance, AWS(Amazon Web Services), Unreal Engine, Interaction***

## I. INTRODUCTION

In response to the global pandemic, there has been a growing demand for virtual experiences that allow individuals to engage in events and performances from the safety of their homes[1][2]. In January 2023, *Wave*(wavexr.com) collaborated with *PICO* to hold a virtual performance featuring ‘Calvin Harris’. *Wave* is spearheading an immersive 3D concert series, featuring greatest artists like ‘Justin Bieber’, ‘John Legend’ and etc[3].

Fig. 1. ‘Calvin Harris’’s virtual performance on *Wave*(wavexr.com)

In South Korea, during July 2022, *SK Telecom (SKT)* showcased a virtual concert boasting ultra-high-definition volumetric content on the metaverse platform 'ifland'[4]. Subsequently, in 2023, *LOTTE*(lotte.co.kr) successfully staged immersive virtual performances featuring 'NMIXX' and 'DJ ALOK' at ‘CES 2023’, executed through the CALIVERSE(metaverse) platform, closely resembling real performances and allowing a large number of audiences to connect simultaneously[5].

Fig. 2. ‘DJ ALOK’’s virtual performance through CALIVERSE Platform

As such, existing virtual performance services enable large numbers of audiences to participate simultaneously, but has the drawback that all audiences are not connected to one server, but are distributed and connected to multiple servers, and interaction is only possible between audiences connected to the same server. In order to solve this problem, this paper aims to design and implement a metaverse platform where tens of thousands of audiences can freely interact with each other and performer by participating in virtual performances in real time. This paper introduces virtual performances through the utilization of a cloud-based metaverse platform, which offers a scalable and dynamic infrastructure for hosting immersive events.

## II. PROPOSED METAVERSE VIRTUAL PERFORMANCE PLATFORM

The metaverse virtual performance platform to be proposed in this paper is classified according to the physical space as follows. It can be classified into an actual venue space where performers are located, a space where audiences are located, and a network space located between performers and audience.

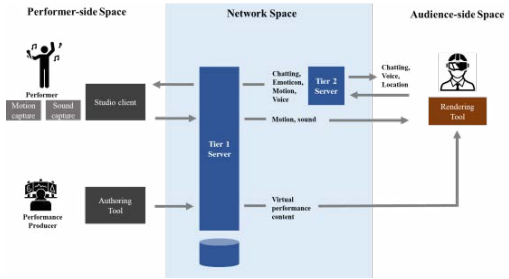


Fig. 3. An overview of the proposed virtual performance platform

### A. Performer-side Space

The performer space where performers are located should be equipped with the necessary equipment and systems to capture the performers' motions and sound[6]. The motion capture system and sound capture system can exist as separate systems or can be integrated into a single system. They capture the performer's motion and sound in real-time and transmit the

data to the studio client. The studio client forwards the performer's motion and sound through the network infrastructure and visualizes and transmits the interaction data received from the audience to the performer. The motion information acquired through the motion capture system provides position information (3D vectors) for each facial and body joint corresponding to each frame. To maintain immersion, the synchronization of sound and motion is very important factor. Synchronization can be achieved by encoding sound and motion information and transmitting them together as one combined data stream, or by separately transmitting each data and synchronizing them at the audience side. In this paper, synchronization is designed to be managed by the studio client and transmitted through the network infrastructure. The performance producer must implement virtual stage and various contents and assets including avatar, etc. composed virtual performance, before starting the performance. and then they need to register implemented virtual performance content and information about the performance on the metaverse platform so that both performers and audiences can access the performance. Additionally, they must respond promptly to control the performance in real-time and address any possible issues that arise. The performance producer can be physically located in the space where the performers are.

### *B. Network Space*

The Tier 1 server assumes responsibility for overseeing all systems and audience-related data within the virtual performance platform. Tier 2 servers are the interface through which actual audiences connect. Upon connecting to the Tier 2 server, audiences are granted access to real-time motion and sound data from the performer. In case of interaction, interaction data from audiences that are globally shareable are routed to the Tier 1 server. Conversely, interaction data intended for local broadcasting among proximate audiences is processed within the Tier 2 server. This approach ensures streamlined management of transmitted data. In audience-centric context, the Tier 2 server receives and promptly updates real-time positional and interaction data for connected audiences. Subsequently, the server broadcast the updated positional information, thereby enable nearby audiences' movements rendering in virtual performance content immediately. Interaction data can be broadcasted to all audiences connected to Tier 2 servers either via the Tier 1 server or directed towards specific audiences or groups. The proposed interaction framework includes audience movement and voice expression, as well as interactions in traditional virtual performances, such as chat, emoticons, and voting. Audience position data includes both the absolute position (3D vector) and HMD direction (3D vector) within the virtual performance using data from a head-mounted display (HMD).

### *C. Audience-side Space*

Audiences need to download and install a virtual performance rendering tool. When the virtual performance begins, the audience rendering tool renders the virtual performance content implemented by performance producer. This performance content also be downloaded in advance.

## III. Detailed Design of our Proposed Platform

In this section, more detailed design of metaverse virtual performance platform proposed in this paper is described.

The system components that constitute the metaverse virtual performance platform proposed in this paper are as follows: 1) Virtual Performance Content Authoring Tool, 2) Motion Capture System, 3) Sound Capture System, 4) Studio Client, 5) Tier 1 Server/Tier 2 Server, 6) Audience Rendering Tool.

#### 1) Virtual Performance Content Authoring Tool

The virtual performance content authoring tool creates Unreal5 based virtual performance content that includes various assets, sequences, and scenarios required for virtual performances. It constructs content based on the performance scenario, including the stage, special effects assets, avatars and etc.

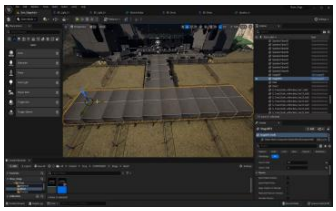

Fig. 4. Unreal5 based Authoring Tool (unreal editor plugin)

#### 2) Motion Capture System

We utilized *Vicon*(vicon.com) motion capture system in this paper. The acquired motion data is treated in a comparable format to the joint position data utilized within the Unreal Live Link plugin.

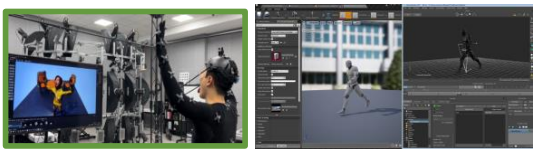

Fig. 5. Motion capture studio(left), Unreal Engine Live Link using MotionBuilder animation(right)

#### 3) Sound Capture System

Acquired sound data through the sound capture system is streamed in real time in a lossless .wav file format.

#### 4) Studio Client

Before transmitting the data acquired through the motion capture system and sound capture system to the network, it is checked that there are no errors, and the performer's motion and sound are streamed in real time and reflected in the virtual performance contents to verify that they are successfully rendered. It also provides a function of synchronizing motion data and sound and transmitting them over the network, and visualizing interactive data transmitted from the audiences.

#### 5) Tier 1 server/Tier 2 Server

The proposed platform is built upon the AWS cloud infrastructure[7]. Tier 1 servers manage all systems and user information, while Tier 2 servers are allocated based on user counts and handle the real-time interactions. AWS services such as EC2 instances, RDS, and REDIS are leveraged to ensure efficient and reliable performance. As shown in Fig. 6, it is designed so that Tier 2 servers can be created flexibly through the load balancer according to the number of audiences connected to the platform and the network bandwidth.

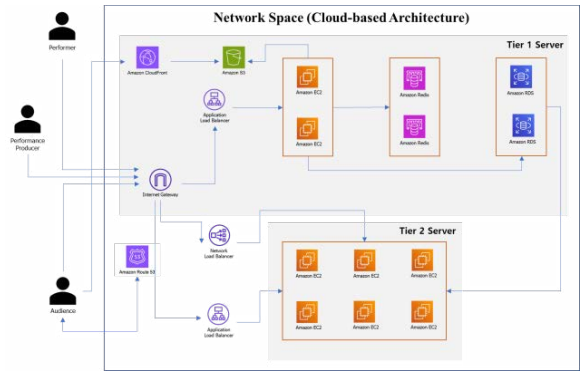


Fig. 6. Proposed design of network architecture based AWS(Amazon Web Services)

In designing and implementing the network architecture, it is important to define data transmission specifications as well as fast data transmission. The data transmission specification should be able to easily add data types and support various data sizes. So, in this paper, we used json as a data transmission specification. However, in order to support real-time rendering of more than 20fps in the audience rendering tool, the computation time required to parse the motion and sound data transmitted from the performer from json must be reduced. Motion and sound data are transmitted in fixed-size binary format, and other interaction data and virtual performance control data are transmitted in json format. Fig. 7 shows JSON data sample used the case of transmission from audience to Tier 2 server.

```
{
  "dataType" : "location",
   // voice, chat, emoticon, motion and etc.
  "data" : {
    "UserID" : 1,
      "WorldPosition" : {
          "x" : 3.0,
          "y" : 3.0,
          "z" : 3.0
      },
      "WorldRotation" : {
          "x" : 4.0,
          "y" : 4.0,
          "z" : 4.0
      },
      "Direction" : {
          "x" : 1.0,
          "y" : 1.0,
          "z" : 1.0
      },
      ...
  }
}
```

Fig. 7. JSON data from audience to Tier 2 server (location)

### 7) Audience Rendering Tool

In this paper, virtual performance contents are rendered in the audience's rendering tool. It receives the motion and sound data of the performers in real time and reflects them in the virtual performance contents and renders them. The rendering tool runs on the Windows operating system through an HMD like Meta Quest 2. Based on the position and direction data of the audience's HMD, they can move or change their direction within the virtual performance, and interact with the audience located nearby.

Fig. 8. Virtual performance content rendering using audience rendering tool development version

Currently, each system component is implemented as a development version, and system integration and demonstration services for tens of thousands of audiences are scheduled at the end of this year. As the number of audiences increases during the integration process, there is a possibility of network modification considering data transmission delay time, so our proposed platform configuration will be modified after the integration test.

## IV. Conclusion

This paper introduced a metaverse virtual performance platform based on AWS and explained the detailed design and implementation. We proposed a method to effectively support interaction between audiences while supporting real-time virtual performance content data transmission based on a network structure with two tier servers. In addition, cloud services were used to support simultaneous access to virtual performances by more than tens of thousands of audiences, and network design using AWS instances was introduced. In the future, we plan to integrate the system components implemented in the current development version and modify and supplement the platform to support more than 100,000 audiences. In preparation for the difference in data size transmitted depending on the type and configuration of virtual performance content, we will continue to study how to minimize data transmission delay by conducting tests under various environmental conditions.

## Acknowledgment

This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2022 (Project Name: Development of real-time interactive metaverse performance experience platform technology on the scale of a large concert hall Project Number: RS-2022-050002, Contribution Rate: 100%)

## References

[1] Go. Byeong-Su, Kim. Myeong-Ha, "Metaverse-based immersive content R&D support project trend", Broadcasting and Media Magazine, Vol. 27 Issue 1, 2022, pp. 21-26.

[2] Kim. Gwang-jip, "The evolution of the real and virtual world through the case of the metaverse", Broadcasting and Media Magazine, Vol. 26 Issue 3, 2021, pp. 10-19.

[3] https://wavexr.com/

[4] https://ifland.io/

[5] https://www.caliverse.co.kr/

[6] Lee. S.B., Lee. G.M., Han. S.K., Jeong. S.H., Lee. J.J., "A Simulcast System for Live Streaming and Virtual Avatar Concerts", J Korea Comput Graph Soc 2023 29(2), pp.21-30.

Yu. C.R., Gil. Y.H., Jeong. I.K., "A Study on the metaverse network design for accommodating large-scale virtual performance users", In Proceedings of Symposium of the Korean Institute of communications and Information Sciences, Gyeongju, Korea, 16-18 November 2022, pp. 749-751.