

# Games Are Not Equal: Classifying Cloud Gaming Contexts for Effective User Experience Measurement

Yifan Wang  
University of New South Wales  
Sydney, NSW, Australia  
wangyifan.frank@student.unsw.edu.au

Minzhao Lyu  
University of New South Wales  
Sydney, NSW, Australia  
minzhao.lyu@unsw.edu.au

Vijay Sivaraman  
University of New South Wales  
Sydney, NSW, Australia  
vijay@unsw.edu.au

## Abstract

To tap into the growing market of cloud gaming, whereby game graphics is rendered in the cloud and streamed back to the user as a video feed, network operators are creating monetizable assurance services that dynamically provision network resources. However, without accurately measuring cloud gaming user experience, they cannot assess the effectiveness of their provisioning methods. Basic measures such as bandwidth and frame rate by themselves do not suffice, and can only be interpreted in the context of the game played and the player activity within the game. This paper equips the network operator with a method to obtain a real-time measure of cloud gaming experience by analyzing network traffic, including contextual factors such as the game title and player activity stage. Our method is able to classify the game title within the first five seconds of game launch, and continuously assess the player activity stage as being active, passive, or idle. We deploy it in an ISP hosting NVIDIA cloud gaming servers for the region. We provide insights from hundreds of thousands of cloud game streaming sessions over a three-month period into the dependence of bandwidth consumption and experience level on the gameplay contexts.

## CCS Concepts

• **Networks** → **Network measurement; Network monitoring;**  
• **Information systems** → *Multimedia streaming*; • **Computing methodologies** → *Machine learning approaches*.

## Keywords

Network traffic analysis; cloud gaming; Quality-of-Experience; machine learning

## ACM Reference Format:

Yifan Wang, Minzhao Lyu, and Vijay Sivaraman. 2025. Games Are Not Equal: Classifying Cloud Gaming Contexts for Effective User Experience Measurement. In *Proceedings of the 2025 ACM Internet Measurement Conference (IMC '25)*, October 28–31, 2025, Madison, WI, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3730567.3764455>

Corresponding author: Minzhao Lyu (minzhao.lyu@unsw.edu.au).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
IMC '25, Madison, WI, USA

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 979-8-4007-1860-1/2025/10  
<https://doi.org/10.1145/3730567.3764455>

## 1 Introduction

Cloud gaming has gained significant popularity as an affordable and accessible alternative to traditional gaming by shifting intensive real-time graphics rendering and gaming computation from client hardware to cloud GPU servers. Users on high-speed networks can play the latest game releases without high-end graphics computing hardware. Driven by the increasing demand for high-quality gaming, the maturity of cloud GPU infrastructure, and the increase in the capacity of carrier networks [20], the cloud gaming industry, led by tech giants such as NVIDIA [45], Microsoft [41], Amazon [2] and Sony [57], is expected to soar to US\$143bn by 2032 [67].

Network operators are experimenting with ways to monetize cloud gaming, by providing network assurance for the service. Unlike traditional online games that only require a few hundred Kbps of bandwidth from the network [36], cloud games need much higher bandwidth of tens of Mbps; simultaneously, unlike traditional video streaming that can buffer seconds or even minutes of content, cloud game streaming experience is impacted by sub-second glitches in latency or packet drops [34]. This highly demanding nature of cloud gaming in terms of both bandwidth and latency/loss renders it amenable for a premium (monetizable) assurance service, such as via a 5G slice dedicated to cloud gaming [61], or via dynamic provisioning *e.g.*, using the Quality-on-Demand API being standardized by the CAMARA group in the TM Forum [58]. As an example, Telefonica announced in 2023 that it was conducting such a trial with the Blacknut Cloud Gaming service [9], followed more recently by Singtel's collaboration with Tencent Games to launch a specialized 5G slice service for cloud game streaming [53] in 2025.

For a network operator to assure the gaming experience, they have to be able to measure the gaming experience accurately. This is not as trivial as measuring the throughput, latency, and loss of the game video stream, because these vary depending on the title/genre of the game [8, 48], as well as the activity stage of the player within the game [13, 43], namely whether they are active (*e.g.*, shooting), passive (*e.g.*, spectating), or idle (*e.g.*, between rounds). In the absence of this context, drops in bandwidth or streaming frame rate may wrongly be associated with poor experience, when in reality the user is engaged in a less intensive role-playing game or a strategy card game, or spectating others in a shooting/sports game. For example, players in a First Person Shooter (FPS) game often actively engage in gameplay when a competitive match starts with stringent requirements on high streaming frame rate, and have relatively lower network requirements during passive activity stage such as matchmaking and observation. As another example, players of a card game are likely to have nearly static gaming scenarios with low expectations on streaming frame rate and bandwidth during the entire gameplay.



**Figure 1: Examples of the two types of gameplay activity patterns including (a) a shooter game session where player activities follow a spectate-and-play pattern and (b) a role-playing game session where player activities follow a continuous-play pattern.**

In this paper, we develop a novel method to help network operators measure the **context** of cloud gaming sessions, specifically the title/genre of the game and the player activity stage (*i.e.*, active, passive or idle) within the gameplay. This can be combined with prior works that have developed traffic analysis methods to classify cloud gaming sessions, identify user platforms, and measure objective Quality-of-Experience (QoE) metrics from network Quality-of-Service (QoS) parameters [13, 18, 32], as well as studies that have benchmarked the degradation of game streaming quality due to network conditions [4, 24, 29, 34, 38]. With our three contributions listed below, this paper fills a gap by providing network operators with an accurate measure of the **effective** user experience that takes into account the context of game title/genre and player activity stage.

Our **first** contribution (§3) systematically identifies the unique network traffic characteristics of cloud game streaming sessions in various gameplay contexts. We collect and study traffic traces of over 500 cloud gaming sessions across 13 most popular games spanning 5 genres and diverse game streaming settings. The labeled traffic dataset will be shared with the community. We reveal the unique statistical distributions in packet sizes and timings during the initial launch stage of each cloud game title/genre, as well as the relative volumetric patterns within each streaming session that relate to the player activity stages in the game.

In the **second** contribution (§4), we develop and evaluate a real-time network traffic analysis method to classify the cloud game title within the first five seconds of game launch, and continuously measure the player activity stage thereafter. Game title classification is achieved using pre-trained machine learning models on packet sizes and timing attributes from the first five seconds, whereas the player activity stage is continuously assessed from bidirectional volumetric profiles. When the fine-grained game title cannot be confidently inferred, our method infers the coarse-grained game genre (*i.e.*, gameplay activity pattern) from the transition behaviors between player activity stages.

Our **third** contribution (§5) deploys our cloud gameplay context classification into a live network operated by our partner Internet service provider (ISP) that hosts NVIDIA’s GeForce NOW cloud gaming servers for our region. We collect data over a three-month period, spanning hundreds of thousands of cloud gaming hours. We validate our game title/genre classification method determined in real-time against ground truth from the game server logs produced offline after the conclusion of each gameplay session. We then highlight how bandwidth demand varies depending on the game title/genre, as well as the player activity stage within the game. Lastly, we show that by providing appropriate game contexts, our method can help network operators avoid mislabeling

experience drops that are attributable to the game title/genre or player activity rather than poor network conditions, thereby helping them accurately gauge and assure cloud gaming experience over their network.

## 2 Preliminary of User Experience Contexts in Cloud Gaming

In this section, we provide preliminaries on contextual factors of cloud gameplay that can affect user expectations on their perceived game streaming quality, starting from an intuitive understanding of typical player activity patterns during cloud gameplay sessions (§2.1), to game contexts (*i.e.*, titles/genres) that exhibit unique activity patterns (§2.2).

### 2.1 Player Activity Stage and Gameplay Activity Pattern

Prior works [32, 38] have revealed that players who access commercial cloud gaming services first connect to the respective cloud platform for administrative services, game selection, and allocation of cloud game servers before actual gameplay. However, diversified user activities during cloud gameplay sessions that can lead to different user expectations on their perceived streaming quality, network traffic characteristics, and network demands to be satisfied by network operators have not yet been studied. Therefore, we now zoom into the gameplay session to understand the various types of user activities categorized into three levels of **player activity stages**, namely **active**, **passive** and **idle**, which can lead to various expectations on the streaming quality.

As visually shown in Fig. 1, we use two representative gameplay sessions on NVIDIA’s GeForce NOW platform to discuss the **gameplay activity patterns** each with a different profile of player activity stages. Our discussed insights are generalizable to other commercial platforms.

The first example in Fig. 1(a) shows a CS:GO shooter game session where players’ gameplay activities can be divided into multiple **spectate-and-play** slots. In this shooter gameplay, after the game has been selected on the cloud gaming platform, the player first waits for the game to launch with a transition scene and then stays in the lobby until a match starts. During these periods, the player maintains an almost idle level of gameplay activity. As will be discussed later in §3, such idle activities also exhibit low-profile traffic characteristics. Once the player starts a match, the gameplay activity becomes very active, including both high-frequency graphics refreshment and user input updates. During this shooter game match, the player sometimes gets eliminated by opponents and waits to respawn, becoming passive in terms of gameplay activity

**Table 1: Thirteen popular cloud game titles played on NVIDIA’s GeForce NOW platform in our geography, with their game genre, gameplay activity pattern, and popularity by their fraction of total playtime.**

Game title	Game genre	Activity pattern	Popularity
Fortnite	Shooter	Spectate-and-play	37.80%
Genshin Impact	Role-playing	Continuous-play	20.10%
Baldur’s Gate 3	Role-playing	Continuous-play	3.30%
R6: Siege	Shooter	Spectate-and-play	1.24%
Honkai: Star Rail	Role-playing	Continuous-play	1.16%
Destiny 2	Shooter	Spectate-and-play	1.15%
Call of Duty	Shooter	Spectate-and-play	0.97%
Cyberpunk 2077	Role-playing	Continuous-play	0.84%
Overwatch 2	Shooter	Spectate-and-play	0.74%
Rocket League	Sports	Spectate-and-play	0.64%
CS:GO/CS2	Shooter	Spectate-and-play	0.61%
Dota 2	MOBA	Spectate-and-play	0.55%
Hearthstone	Card	Spectate-and-play	0.04%

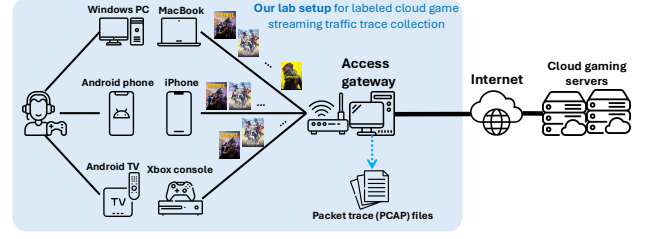
by spectating with decent graphics refresh frequency but very limited user interaction. The player goes back to the idle activity stage in the lobby before starting the second match. In realistic gameplay sessions, the spectate-and-play pattern often occurs multiple times (twice in this example) before the session ends.

The second representative gameplay is shown in Fig. 1(b), where the user plays Cyberpunk 2077, a role-playing game that follows a **continuous-play** activity pattern. For this type of gameplay activity pattern, players often engage in active gameplay activity in a continuous manner for a relatively long period, except at the beginning (e.g., login and character selection) and end of the session. Under such game mechanism, during the active gameplay stage, the player explores the game world without interruption and completes multiple missions, with only occasional short occurrences of passive or idle stages (e.g., static dialogue or transition animations) featuring infrequent graphics refresh and user motion updates. As will be discussed in §3, gameplay following the continuous-play player activity pattern has its traffic profile at a constantly high level with small fluctuations.

## 2.2 Cloud Game Titles and Genres

Having discussed the three categories of player activity stages (i.e., idle, passive, and active) observable within the two types of gameplay activity patterns (i.e., spectate-and-play and continuous-play), each of which can result in unique expectations on game streaming quality such as frame rate and, therefore, network bandwidth demands, we now discuss the gameplay contexts including game titles and genres that inherently lead to these distinct gameplay activity patterns.

We have collected an exhaustive list of cloud game catalog from our partnered network operator hosting NVIDIA’s GeForce NOW cloud gaming servers in our geography. While we cannot disclose the full list as requested by our industry partner, we present in Table 1 the top 13 game titles that contribute to over 69% of total gameplay time during a half-year period encompassing millions of gameplay sessions, with their genres defined by the gaming



**Figure 2: Lab experiment setup for collecting traffic traces of cloud gaming sessions.**

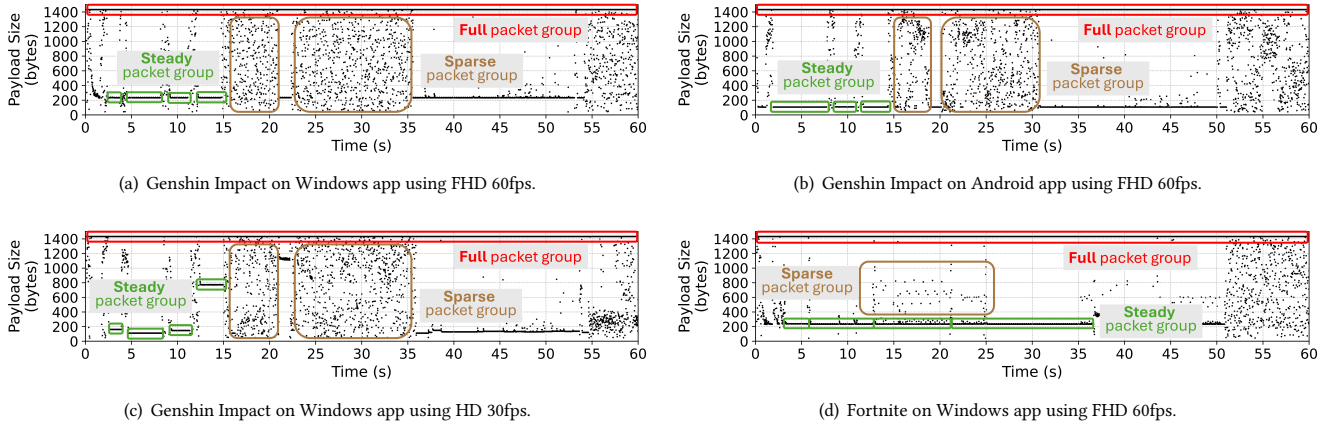
**Table 2: Our lab traffic capture dataset of NVIDIA’s GeForce NOW cloud gameplay sessions across the thirteen popular game titles on a diverse profiles of user configuration.**

Device	OS	Software	Streaming settings	#Sessions	Playtime
PC	Windows	Native app	UHD-SD; 30-120 fps	89	10.9 hours
		Browser	QHD-SD; 30-120 fps	60	6.8 hours
	macOS	Native app	UHD-SD; 30-120 fps	76	10.5 hours
		Browser	QHD-SD; 30-120 fps	61	7.7 hours
Mobile	Android	Native app	QHD-FHD; 30-120 fps	73	9.1 hours
	iOS	Browser	FHD-SD; 30-120 fps	70	8.8 hours
TV	AndroidTV	Native app	FHD-SD; 30-120 fps	48	6.1 hours
Console	Xbox	Browser	FHD-SD; 30-120 fps	54	7.1 hours

community, popularity in percentage of gameplay time, and their respective gameplay activity patterns observed in our study listed.

The top 13 game titles belong to five game genres as defined by the gaming community based on their content, including shooter, role-playing, sports, MOBA, and card games. The two examples just discussed in §2.1 are both in the list. By investigating into all the thirteen game titles on the GeForce NOW platform as well as popular games on three other major cloud gaming platforms, we validated that the gameplay activity patterns (i.e., spectate-and-play and continuous-play) as articulated in §2.1 are directly correlated with their game title and genre. For example, all the six shooter games in the list have their gameplay activities following the spectate-and-play pattern with a repeating combination of idle, active, and/or passive player activity stages, so do MOBA and card games. Player activities in all cloud gameplay sessions of role-playing games are constantly active and thus follow the continuous-play pattern.

As measured in [8, 34], cloud gaming servers render gaming graphics that are streamed back to players’ devices with different quality and encoding granularity depending on the game titles. Soon in §3, we will discuss that the network traffic characteristics (e.g., bandwidth usage) vary across not only game titles and gameplay activity patterns, but also player activity stages within each session, leading to diversified user expectations on streaming quality and thus network demands of the particular session. Therefore, knowing the cloud game contexts (i.e., game titles/genres and player activity stages) is essential for network operators to accurately understand the effective cloud gaming quality delivered over their network infrastructure to the subscribers, so that they can precisely troubleshoot those that are indeed impacted by poor network conditions for proactive network optimization.



**Figure 3: Time-series scatter plots for the downstream packet payload sizes arriving in the first 60 seconds (covering the initial game launching stage) of four representative cloud game streaming sessions, where the packets can be categorized into the full, steady, or sparse groups.**

### 3 Traffic Characterization across Cloud Game Contexts

In this section, using our ground-truth traffic traces of cloud gaming sessions labeled by game titles and in-game player activity stages (§3.1), we discuss our findings into the unique group characteristics of packets during game launch across titles (§3.2) and the volumetric characteristics of streaming flows in the upstream and downstream directions across player activity stages (§3.3).

#### 3.1 Dataset Collection and Overview

To the best of our knowledge, existing public traffic trace datasets (e.g., [13, 38, 55, 56]) on cloud gaming do not contain sufficient labels of the game contexts articulated in §2, and/or not have comprehensive coverage of game streaming configurations. Therefore, to obtain a holistic understanding of the network traffic characteristics of cloud gameplay sessions with diversified configurations (i.e., user devices, operating systems, software applications and streaming settings) across different types of cloud game contexts (i.e., game titles, genres and player activity stages), we built a lab setup to collect ground-truth traffic capture (PCAP) files of cloud game streaming sessions, as visually shown in Fig. 2.

The setup consists of a diverse collection of user devices that support cloud game streaming services, including mobile phones, PCs, a gaming console and a smart TV. The devices are connected to our regional NVIDIA GeForce NOW cloud servers hosted by our partnered Internet service provider with network conditions of nearly 1 Gbps bandwidth, less than 10 ms latency and less than 0.1% packet loss rate. They can also access cloud gaming services hosted by other platforms such as Microsoft’s Xbox Cloud Gaming with nearly ideal network conditions. The Linux-based access gateway connecting user devices is configured with packet capturing capability through Wireshark and TCPdump software.

With a team comprising two of the authors and four student volunteers, we have collected labeled PCAP files for 531 cloud game streaming sessions with 67 hours of total playtime across the 13 popular game titles listed in Table 1. The number of game

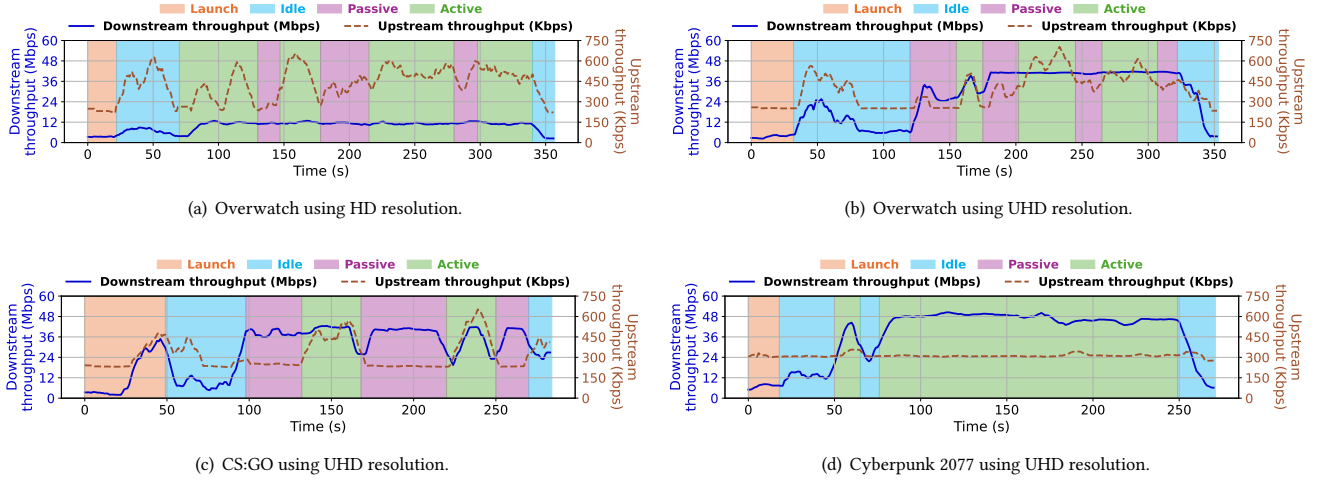
streaming sessions and their total playtime of the collected traffic traces labeled by their device genres, operating systems, software types, and streaming settings are presented in Table 2. To enable us to analyze traffic characteristics associated not only with per-session game titles but also with in-game player activity stages, each PCAP file that contains packet streams of an entire gameplay session is also labeled by the changing player activity stages (i.e., idle, active, or passive) with timestamps.

To validate the generalizability of our insights to other commercial cloud gaming platforms, we have also collected labeled PCAP files for a representative range of gameplay sessions on three other major platforms including Xbox Cloud Gaming, Amazon Luna, and PS5 Cloud Streaming, in addition to the PCAPs for GeForce NOW. The dataset is about 800 GB in size and is shared with the Internet research community through our university cloud drive with the availability information provided in Appendix §B.

#### 3.2 Packet Group Profiles across Game Titles

During cloud gameplay, cloud servers stream the rendered gaming video and receive user inputs in real-time through standard Real-time Transport Protocol (RTP) flows [18, 32, 46]. Specifically, gaming graphics and audio are sent from cloud servers to the client devices via downstream RTP packets of the respective streaming flows, whereas player inputs such as mouse clicks, key strokes, mobile screen touches, and voice are carried in upstream RTP packets from the client devices to cloud servers. Prior works [18, 32] have also developed signatures to detect RTP flows for cloud game streaming using flow metadata, which are used as preliminary filtering conditions on the collected traffic traces for us to focus on the game streaming flows and investigate their behavioral profiles across game contexts.

As already illustrated in Fig. 1, before a player starts actively or passively engaging in gameplay, there is always a game **launch** stage lasting from tens of seconds to one minute, in the form of an opening animation streamed from the cloud server, to keep players waiting for the initialization of the game. In traditional



**Figure 4: Time-series line chart showing the downstream throughput and upstream packet rate of game streaming flows in four representative cloud gaming sessions with background color-coded by player activity stages as **launch**, **idle**, **passive** and **active**.**

gaming, such launch animations are installed and rendered on local gaming devices, while in cloud gaming, the animations that are made differently for each game title are live-streamed to players, leading to unique traffic characteristics across cloud game titles.

By analyzing our traffic traces labeled with their game titles, we observe unique packet group profiles in the downstream direction for their arrival times and payload sizes that carry launching scenes. Fig. 3(a) visually shows the payload size and arrival time of the downstream packets in the RTP streaming flow for the first 60 seconds of a Genshin Impact gameplay session on a Windows PC with full HD graphics and 60fps streaming frame rate. From the scatter plot, we intuitively see that the packets fall into three groups, including **full** packets that have the same fixed (maximum) payload size and are constantly streamed to the local device; **steady** packets that hold similar payload sizes to their adjacent packets arriving within certain seconds; and **sparse** packets with high variations in payload sizes compared to their neighboring packets.

The presence of the three packet groups holds similar profiles (*i.e.*, timing characteristics of each packet group and relative payload sizes across intervals of the same packet group) in all gameplay sessions of the same game titles, regardless of their client device types and streaming configurations. The packet statistics become unpredictable after the launching stage when players are involved in gameplay. As evidence, Fig. 3(b) and Fig. 3(c) show the scatter plots for the same game title (*i.e.*, Genshin Impact) with different user settings. It is clear that the relative payload sizes and arrival time slots of the full, sparse, and steady packet groups in Fig. 3(b) remain almost identical compared to Fig. 3(a). For Fig. 3(c), the arrival time slots of the three packet groups are identical compared to Fig. 3(a), with tiny variations in relative payload sizes during three time slots (4.1 seconds in total) for the steady packet group.

Gameplay sessions of different game titles exhibit distinctive profiles for the three groups of packets. An example is provided in Fig. 3(d) for a Fortnite gameplay session. The arrival density of full packets, the arrival time slots of sparse and steady packet groups,

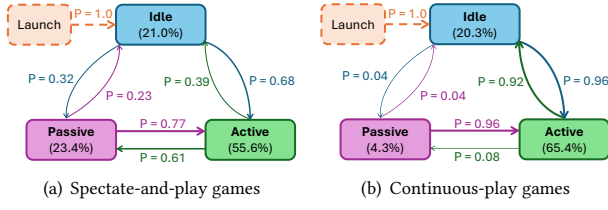
and their relative payload sizes are all different compared to game streaming sessions of the just-discussed Genshin Impact and other popular game titles in our dataset.

### 3.3 Flow Volumetric Profiles across Player Activity Stages

After understanding the packet characteristics across game titles that indicate the overall QoE expectation of a game streaming session [8, 48], we now discuss the volumetric profiles of streaming flows that dynamically change on player activity stages in a cloud game. As a common practice in the gaming industry, graphics rendering and gaming computation are dynamically optimized at the software level for efficient resource consumption [40, 59]. For cloud game streaming, such optimizations lead to reduced demands on QoE metrics (*e.g.*, streaming frame rate) and network resource consumption (*e.g.*, bandwidth) for delivering a good user experience in real time [34, 38].

By analyzing the collected PCAP files of cloud gameplay with labeled player activities, we identify unique volumetric patterns in both upstream and downstream directions depending on the stages of player activity from idle, passive to active, with a visual example provided in Fig. 4(a). The figure shows the upstream and downstream throughput of game streaming flows for an Overwatch gameplay session with background colors indicating the ground-truth player activity stages.

For the **idle** player activity stage, we can see that the throughput in the downstream direction remains relatively low during the first idle stage when the player was “browsing available matching options” and drops to a low level in both directions when the player was “back to the hub from the matching room” at the end of this streaming session. In contrast, when the player activity stage becomes **active** for “combating in progress”, the streaming throughput jumps to the highest level of the entire session, especially in the downstream direction. During the **passive** stage when the player was passively watching the progress of gameplay such



**Figure 5: The average percentage of total playtime spent in the idle, passive, and active player activity stages with their transition probabilities per game streaming session for (a) spectate-and-play and (b) continuous-play cloud games in our labeled dataset.**

as “being defeated by opponents in the match and watching the activity of teammates”, the downstream throughput remains at a similarly high level to the active stage but with reduced throughput in the upstream direction from the client to the cloud server. The observations in the relative (not absolute) changes of upstream and downstream volumetric profiles during a cloud gameplay session are consistent across all 13 popular game titles. Later in §4, we will leverage the bidirectional volumetric patterns as discussed to develop our method to reversely infer the player activity stages.

As discussed in §2.1, due to the designed gameplay mechanism of each game title, the 13 popular games can be categorized by their gameplay activity patterns as either “continuous-play” or “spectate-and-play”. We observe that the streaming sessions of the **continuous-play** game type, e.g., *Cyberpunk 2077* in Fig. 4(d), often have a larger fraction of playtime in the active player activity stage compared to that of **spectate-and-play** games like *Overwatch* in Fig. 4(b) and *CS:GO* in Fig. 4(c). Also, the **transitions among player activity stages** hold similarities for gaming sessions of the same type of gameplay activity pattern. The two state transition diagrams in Fig. 5 show the average fraction of playtime spent in the active, passive, and idle player activity stages per gameplay session, with their transition probabilities computed from all ground-truth sessions. For game streaming sessions of the “spectate-and-play” games, the active stage often takes about 40% to 60% of the total playtime and the passive stage accounts for the majority of the rest of the playtime. As for the “continuous-play” games, more than 95% playtime is spent in the active or idle player activity stages and less than 5% playtime is spent in the passive stage.

Our empirical findings in the differentiation of gameplay activity patterns show the ineffectiveness of existing cloud gaming user experience measurements relying solely on objective QoE metrics, which can mislabel “spectate-and-play” game streaming sessions as having poor user experience due to their low volumetric profiles during passive and idle stages. The findings also inspire us to leverage the transition behaviors of the player activity stages to classify the gameplay activity patterns.

## 4 Classifying Cloud Game Contexts from Real-time Network Traffic

Driven by the insights into packet group characteristics and flow volumetric profiles across game titles and player activity stages in §3, we develop our traffic analysis method (§4.1) to classify cloud

game contexts for network operators in real time, including two classification processes that use statistical machine learning models to classify game titles (§4.2) and player activity stages (§4.3). The classification performance of our method is extensively evaluated in our lab using ground-truth sessions (§4.4) before deployment in the wild.

### 4.1 Methodology Overview

Our network traffic analysis method (overviewed in Fig. 6) is designed to process real-time network traffic through an operational network to classify the gameplay contexts of cloud game streaming flows.

Upon receiving real-time packet streams, the cloud gaming packet filter module selects only packets belonging to cloud game streaming flows for further processing and classification. We use adapted state-of-the-art cloud game streaming flow detection signatures [23, 32, 52] that achieve 100% accuracy from our lab validation in detecting RTP flows of four major cloud gaming platforms including NVIDIA’s GeForce NOW, Microsoft’s Xbox Cloud Gaming, Amazon Luna, and Sony’s PS5 Cloud Streaming.

The packet streams of game streaming flows are then forwarded to three parallel processes. The bottom one, in the gray region of Fig. 6, continuously measures objective QoE metrics, including game streaming frame rate, game streaming lag, and graphic resolution from flow QoS attributes using an established method by prior work [32]. The process highlighted in the light blue region detects game titles, while the one shown in the orange region continuously measures player activity stages and infers gameplay activity patterns. These two processes are novel in this work and will be discussed in detail.

The first novel process analyzes the packet streams that belong to the first  $N$  seconds of a game streaming flow to classify the respective game title being played, leveraging the packet group characteristics during the game launching stages. As will be discussed in §4.2, we select  $N = 5$  for a balanced classification accuracy and responsiveness. The key technical components in this module including statistical attributes and classification models, will be elaborated soon in §4.2.

Driven by the insights of flow volumetric patterns across player activity stages from game launch to idle, active or passive, our second novel process (orange box in Fig. 6) analyzes packet streams of both upstream and downstream directions in each game streaming flow to continuously label player activity stage. As will be discussed in §4.3, upon receiving a sufficient number of past states and their transitions, this module will make a confident inference on the gameplay activity pattern of the session, which will be useful for estimating the QoE demands of a game streaming session when an inference result from the game title classification module does not fall into the known catalog. In our implementation, we operate both modules independently in parallel at their respective granularities, with the objective of cross-validating the classification results for each candidate game streaming session.

As the output of our method, the gameplay contexts will be correlated with state-of-the-art objective QoE metrics and/or network QoS attributes per streaming session, so that network operators can calibrate their measurement of user experience (which we refer

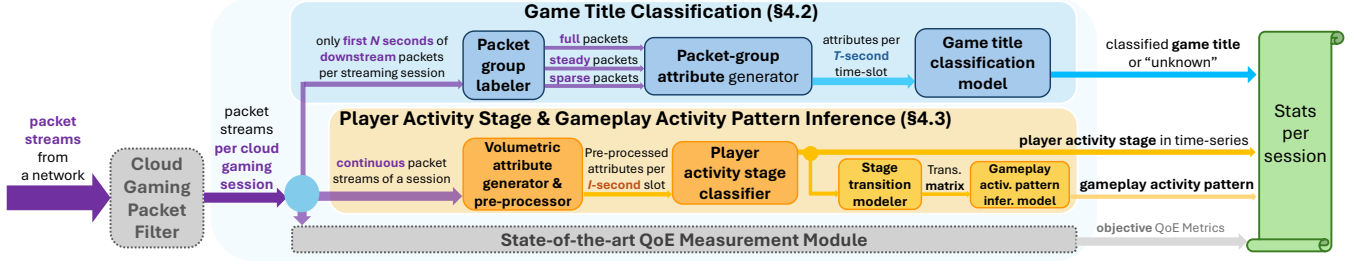


Figure 6: Our real-time network traffic analysis methodology to classify gameplay contexts of cloud game streaming sessions for effective user experience measurement.

to as effective QoE) to identify issues caused by network conditions rather than less demanding game titles and player activities. Later in §5.3, we will show an example of such context-based calibration deployed for our partnered network operator, which corrects mislabeled poor game streaming user experience due to low frame rate and throughput in less demanding scenarios.

## 4.2 Classifying Game Titles

The game title classification process, as shown in the blue region of Fig. 6, classifies the game title of a streaming session based on the first few seconds of its downstream packets delivering the game launch content. In §3.2, by analyzing the arrival time slots and payload sizes, we have empirically clustered three groups of **downstream** packets, namely full, steady, and sparse that arrive in the game launch stages. This drives our development of the game title classification process, which uses well-trained machine learning models on statistical attributes formulated per time slot from the three unique packet groups. Its key technical components are discussed below.

**4.2.1 Labeling downstream packet groups.** The downstream packets during the first  $N$  seconds of each game streaming session will first be categorized into one of the three packet groups (*i.e.*, full, steady and sparse) by their relative payload sizes in the respective arrival time slot of  $T$  seconds. If  $T$  is properly configured, the “full” packets that have the maximum payload size (*e.g.*, 1432 bytes) will consistently arrive in all time slots, as shown in Fig. 3. Additionally, in a  $T$ -second time slot, packets that are not labeled as “full” will be labeled as either “steady” or “sparse” using a majority-voting method based on their payload sizes, with a tunable parameter  $V$  that specifies the numerical range of the payload size variation for a packet compared to its adjacent packets. Specifically, a packet having a payload size within  $\pm V$  difference than its neighbors within the time slot is identified as steady, otherwise sparse. This is driven by our findings in §3.2 that each discrete time slot (*e.g.*, hundreds of milliseconds to tens of seconds) can contain steady packets densely located in one or several narrow bands of payload sizes; and/or sparse packets following a random distribution for their payload sizes. As will be discussed in §4.4.1, in our implementation, the packet group labeler achieves its best performance with the value of  $V$  set to 10% after evaluating a range of options from 1% to 20%.

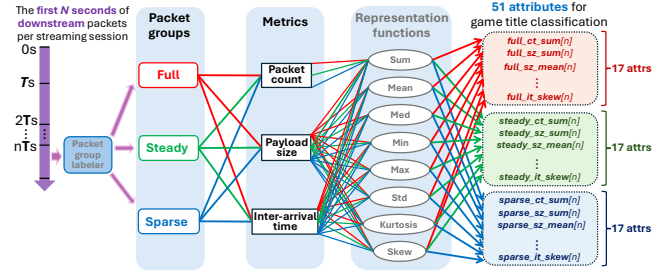


Figure 7: Formulating attributes from the full, steady and sparse packet groups per time slot.

**4.2.2 Attributes of packet groups per time slot.** Packets labeled full, steady, or sparse are then used to generate statistical attributes per packet group for the respective  $T$ -second time slot. Fig. 7 visually depicts the formulation of our per-time-slot attributes describing statistical representations of packet count, payload size, and inter-arrival time for the packets within the “full”, “steady”, or “sparse” group. For example, the attribute annotated as *full\_ct\_sum[0]* shows the total count of packets in the “full” group during the first  $T$ -second time slot of this game launching stage. As will be discussed in §4.4, in our implementation, we process the first 5 seconds of packets in a 1-second time slot to achieve good classification performance.

**4.2.3 Classification model.** For each game streaming session, the attributes generated from all  $T$ -second time slots for the first  $N$  seconds will be processed in a batched manner by our classification model. The model will predict a (popular) game title that exists in our pre-trained dataset or the “unknown” type if a confident result cannot be made. In §4.4, we discuss our evaluation of the classifier trained with common machine learning algorithms, with the best-performing random-forest classifier selected in our implementation.

We acknowledge that flow classification for network operators is often expected to be as early as possible for subsequent network operational tasks [5, 47]. In this work, our game title classification takes only a few seconds (*i.e.*, 5 seconds in our implementation) from the game launch stage, allowing the network operator to seamlessly start subsequent measurement tasks and/or prioritize game streaming flows of high-demanding games prior to the start of actual gameplay.

### 4.3 Classifying Player Activity Stages and Inferring Gameplay Activity Patterns

The second novel process in our method, as outlined in the yellow region of Fig. 6, continuously classifies the player activity stages of a game streaming session using its bidirectional volumetric attributes and infers the activity pattern of this gameplay from the transition pattern of the classified player activity stages.

**4.3.1 Classifying player activity stages.** This process first classifies and tracks the player activity stages as active, passive, or idle over time by processing the bidirectional packet streams of the cloud game streaming session. Standard volumetric attributes for a game streaming session, including packet rate and throughput in both upstream and downstream directions, are computed per time slot of  $I$  second, e.g., one second in our implementation after balancing classification responsiveness and accuracy.

As discussed in §3.3, each game streaming session can exhibit different absolute value ranges of their volumetric attributes across the three player activity stages, while their relative volumetric levels remain consistent regardless of the actual game titles and streaming configurations. Therefore, in each  $I$ -second classification time slot, the four volumetric attributes are converted to the relative fraction compared to their peak value (above a threshold dynamically decided during the game launch) observed from previous slots.

To eliminate the effect of noisy attribute values caused by unexpected short behavior for a certain type of player activity stage, such as accidental mouse movement (i.e., high upstream throughput) when spectating teammates (i.e., passive stage), which will lead to a classification result as active for this slot rather than the correct passive stage, the volumetric attributes are processed by exponential moving average (EMA) to consider the attribute values from the past slots as shown in Equation 1. The weight of the current stage  $\alpha$  is selected as 0.4 from the range of (0,1) for the best accuracy, as will be evaluated soon in §4.4.

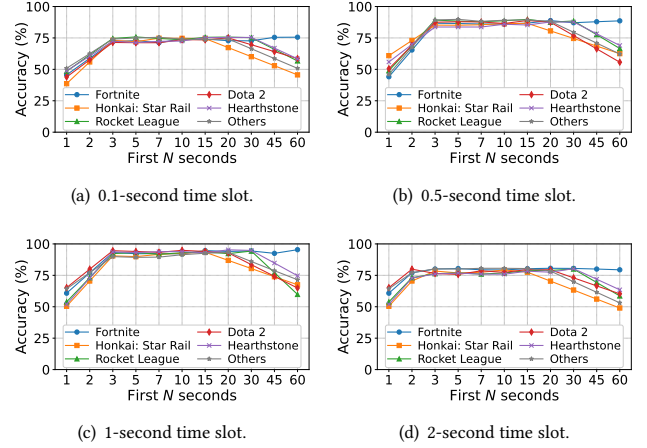
$$\text{attr}_t = \alpha \cdot \text{attr}_t + (1 - \alpha) \cdot \text{attr}_{t-1} \quad (1)$$

A well-trained machine learning classifier developed using the Random Forest algorithm then consumes preprocessed volumetric attributes in their EMA-smoothed relative values to classify player activity stages as idle, passive, or active for each slot accordingly.

**4.3.2 Inferring gameplay activity pattern.** As discussed in §2.1, depending on the game titles, cloud game sessions exhibit two distinctive patterns for their gameplay activities, namely “continuous-play” and “spectate-and-play”. If the fine-grained game title cannot be confidently classified for a game streaming session, our method can provide network operators with visibility into the gameplay activity pattern, revealing coarse-grained profiles and network demands for the respective game streaming sessions.

As quantitatively shown in Fig. 5, the two types of gameplay activity patterns hold different proportions of player activity stages and transition probabilities between these stages. Therefore, our inference model captures the stochastic transition behaviors between player activity stages in a gameplay for classification.

As shown in the process in Fig. 6, the continuously classified player activity stages are also received by the stage transition model. In our implementation, for each game streaming session, we



**Figure 8: Accuracy of game title classification using attributes computed from packets from the first 1 to 60 seconds of game launch with time slot sizes as (a) 0.1 seconds, (b) 0.5 seconds, (c) 1 second, and (d) 2 seconds.**

use a 3x3 matrix with each cell representing the occurrence of transition per slot from one stage to another or retaining itself. The inference model (e.g., using random forest algorithm) makes a prediction based on the nine values of the matrix normalized to their probabilities across time slots within the monitored duration when its confidence level is above a threshold (e.g., 75%) tuned for balanced prediction responsiveness and accuracy.

### 4.4 Parameter and Model Evaluation

After having a comprehensive understanding of our network traffic analysis method for cloud gameplay context classification as depicted in Fig. 6, we now evaluate the options for inference models and key parameters in the pipeline, as well as its overall classification performance for game titles, player activity stages and gameplay activity patterns. The lab evaluation uses our ground-truth dataset (described in §3.1). Similar to prior works [5, 28, 63] in traffic classification, we augment our dataset for larger sample sizes using variation-based statistical techniques, i.e., by synthesizing packet data with randomly varied sizes and arrival times based on the original ground-truth data, especially for classes with fewer samples. We now present evaluation results for key design choices in developing the game title and player activity stage classification processes.

**4.4.1 Evaluating game title classification.** As discussed in §4.2, the game title classification uses three tunable parameters  $N$ ,  $T$  and  $V$ , along with a machine learning classifier, which were selected for our implemented system after lab evaluations.

**Time window  $N$  and time slot  $T$ :** Our attributes are calculated from the packet statistics per time slot  $T$  within the first  $N$  seconds of the game launch stage. A sufficiently large time window  $N$  allows more identical packet statistics to be captured by the attributes. Similarly, a properly selected time slot  $T$  helps the attributes better profile the statistical features while minimizing the impact of mild delays and packet loss that can be diluted in a sufficiently large

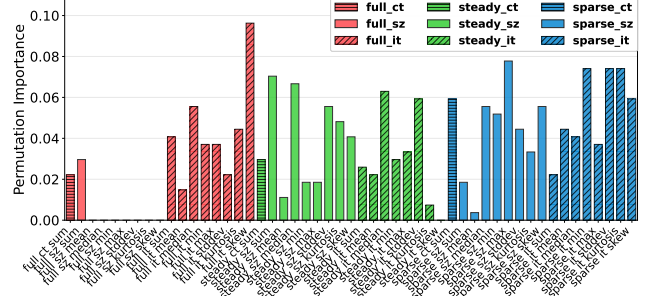
**Table 3: Game title classification accuracy of the best-performing classifier using our specialized packet-group-based attributes vs standard flow volumetric attributes.**

Game title	Accur. (pkt. group)	Accur. (flow vol.)
Baldur's Gate 3	96.9%	82.1%
Cyberpunk 2077	94.9%	81.5%
Genshin Impact	98.0%	89.8%
Honkai: Star Rail	94.4%	85.9%
Call of Duty	95.2%	80.5%
CSGO	97.5%	81.8%
Destiny 2	94.1%	86.0%
DOTA 2	92.7%	84.9%
Fortnite	97.3%	91.5%
Hearthstone	94.8%	88.0%
Overwatch 2	95.6%	85.4%
R6: Siege	94.3%	83.7%
Rocket League	92.8%	83.5%

time slot. Therefore, we tune the best-performing classification models (processes to be discussed later) for different options of  $T$  and  $N$ , with their accuracy for five representative game titles provided in Fig. 8. It is clear that the classification accuracy follows an increasing trend for both  $T$  and  $N$  until a certain value, *i.e.*, 3 seconds for  $T$  and 1 second for  $N$  with over 95% accuracy on average, which are the suitable options to be used in our system implementation. Notably, we use the downstream packets from the first five (instead of three) seconds of a game streaming session to accommodate possible packet arrival delays in the game launch stage. It is worth noting that the implemented values of the tunable parameters  $N$  and  $T$  are selected based on our lab dataset without additionally introduced latency or packet loss. We acknowledge that different values can be selected for extended datasets with varying network conditions, which is beyond our current scope.

**Payload size variation  $V$  for packet group labeling:** While full packets can be simply labeled by those having maximum payload size, to precisely distinguish steady and sparse packets for game title classification, we tune the parameter  $V$  used in our majority-voting algorithm as discussed in §4.2.1, which represents the range of payload size variation allowed among adjacent steady packets in a time slot. Since sessions of a certain game title show consistent packet patterns, we randomly pick one representative session from each of the thirteen games and visually inspect the labeled packet groups using  $V$  between 1% and 20%. From our evaluation results, we conclude that a high range of allowed variation (*i.e.*, 15% and 20%) leads to sparse packets being mistakenly labeled as steady packets, while low variation (*i.e.*, 1% and 5%) mislabels certain steady packets with only slight discrepancy as sparse packets. The observation is consistent across all thirteen games, therefore, we use  $V$  of 10% in our implementation, which yields the best labeling results.

**Machine learning classifier:** To select a suitable machine learning model as our game title classifier, we fine-tune and evaluate three commonly used machine learning algorithms, namely margin-based Support Vector Machine (SVM), distance-based K-Nearest Neighbors (KNN), and Random Forest (RF). After fine-tuning the hyperparameters of each algorithm – such as the regularization

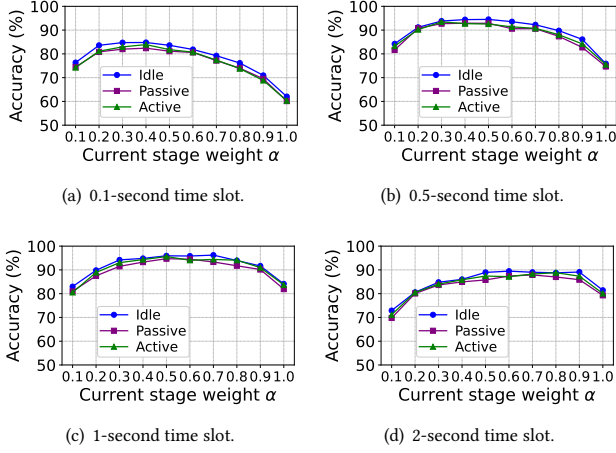
**Figure 9: Importance of the 51 attributes in classifying game titles by the best-performing Random Forest model. Each of them is color-coded by its packet group (full, steady, or sparse) and pattern-coded by the metric (count, size, or inter-arrival time).**

and kernel type for SVM, the number of neighbors and distance metric for KNN, and the number of trees and maximum tree depth for random forest – the classifier that achieves the best overall accuracy (*i.e.*, over 95%) in classifying all game titles is selected, with its accuracy for each game title provided in the second column of Table 3. The details of our hyperparameter tuning and model evaluation are provided in Appendix §C.1.

We evaluate the importance of the attributes consumed by the selected random forest classifier using the permutation importance metric [10], which measures the drop in the accuracy of the model when the values of an evaluated attribute are shuffled randomly. As visually shown in Fig. 9, 43 attributes exhibit certain predictive powers in classifying game titles, while the other eight attributes, including seven for the full packet group and one for the steady group, have little significance with their measured importance being 0. Unsurprisingly, the eight attributes describe less varying statistics among game titles, such as the mean packet sizes of the full packet group, which can be excluded in the classification pipeline to optimize the processing cost [62].

For comparison, we also trained classifiers that take the two standard flow volumetric attributes (*i.e.*, packet rate and throughput) per time interval rather than our well-articulated attributes from the three packet groups. Not surprisingly, the accuracy drops significantly (*e.g.*, more than 10%) for each game title, as shown in the rightmost column of Table 3. We have observed that most misclassified game steaming sessions have label confidence less than 40%. Therefore, in our implementation, sessions with low-confidence results are labeled as “unknown” game title, suggesting that network operators should include more game titles in their databases if they wish, or refer to the coarse-grained gameplay activity patterns for those streaming sessions.

**4.4.2 Evaluating player activity stage classification and gameplay activity pattern inference.** Our second process (§4.3) continuously classifies the player activity stage and infers the gameplay activity pattern of a streaming session. We now discuss the selection of key options, including the time slot  $I$  for player activity stage classification, the weight of the current activity stage  $\alpha$ , the confidence threshold for outputting gameplay activity pattern inference results, and the machine learning classification models.



**Figure 10: Accuracy of player activity stage classification using current stage weights between 0.1 and 1 with time slot sizes of (a) 0.1 seconds, (b) 0.5 seconds, (c) 1 second, and (d) 2 seconds.**

**Time slot  $I$  and current stage weight  $\alpha$ :** In the process that continuously classifies the player activity stage in a cloud game session, as discussed in §4.3.1, the intermediate labels are produced per  $I$ -second time slot and smoothed by an exponential moving average (EMA) function with the weight of the current stage as  $\alpha$ . Therefore, we evaluate the accuracy produced by our best-performing classifiers that use different options of the two parameters, with representative results shown in Fig. 10. We can see in Fig. 10(c) that classification accuracy reaches its best performance with a 1-second time slot, as a small time slot can be overly granular to capture identical volumetric characteristics of the respective player activity stage, while a large time slot can be mixed with multiple stages. Regarding the current stage weight  $\alpha$ , a value between 0.5 and 0.6 has the best overall performance. In our implementation, we use a 1-second time slot with an  $\alpha$  value of 0.5 to achieve classification robustness to short noises in streaming volumetric profiles.

**Confidence threshold for gameplay activity pattern inference:** In our second process, we continuously infer the gameplay activity pattern based on the stage transition/retention rates upon each newly classified stage at a time slot, until the confidence level of the inferred pattern exceeds a certain threshold. To balance accuracy (by using more data from more time slots) and responsiveness (*i.e.*, shorter inference time), we evaluate the per-session average accuracy and time needed to generate a confident inference result above confidence thresholds from 0% up to 95% for continuous-play and spectate-and-play games, respectively, using the optimal settings for other parameters. By using very low confidence thresholds (*e.g.*, 0% to 40%), the model tends to generate results very early (*e.g.*, 1 to 30 seconds) but they are highly inaccurate (*i.e.*, less than 50% accuracy). On the other hand, overly high thresholds (*e.g.*, 95%) may lead to inference results not being generated until close to the end of sessions, which are not useful for real-time monitoring and performance boost. Therefore, we choose 75% as the confidence threshold since it produces results with around 90% accuracy for both types of games while only needing five minutes on average, at which time the player has just begun the gameplay.

**Table 4: Player activity stage classification (by time slot) and gameplay activity pattern inference (by session) accuracy for continuous-play and spectate-and-play games.**

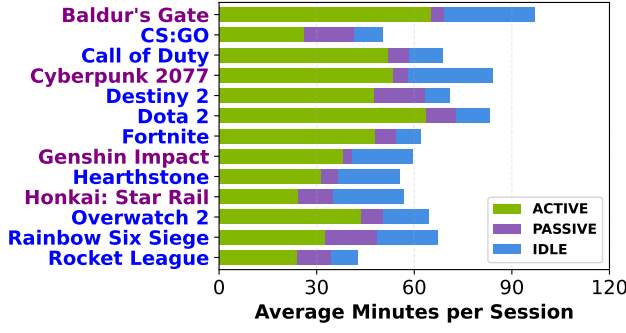
Gameplay actv. pattern	Accur.	Player actv. stage	Accur.
Continuous-play	95.7%	Active	94.1%
		Passive	92.5%
		Idle	97.6%
Spectate-and-play	97.2%	Active	96.8%
		Passive	95.9%
		Idle	98.4%

**Machine learning classifiers:** In this process, we have developed two machine learning models to classify the player activity stage in a cloud game session using bidirectional volumetric attributes of the streaming session and to infer the overall gameplay activity pattern. We follow a similar process as described in §4.4.1, evaluating commonly used machine learning algorithms and their hyperparameters to select the best-performing models. To classify player activity stages using volumetric attributes, both throughput and packet rate demonstrate similar importance. The downstream attributes contribute more to the classification of the idle stage, while the upstream attributes are more important for the active and passive stages. To classify gameplay activity patterns using the stage transition behaviors, the attributes that capture transitions from active to idle stages and from idle to passive stages are the most important ones. The details of our model and attribute evaluation are provided in Appendix §C.2. The accuracies of the best-performing models for both classification tasks are provided in Table 4. We observe that decent accuracy (*e.g.*, mostly over 95%) is achieved for all classification labels in both player activity stages (*i.e.*, active, passive and idle) and gameplay activity pattern types (*i.e.*, continuous-play and spectate-and-play), meeting the requirement for the field deployment in our partnered ISP network as discussed next.

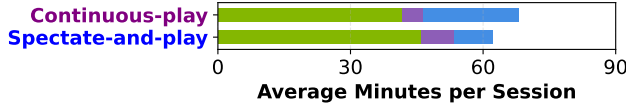
## 5 Measuring Cloud Gaming Contexts at a Large Scale

We have implemented and deployed our cloud gaming context measurement method in our partner ISP that hosts NVIDIA's GeForce NOW cloud gaming servers for our geography, providing valuable insights to precisely identify and troubleshoot user experience degradations that are indeed caused by network factors rather than mistaking low-throughput sessions caused by game titles and gameplay activity patterns as having poor experience. The real-time inference of game contexts is also meant to help our partner ISP in assessing the efficacy of their prioritization techniques in assuring cloud gaming experience over their 5G mobile broadband network.

In this section, we discuss representative insights from a three-month (1 December 2024 to 1 March 2025) measurement of cloud game contexts. The results are selected to comply with commercial confidentiality restrictions and our university ethical approval conditions as provided in Appendix §A. One month before our field deployment, we evaluated our game title classification method in the deployment network by validating our classified game titles



(a) Average durations of player activity stages across 13 popular game titles.



(b) Average durations of player activity stages across 2 activity patterns.

**Figure 11: Average number of minutes spent in active, passive, and idle player activity stages per cloud gaming session for (a) 13 popular game titles and (b) 2 gameplay activity pattern types.**

with the offline cloud game server logs that were generated after the conclusion of each game session and are not available to ISPs in daily operations. The evaluation results showed that our method can effectively classify the game titles for the thirteen popular games with an overall accuracy above 95%, consistent with our lab evaluation results. As for the player activity stages, there is no record from cloud game server logs for us to validate outside the lab environment.

We now demonstrate how our measurement of cloud game contexts can help network operators understand player activity profiles across game titles (§5.1) which can lead to different expectations on game streaming quality; benchmark bandwidth demands that are unique to game titles and gameplay activity patterns (§5.2); and measure effective cloud gameplay experience by corroborating gameplay contexts with objective QoE and network QoS metrics (§5.3).

### 5.1 Player Activity Stages across Cloud Game Contexts

Our measurement of cloud gaming contexts provides network operators with player activity profiles of streaming sessions per cloud game title and genre, which serve as references for network operators to dynamically provision network resources, *e.g.*, allocate 5G eMBB slices with prioritized QoS profiles by our partnered ISP with an expected session duration and slice capacity, upon detecting a newly commenced game streaming session.

We report the average duration of each game streaming session and each player activity stage (*i.e.*, active, passive, and idle) across the 13 popular game titles in Fig. 11(a). The results are aggregated

for the entire three-month deployment period. We can clearly observe significant differences in both streaming durations and player activity stages.

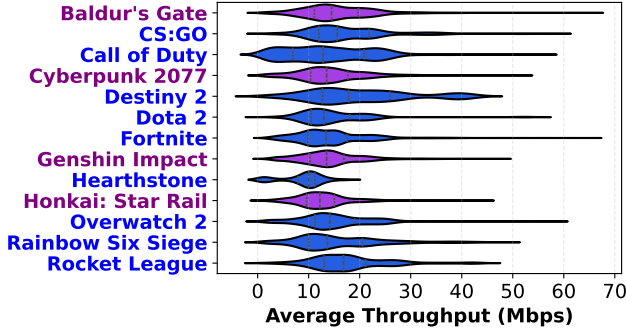
For the session duration, nine games have long average durations near or above 1 hour. Those game titles with long session durations are either role-playing games that involve extensive dialogue-based content like Baldur's Gate (95 minutes) and Cyberpunk 2077 (82 minutes), or MOBA, Sports and Shooter games like Dota 2 and Rainbow Six Siege that have either long matchmaking or match durations. Rocket League and CS:GO are with short per-match duration and have the shortest average game streaming durations among the popular games. Depending on the gameplay content, each game title has its unique distribution of player activity stages, such as a relatively larger fraction (*e.g.*, 25 – 55%) of duration in idle and passive stages for Baldur's Gate, Cyberpunk 2077, Honkai: Star Rail and Hearthstone due to their extensive static content or unskippable dialogues. On the contrary, Fortnite and Dota 2 have most of their gameplay duration in the active stage.

For game sessions that do not belong to the thirteen popular titles and thus are coarsely categorized as either continuous-play or spectate-and-play games, from Fig. 11(b), we can also observe different expectations on the streaming session durations and the composition of player activity stages. Continuous-play games, which are often role-playing, have a longer session duration with a significant fraction of time spent in idle (26%) stages. Spectate-and-play games are often match-based shooter, card, strategy or sports games without lengthy in-game static scenes and dialogues, therefore, have a larger fraction of time spent in the active stage.

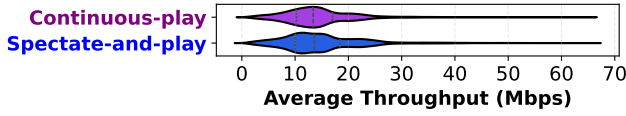
### 5.2 Network Bandwidth Demands across Cloud Game Contexts

Among the network resource provisioning tasks, ensuring sufficient bandwidth for multimedia streaming sessions to support good quality-of-experience (QoE) metrics is often very challenging for network operators who operate with constrained bandwidth resources for massive subscribers, especially in high-density regions (*e.g.*, a 5G base station serving a metropolitan area) during peak hours. Cloud game streaming sessions with different contexts exhibit various network bandwidth demands, determined not only by their streaming settings but also by the gameplay activity patterns that are inherently different across game titles.

The average bandwidth demands of streaming sessions that belong to the thirteen popular game titles and the two types of gameplay activity patterns are provided in Fig. 12(a) and Fig. 12(b) as representative examples, respectively. According to the throughput distributions of each game title, we can observe unique ranges across different games. For example, high-demand games like Baldur's Gate and Fortnite have their maximum session-level average throughput up to 68 Mbps; while low-demand games like Hearthstone require only 20 Mbps for gameplay sessions with the best streaming settings. There are many streaming sessions with very low streaming throughput (*e.g.*, less than 1 Mbps), which likely suffer from poor user experience due to constrained network conditions with large game streaming lag mostly over 70ms from the measured objective QoE metrics. Therefore, they are excluded from our analysis of network bandwidth demands.



(a) Average throughput per session across 13 popular game titles.



(b) Average throughput per session across 2 gameplay activity patterns.

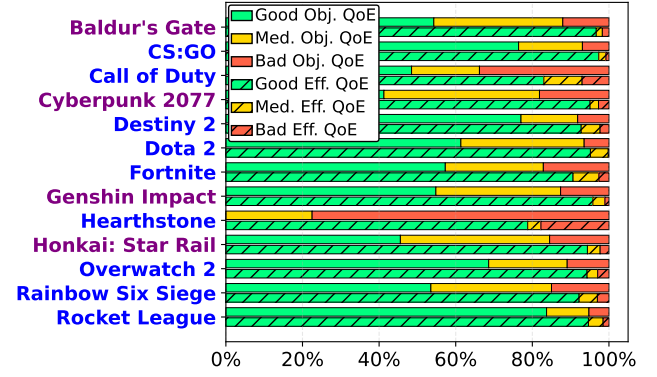
**Figure 12: Average throughput per game streaming session for (a) 13 popular game titles and (b) 2 gameplay activity patterns.**

Depending on the game streaming settings such as graphic resolutions and user setups (e.g., PC or mobile devices), we have observed several (two to four) clusters of bandwidth demands for each game title. For example, as shown in Fig. 12(a), *Destiny 2* has three clusters of bandwidth in the ranges of 8 – 18 Mbps, 20 – 30 Mbps, and 35 – 47 Mbps, respectively, each of which is mapped to a group of graphic resolution settings depending on the player devices. Detection of the player devices has been done in prior work [32] and is not discussed in this paper. In Fig. 12(b), similar throughput distributions, which are dominantly in the range of 10 – 25 Mbps with the highest bandwidth demand exceeding 65 Mbps for streaming in very high graphic resolutions, are observed for both continuous-play and spectate-and-play games, with slightly higher bandwidth demands seen for spectate-and-play games.

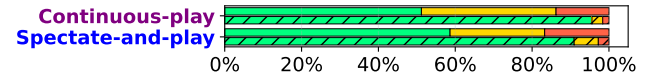
With a concrete understanding of the throughput requested for streaming sessions of certain game titles, our partnered ISP is better equipped to prioritize premium users with the appropriate QoS (bandwidth) profiles for service quality assurance without over-provisioning the constrained network resources.

### 5.3 Measuring Effective Game Streaming Experience with Contexts

Most importantly, by combining our measurement of cloud gaming contexts with objective QoE metrics (e.g., streaming frame rate) and standard network QoS metrics (i.e., latency, packet drop rate, and throughput), our partnered network operator is now able to measure the effective game streaming experience as perceived by subscribers, without mislabeling streaming sessions of certain game titles or periods of gameplay in idle/passive player activity stages that inherently require low streaming frame rate and network bandwidth as degraded user experience.



(a) Objective/effective session QoE across 13 popular game titles.



(b) Objective/effective session QoE across 2 gameplay activity pattern types.

**Figure 13: Average fraction of gameplay duration per game streaming session in good, medium and bad user experience for both objective QoE and effective QoE levels, i.e., before and after calibration with gameplay contexts, across (a) 13 popular game titles and (b) 2 gameplay activity pattern types.**

**Calibrating user experience measurement:** In the three-month deployment, our measurement system is operated in parallel with an existing network observability module (the bottom gray box in Fig. 6) operated by our partnered ISP, which labels the objective quality-of-experience (QoE) of game streaming sessions into three levels as bad, medium or good, by mapping the measured frame rate, throughput, latency and packet drop rate of streaming flows into the expected value ranges that are internally maintained in the observability system. For example, a session with a streaming frame rate lower than 30 FPS and/or a throughput below 8 Mbps will be labeled with bad objective QoE. Augmented with the results produced in this work, the expected value ranges of the objective QoE levels are empirically calibrated according to our real-time measurement of cloud game contexts, annotated as **effective QoE levels**. After calibration, reasonable drops in **streaming frame rate** and **throughput** due to less demanding game titles (e.g., *Hearthstone* and *Honkai: Star Rail*) or player activity stages (i.e., idle and passive), which intrinsically have less dynamic graphics or infrequent scene changes and hence require lower network resources, will not be mislabeled with degraded QoE levels. The expected value ranges of **low latency** and **packet drop rate** between players and cloud servers remain unchanged from the objective QoE levels to the effective QoE levels after the context-based calibration.

In Fig. 13, we show how cloud gaming contexts help our partnered ISP reduce falsely labeled sessions with degraded game streaming user experience, i.e., by transitioning from objective QoE to effective QoE levels. For simplicity, we show the overall QoE levels of each game streaming session, representing the majority QoE labels measured for the respective session in real-time. It is clear

from Fig. 13(a) that all game titles have significantly more sessions with good QoE levels after being calibrated with the gameplay contexts. For example, all game streaming sessions of Hearthstone, a low-demanding card game, have medium or bad objective QoE levels, 80% of which are corrected to the good effective QoE level. Another example is Cyberpunk 2077, which has 40% and 16% of its streaming sessions labeled as medium and bad for the objective QoE levels. Although this game title has high demands for network bandwidth and frame rate during the active stage, a large fraction of its player activity stages are passive and idle, leading to 95% sessions with good effective QoE levels after calibration.

For streaming sessions of less popular game titles that are classified as continuous-play or spectate-and-play games, in Fig. 13(b), we can also observe that about half of the total sessions are corrected from medium or poor in objective QoE levels to good in effective QoE levels.

By calibrating the mapping criteria of game streaming sessions with our real-time classification of cloud gameplay contexts for effective QoE measurement, our partnered network operator is able to precisely identify the groups of under-performing game streaming sessions in a much smaller (and therefore manageable) quantity, such as those from clients connected to its 5G home broadband network via a poorly configured mobile cell. It enables precise troubleshooting without mislabeling streaming sessions with low network demands as poor user experience. In addition, knowing the effective QoE after our context-based calibration helps the ISP to reactively improve under-performing game streaming sessions served by its 5G broadband network by enforcing QoS profiles for the respective sessions via network slices with higher capacity.

## 6 Related Work

**Network traffic analysis of cloud gaming services:** As an emerging type of multimedia streaming service that exhibits high demands on network conditions, cloud gaming services have been analyzed in prior works for their network traffic characteristics, including flow anatomy and volumetric profiles. For example, Lyu *et al.* [32] discussed the use of network flows during entire user sessions, from platform administration to server selection and gameplay, which serve as indicative signatures for network operators to detect cloud gaming sessions played on different types of user setups by their broadband/mobile subscribers. The works in [13, 18, 34] analyzed the bandwidth demands of cloud gaming sessions with different levels of streaming quality settings. Furthermore, objective quality-of-experience metrics (e.g., user input lag, graphic resolution, and streaming frame rates) have also been derived from network quality-of-service (QoS) attributes of game streaming flows [6, 32, 44]. For the network observability industry, it has been highlighted that purely using objective QoE metrics such as frame rate cannot effectively measure the quality of game streaming perceived by players, as gameplay contexts, including game titles and player activities, can significantly vary the expected/requested QoE and network QoS levels for good user experience [3, 14, 22, 25, 27, 49, 66]. To bridge this gap, we develop a traffic analysis method to classify the contexts of cloud gameplay, enabling network operators to effectively measure user-perceived streaming quality when correlating with the objective QoE metrics.

**Measuring multimedia services with contexts:** Gaining visibility into the coarse categories of multimedia streaming contexts has become important for the network observability industry to effectively measure user-perceived streaming quality as delivered by networks [7, 12, 21, 42, 54, 60, 65]. Prior works have developed methods to classify contexts in popular types of multimedia services such as video streaming [1, 11, 64, 68], live streaming [30, 37], video conferencing [35, 39, 51], voice calling [19, 26], online gaming [36, 50], and virtual reality applications [17, 33]. For example, the works in [15, 16] focus on the contexts in online gaming that vary the expected network conditions for a good gameplay experience including game titles, genres, the number of players, peripheral status and team communication model. In [17, 33], the authors measured the network demands for a good VR experience that depend on the user activity status and the number of surrounding users. This work is specialized in classifying the contexts (*i.e.*, game titles, gameplay activity patterns, and player activity stages) of cloud game streaming sessions. As demonstrated in a large-scale deployment at an ISP hosting NVIDIA's cloud gaming servers, our method, combined with objective QoE measurement, holds unique value for network operators to better support this highly demanding service by effectively measuring the user-perceived streaming quality.

## 7 Conclusion

In this paper, we have developed a real-time network traffic analysis method that classifies the gameplay contexts of cloud game streaming sessions for network operators to effectively measure the user experience of this emerging application that exhibits high demands on network resources. We first systematically analyze unique network traffic characteristics of cloud game streaming sessions across various gameplay contexts, including packet group profiles for different game titles and flow volumetric behaviors for player activity stages. Driven by these insights, our real-time network traffic analysis method is developed with two novel processes leveraging machine learning models to classify game titles during the game launching stage and continuously measure player activity stages, respectively. Our classification method is implemented in the network observability platform operated for an ISP hosting NVIDIA's cloud gaming servers with representative insights over a three-month period discussed. By integrating with the existing QoE and QoS metrics of cloud gaming sessions, our method helps the ISP significantly reduce the instances of poor user experience in cloud gameplay that are incorrectly labeled due to less demanding game titles and inactive player activity stages.

## Acknowledgments

We thank our anonymous shepherd and the six reviewers for their comprehensive and insightful feedback. We thank Maheesha Perera and Craig Russell from Canopus Networks Pty Ltd, and Jeremy Hall, Matthew Kocoski and Dylan Ryan from Pentanet Ltd for their system engineering and infrastructure support. We thank our Masters Project students Qixin Lu, Bingxian Min, Chi Zhang and Haoran Zheng, who participated in the lab dataset collection. This work is supported by Australian Government's National Industry PhD Program award reference number 35063 and Cooperative Research Centres Projects (CRC-P) Grant CRCPXIV000099.

## References

- [1] Vivek Adarsh, Michael Nekrasov, Udit Paul, Alex Ermakov, Arpit Gupta, Morgan Vigil-Hayes, Ellen Zegura, and Elizabeth Belding. 2021. Too Late for Playback: Estimation of Video Stream Quality in Rural and Urban Contexts. In *Proc. Passive and Active Measurement*. Virtual Event, 141–157.
- [2] Amazon. 2025. Amazon Luna Cloud Gaming. <https://luna.amazon.com/> Accessed: 2025-04-20.
- [3] Congkai An, Huanhuan Zhang, Shibo Wang, Jingyang Kang, Anfu Zhou, Liang Liu, Huadong Ma, Zili Meng, Delei Ma, Yusheng Dong, and Xiaogang Lei. 2025. Tooth: Toward Optimal Balance of Video QoE and Redundancy Cost by Fine-Grained FEC in Cloud Gaming Streaming. In *Proc. USENIX NSDI*. Philadelphia, PA, USA, 635–651.
- [4] Franck Aumont, Frédérique Humbert, Christoph Neumann, Charles Salmon-Legagneur, and Charline Taibi. 2021. Dissecting Cloud Game Streaming Platforms regarding the Impacts of Video Encoding and Networking Constraints on QoE. In *Proc. Workshop on Game Systems*. Istanbul, Turkey, 13–19.
- [5] Rushi Jayeshkumar Babaria, Minzhao Lyu, Gustavo Batista, and Vijay Sivaraman. 2025. FastFlow: Early Yet Robust Network Flow Classification using the Minimal Number of Time-Series Packets. *Proc. ACM Meas. Anal. Comput. Syst.* 9, 2, Article 23 (June 2025).
- [6] Carlos Baena, O.S. Peñaherrera-Pulla, Raquel Barco, and Sergio Fortes. 2023. Measuring and Estimating Key Quality Indicators in Cloud Gaming Services. *Computer Networks* 231, Article 109808 (2023).
- [7] Alcardo Alex Barakabitze, Nabajeet Barman, Arslan Ahmad, Saman Zadtootaghaj, Lingfen Sun, Maria G. Martini, and Luigi Atzori. 2020. QoE Management of Multimedia Streaming Services in Future Networks: A Tutorial and Survey. *IEEE Communications Surveys & Tutorials* 22, 1 (2020), 526–565.
- [8] Sandeepa Bhuyan, Shulin Zhao, Ziyu Ying, Mahmut T. Kandemir, and Chita R. Das. 2022. End-to-end Characterization of Game Streaming Applications on Mobile Platforms. *Proc. ACM Meas. Anal. Comput. Syst.* 6, 1, Article 10 (Feb. 2022).
- [9] Blacknut. 2023. Operators Are Opening Up 5G Networks to Application Developers to Drive Innovation. <https://www.blacknut.biz/press-release/operators-are-opening-up-5-g-networks-to-application-developers-to-drive-innovation> Accessed: 2025-05-06.
- [10] Leo Breiman. 2001. Random Forests. *Machine Learning* 45 (Oct. 2001), 5–32.
- [11] Francesco Bronzino, Paul Schmitt, Sara Ayoubi, Hyojoon Kim, Renata Teixeira, and Nick Feamster. 2021. Traffic Refinery: Cost-Aware Data Representation for Machine Learning on Network Traffic. *Proc. ACM Meas. Anal. Comput. Syst.* 5, 3, Article 40 (Dec. 2021).
- [12] Francesco Bronzino, Paul Schmitt, Sara Ayoubi, Guilherme Martins, Renata Teixeira, and Nick Feamster. 2019. Inferring Streaming Video Quality from Encrypted Traffic: Practical Models and Deployment Experience. *Proc. ACM Meas. Anal. Comput. Syst.* 3, 3, Article 56 (Dec. 2019).
- [13] Marc Carrascosa and Boris Bellalta. 2022. Cloud-gaming: Analysis of Google Stadia traffic. *Computer Communications* 188 (April 2022), 99–116.
- [14] Marcos Carvalho, Daniel Soares, and Daniel Fernandes Macedo. 2024. QoE Estimation Across Different Cloud Gaming Services Using Transfer Learning. *IEEE Transactions on Network and Service Management* 21, 6 (2024), 5935–5946.
- [15] Xianhui Che and Barry Ip. 2012. Packet-level Traffic Analysis of Online Games from the Genre Characteristics Perspective. *Journal of Network and Computer Applications* 35, 1 (2012), 240–252.
- [16] Kuan-Ta Chen and Chin-Laung Lei. 2006. Network Game Design: Hints and Implications of Player Interaction. In *Proc. ACM SIGCOMM Workshop on Network and System Support for Games*. Singapore.
- [17] Ruizhi Cheng, Nan Wu, Matteo Varvello, Songqing Chen, and Bo Han. 2022. Are We Eady for Metaverse? A Measurement Study of Social Virtual Reality Platforms. In *Proc. ACM IMC*. Nice, France, 504–518.
- [18] Andrea Di Domenico, Gianluca Perna, Martino Trevisan, Luca Vassio, and Danilo Giordano. 2021. A Network Analysis on Cloud Gaming: Stadia, GeForce Now and PSNow. *Network* 1, 3 (2021), 247–260.
- [19] Mario Di Mauro and Antonio Liotta. 2020. An Experimental Evaluation and Characterization of VoIP Over an LTE-A Network. *IEEE Transactions on Network and Service Management* 17, 3 (2020), 1626–1639.
- [20] Ericsson. 2023. 5G Opportunities in the Cloud Gaming Market. <https://bit.ly/ericsson-cg-report> Accessed: 2025-04-20.
- [21] Alessandro Finamore, Marco Mellia, Maurizio M. Munafò, Ruben Torres, and Sanjay G. Rao. 2011. YouTube Everywhere: Impact of Device and Infrastructure Synergies on User Experience. In *Proc. ACM IMC*. Berlin, Germany, 345–360.
- [22] Sebastian Flinck Lindström, Markus Wetterberg, and Niklas Carlsson. 2020. Cloud Gaming: A QoE Study of Fast-paced Single-player and Multiplayer Gaming. In *Proc. IEEE/ACM International Conference on Utility and Cloud Computing*. Leicester, UK, 34–45.
- [23] Philippe Graff, Xavier Marchal, Thibault Chole, Bertrand Mathieu, and Olivier Festor. 2023. Efficient Identification of Cloud Gaming Traffic at the Edge. In *Proc. IEEE/IFIP Network Operations and Management Symposium*. Miami, FL, USA, 1–10.
- [24] Philippe Graff, Xavier Marchal, Thibault Chole, Stéphane Tuffin, Bertrand Mathieu, and Olivier Festor. 2021. An Analysis of Cloud Gaming Platforms Behavior under Different Network Constraints. In *Proc. International Conference on Network and Service Management*. Izmir, Turkey, 551–557.
- [25] Zhaoyuan He, Yifan Yang, Shuoze Li, Lili Qiu, Diyu Dai, and Yuqing Yang. 2024. VIGOR: Reviving Cloud Gaming Sessions. *Proc. ACM Netw.* 2, CoNEXT4, Article 32 (Nov. 2024).
- [26] Jan Holub, Michael Wallbaum, Noah Smith, and Hakob Avetisyan. 2018. Analysis of the Dependency of Call Duration on the Quality of VoIP Calls. *IEEE Wireless Communications Letters* 7, 4 (2018), 638–641.
- [27] Hassan Iqbal, Ayesha Khalid, and Muhammad Shahzad. 2021. Dissecting Cloud Gaming Performance with DECAF. *Proc. ACM Meas. Anal. Comput. Syst.* 5, 3, Article 31 (Dec. 2021).
- [28] Xi Jiang, Shinan Liu, Aaron Gember-Jacobson, Arjun Nitin Bhagoji, Paul Schmitt, Francesco Bronzino, and Nick Feamster. 2024. NetDiffusion: Network Data Augmentation Through Protocol-Constrained Traffic Generation. *Proc. ACM Meas. Anal. Comput. Syst.* 8, 1, Article 11 (Feb. 2024).
- [29] Joël Roman Ky, Bertrand Mathieu, Abdelkader Lahmadi, and Raouf Boutaba. 2023. ML Models for Detecting QoE Degradation in Low-Latency Applications: A Cloud-Gaming Case Study. *IEEE Transactions on Network and Service Management* 20, 3 (2023), 2295–2308.
- [30] Frank Loh, Kathrin Hildebrand, Florian Wamser, Stefan Geißler, and Tobias Hoffeld. 2023. Machine Learning Based Study of QoE Metrics in Twitch.tv Live Streaming. In *Proc. IEEE/IFIP Network Operations and Management Symposium*. Miami, FL, USA, 1–7.
- [31] Minzhao Lyu. 2025. Multimedia Network Traffic Dataset. <https://minzhaolyu.github.io/dataset/MultimediaNetworkTrafficDataset> Accessed: 2025-09-09.
- [32] Minzhao Lyu, Sharat Chandra Madanapalli, Arun Vishwanath, and Vijay Sivaraman. 2024. Network Anatomy and Real-Time Measurement of Nvidia GeForce NOW Cloud Gaming. In *Proc. Passive and Active Measurement*. Virtual Event.
- [33] Minzhao Lyu, Rahul Dev Tripathi, and Vijay Sivaraman. 2023. MetaVRadar: Measuring Metaverse Virtual Reality Network Activity. *Proc. ACM Meas. Anal. Comput. Syst.* 7, 3 (Dec. 2023).
- [34] Minzhao Lyu, Yifan Wang, and Vijay Sivaraman. 2024. Do Cloud Games Adapt to Client Settings and Network Conditions?. In *Proc. IFIP/IEEE Networking*. Thessaloniki, Greece.
- [35] Kyle MacMillan, Tarun Mangla, James Saxon, and Nick Feamster. 2021. Measuring the Performance and Network Utilization of Popular Video Conferencing Applications. In *Proc. ACM IMC*. Virtual Event, 229–244.
- [36] Sharat Chandra Madanapalli, Hassan Habibi Gharakheili, and Vijay Sivaraman. 2022. Know Thy Lag: In-Network Game Detection and Latency Measurement. In *Proc. Passive and Active Measurement*. Virtual Event.
- [37] Sharat Chandra Madanapalli, Alex Mathai, Hassan Habibi Gharakheili, and Vijay Sivaraman. 2021. ReCLive: Real-Time Classification and QoE Inference of Live Video Streaming Services. In *Proc. IEEE/ACM International Symposium on Quality of Service*. Tokyo, Japan, 1–7.
- [38] Xavier Marchal, Philippe Graff, Joël Roman Ky, Thibault Chole, Stéphane Tuffin, Bertrand Mathieu, and Olivier Festor. 2023. An Analysis of Cloud Gaming Platforms Behaviour Under Synthetic Network Constraints and Real Cellular Networks Conditions. *J. Netw. Syst. Manage.* 31, 2 (Feb. 2023).
- [39] Oliver Michel, Satadal Sengupta, Hyojoon Kim, Ravi Netravali, and Jennifer Rexford. 2022. Enabling Passive Measurement of Zoom Performance in Production Networks. In *Proc. ACM IMC*. Nice, France, 244–260.
- [40] Microsoft. 2024. Xbox Game Energy Efficiency Essentials. <https://learn.microsoft.com/en-us/gaming/sustainability/xbox-game-energy-efficiency-essentials> Accessed: 2025-04-25.
- [41] Microsoft. 2025. Xbox Cloud Gaming (Beta) on Xbox.com. <https://www.xbox.com/en-us/play> Accessed: 2025-04-20.
- [42] Ricky K.P. Mok, Edmond W.W. Chan, Xiapu Luo, and Rocky K.C. Chang. 2011. Inferring the QoE of HTTP Video Streaming from User-viewing Activities. In *Proc. ACM SIGCOMM Workshop on Measurements up the Stack*. Toronto, Ontario, Canada, 31–36.
- [43] Philipp Moll, Mathias Lux, Sebastian Theuermann, and Hermann Hellwagner. 2018. A Network Traffic and Player Movement Model to Improve Networking for Competitive Online Games. In *Proc. Annual Workshop on Network and Systems Support for Games*. Amsterdam, Netherlands, 1–6.
- [44] Dorian Monaco, Alessio Sacco, Daniele Spina, Francesco Strada, Andrea Bottino, Tania Cerquitelli, and Guido Marchetto. 2025. Real-time Latency Prediction for Cloud Gaming Applications. *Computer Networks* 264, Article 111235 (2025).
- [45] NVIDIA. 2025. GeForce NOW | The Next Generation in Cloud Gaming | NVIDIA. <https://www.nvidia.com/geforce-now/> Accessed: 2025-04-20.
- [46] NVIDIA Support. 2023. How can I adjust my network firewall to work with GeForce NOW? [https://nvidia.custhelp.com/app/answers/detail/a\\_id/5482/~how-can-i-adjust-my-network-firewall-to-work-with-geforce-now%3F](https://nvidia.custhelp.com/app/answers/detail/a_id/5482/~how-can-i-adjust-my-network-firewall-to-work-with-geforce-now%3F) Accessed: 2024-07-22.
- [47] Julien Piet, Dubem Nwoji, and Vern Paxson. 2023. GGFAST: Automating Generation of Flexible Network Traffic Classifiers. In *Proc. ACM SIGCOMM*. New York, NY, USA, 850–866.

- [48] Saeed Shafiee Sabet, Steven Schmidt, Saman Zadtootaghaj, Carsten Griwodz, and Sebastian Möller. 2020. Delay sensitivity classification of cloud gaming content. In *Proc. ACM International Workshop on Immersive Mixed and Virtual Environment Systems*. Istanbul, Turkey, 25–30.
- [49] Avinab Saha, Yu-Chih Chen, Chase Davis, Bo Qiu, Xiaoming Wang, Rahul Gowda, Ioannis Katsavounidis, and Alan C. Bovik. 2023. Study of Subjective and Objective Quality Assessment of Mobile Cloud Gaming Videos. *IEEE Transactions on Image Processing* 32 (2023), 3295–3310.
- [50] Steven Schmidt, Saman Zadtootaghaj, Saeed Shafiee Sabet, and Sebastian Möller. 2021. Modeling and Understanding the Quality of Experience of Online Mobile Gaming Services. In *Proc. International Conference on Quality of Multimedia Experience*. Montreal, QC, Canada, 157–162.
- [51] Taveesh Sharma, Tarun Mangla, Arpit Gupta, Junchen Jiang, and Nick Feamster. 2023. Estimating WebRTC Video QoE Metrics Without Using Application Headers. In *Proc. ACM IMC*. Montreal QC, Canada, 485–500.
- [52] Alireza Shirmarz, Fábio Luciano Verdi, Suneet Kumar Singh, and Christian Esteve Rothenberg. 2024. From Pixels to Packets: Traffic Classification of Augmented Reality and Cloud Gaming. In *Proc. IEEE International Conference on Network Softwarization*. Saint Louis, MO, USA, 195–203.
- [53] Singtel. 2025. Singtel and Tencent Games to launch Honor of Kings · Cloud in Singapore with global-first 5G network slicing, available on TxStore. <https://www.singtel.com/about-us/media-centre/news-releases/singtel-and-tencent-games-to-launch-honor-of-kings-cloud> Accessed: 2025-09-05.
- [54] Lea Skorin-Kapov, Martín Varela, Tobias Hoßfeld, and Kuan-Ta Chen. 2018. A Survey of Emerging Concepts and Challenges for QoE Management of Multimedia Services. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 2s, Article 29 (May 2018).
- [55] Ivan Slivar, Kresimir Bacic, Irena Orsolic, Lea Skorin-Kapov, and Mirko Suznjevic. 2022. CGD: a Cloud Gaming Dataset with Gameplay Video and Network Recordings. In *Proc. ACM MMSys*. Athlone, Ireland, 272–278.
- [56] Ivan Slivar, Mirko Suznjevic, and Lea Skorin-Kapov. 2018. Game Categorization for Deriving QoE-Driven Video Encoding Configuration Strategies for Cloud Gaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 3s, Article 56 (June 2018).
- [57] Sony. 2025. Cloud Game Streaming (Beta) on PS Portal. <https://bit.ly/ps5-cloud-streaming> Accessed: 2025-04-20.
- [58] The Linux Foundation. [n. d.]. Quality on Demand — Camara Project. <https://camaraproject.org/quality-on-demand/> Accessed: 2025-05-06.
- [59] Unity. 2025. Reduce Rendering Work on the CPU or GPU. <https://docs.unity3d.com/Manual/OptimizingGraphicsPerformance.html> Accessed: 2025-04-25.
- [60] Matteo Varvello, Hyunseok Chang, and Yasir Zaki. 2022. Performance Characterization of Videoconferencing in the Wild. In *Proc. ACM IMC*. Nice, France, 261–273.
- [61] Vodafone. 2023. Vodafone and Ericsson Successfully Trial Cloud Gaming on 5G Standalone Network Slice. <https://www.vodafone.co.uk/newscentre/press-release/vodafone-and-ericsson-successfully-trial-cloud-gaming-on-5g-standalone-network-slice/> Accessed: 2025-05-06.
- [62] Gerry Wan, Shinan Liu, Francesco Bronzino, Nick Feamster, and Zakir Durumeric. 2025. CATO: End-to-End Optimization of ML-Based Traffic Analysis Pipelines. In *Proc. USENIX NSDI*. Philadelphia, PA, USA, 1523–1540.
- [63] Chao Wang, Alessandro Finamore, Pietro Michiardi, Massimo Gallo, and Dario Rossi. 2024. Data Augmentation for Traffic Classification. In *Proc. Passive and Active Measurement*. Virtual Event, 159–186.
- [64] Yifan Wang, Minzhao Lyu, and Vijay Sivaraman. 2024. Characterizing User Platforms for Video Streaming in Broadband Networks. In *Proc. ACM IMC*. Madrid, Spain.
- [65] Yifan Wang, Minzhao Lyu, and Vijay Sivaraman. 2024. Standardizing Multimedia QoE Telemetry from Telecommunications Networks for Open Analytics. In *Proc. ACM SIGCOMM Workshop on Emerging Multimedia Systems*. Sydney, Australia, 14–20.
- [66] Jiangkai Wu, Yu Guan, Qi Mao, Yong Cui, Zongming Guo, and Xinggong Zhang. 2023. ZGaming: Zero-Latency 3D Cloud Gaming by Image Prediction. In *Proc. ACM SIGCOMM*. New York, NY, USA, 710–723.
- [67] Yahoo!Finance. 2024. Cloud Gaming Market to Soar to USD 143.4 Bn by 2032. <https://finance.yahoo.com/news/cloud-gaming-market-soar-usd-121200504.html> Accessed: 2025-04-25.
- [68] Xu Zhang, Yiyang Ou, Siddhartha Sen, and Junchen Jiang. 2021. SENSEI: Aligning Video Streaming Quality with Dynamic User Sensitivity. In *Proc. USENIX NSDI*. Virtual Event, 303–320.

## A Ethics

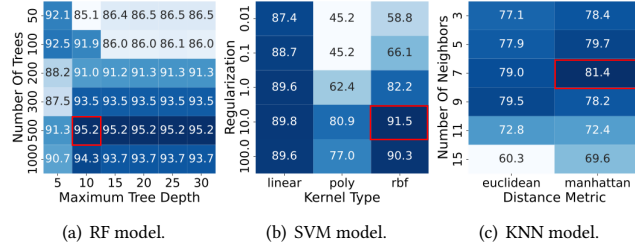
We have obtained ethical clearance from our university ethics board (UNSW Human Research Ethics Advisory Panel approval reference number iRECS5933) to report the measured cloud gaming session characteristics in an aggregated manner, by analyzing real-time network traffic at our partnered Internet service provider that exclusively hosts NVIDIA’s GeForce NOW cloud gaming servers and offline logs. All user identities are anonymous, and we have made no attempt to collect or reveal any personal information. Due to commercial confidentiality restrictions, we only report percentage values after aggregation instead of exact numbers from our field deployment.

## B Data Availability

Our ground-truth traffic traces (PCAP files) of cloud gaming sessions with their context labels (CSV files) that contain game titles, game genres, user platforms, streaming configurations and game activity stages, which are collected in our lab setup and described in §3.1, are publicly shared with the research community via our university cloud drive. We also share a collection of supplementary preprocessing scripts in a GitHub repository. The public links to both the dataset and the scripts are available on our website [31] under the “Cloud Gaming” category.

## C Additional Evaluation Results

In this section, we present additional evaluation results on model selection and attribute importance for game title (§C.1) and gameplay activity pattern (§C.2) classification.

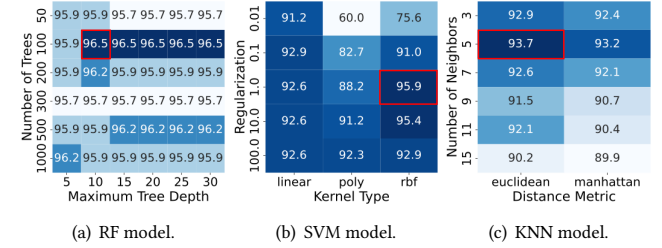


**Figure 14: Three models (i.e., RF, SVM and KNN) are fine-tuned for their hyperparameters for game title classification, with the best accuracy achieved by the RF model.**

### C.1 Game Title Classification

We trained and fine-tuned the hyperparameters of three models that are commonly used in traffic classification tasks, namely Random Forest (RF), Support Vector Machine (SVM) and K-Nearest-Neighbors (KNN). In Fig.14, we show the accuracies of the three models for game title classification while tuning their two representative hyperparameters, including the number of trees and maximum tree depth for Random Forest, regularization parameter (C) and kernel type for SVM, and the number of neighbors and distance metric for KNN. The best hyperparameter combinations from each model are highlighted in red. The overall highest accuracy (i.e., 95.2%) was achieved from the Random Forest model

with the number of trees set to 500 and the maximum tree depth between 10 and 30. In our deployment, we select a maximum tree depth of 10 as it reduces model complexity as well as the risk of overfitting. In comparison, the highest accuracy from SVM and KNN models after fine-tuning are 91.5% and 81.4%, respectively. The substantially lower performance of the KNN model might be due to the limitations of distance-based metrics in capturing the characteristics of high-dimensional attribute space.



**Figure 15: Three models (i.e., RF, SVM and KNN) are fine-tuned for their hyperparameters for the classification of gameplay activity patterns, with the highest accuracy achieved by the RF model.**

### C.2 Gameplay Activity Pattern Classification

Following the same model fine-tuning process as just described for game title classification, we show the accuracy of the three models with different hyperparameters in Fig. 15. Similarly, the highest accuracy of 96.5% was achieved by the Random Forest model with the number of trees set to 100 and the maximum tree depth within the range of 10 and 30. We select a maximum tree depth of 10 in our deployment. Due to the lower dimensionality of the attribute space, the SVM and KNN models demonstrate slightly lower accuracy compared to the Random Forest model, with their highest accuracy being 95.9% and 93.7%, respectively.

**Table 5: Importance of the nine attributes (each representing a transition probability among the three types of player activity stages) in classifying gameplay activity patterns by the best-performing Random Forest model.**

	From	Active	Passive	Idle
To	Active	0.022	0.009	0.018
	Passive	0.052	0.027	0.094
	Idle	0.167	0.022	0.026

Table 5 shows the permutation importance of the nine attributes in classifying the gameplay activity pattern by our best-performing Random Forest model. Each of these attributes represents the probability of one possible stage transition. All nine attributes have certain predictive power in classifying gameplay activity patterns. Among them, the attribute for transitions from active to idle stage has the highest importance.