The Future of Cloud Gaming

Wei Cai,[1] Ryan Shea,[2] Chun-Ying Huang,[3] Kuan-Ta Chen,[4]
Jiangchuan Liu,[2] Victor C. M. Leung,[1] and Cheng-Hsin Hsu[5]

[1] Department of Electrical and Computer Engineering, The University of British Columbia, Canada
[2] School of Computing Science, Simon Fraser University, Canada
[3] Department of Computer Science, National Chiao Tung University, Taiwan
[4] Institute of Information Science, Academia Sinica, Taiwan
[5] Department of Computer Science, National Tsing Hua University, Taiwan

In the late 2000's, several startups including OnLive [7] and GaiKai [8] started to offer *cloud gaming* services. Cloud gaming, as illustrated in Figure 1, refers to the technologies that offload parts of game software from traditional game consoles or Personal Computers (PCs) to powerful and elastic cloud infrastructures. Cloud gaming makes perfect sense to: (i) gamers, who otherwise have to constantly upgrade their consoles or PCs, which is certainly no fun and costly, (ii) cloud service providers, who can sell the already-deployed and idling cloud resources to support the cutting-edge games that are extremely resource-hungry, and (iii) game developers, who no longer need to spend months to port their games to different platforms. As such, cloud gaming has attracted significant attention from both academia and industry.

Initially, given the limited scales of these startups, people were not sure whether cloud gaming could be commercially successful. A fundamental question that came to people, inside and outside of the gaming circle, was: *Will the three dominating console manufacturers, Microsoft, Nintendo, and Sony, be willing to give up their margin of selling the game consoles?* The answer was initially unclear, and many suspected that these console manufacturers would be more than happy to *kill* cloud gaming in order to remain profitable.

The opposite actually happened: Sony acquired GaiKai in 2012 for 380 million USD, announced the new service PlayStation Now (PS Now) in early 2014, and brought PS Now to parts of the world in 2015. Using PS Now, Sony allows their customers to play hundreds of PlayStation 3 (PS3) games on the latest PlayStation 4 (PS4) consoles without porting the games. Sony has also unveiled a plan to: (i) allow gamers to use other PlayStation consoles and smart TV to access PS Now and (ii) include game titles developed for other generations of PlayStations on PS Now. The emerging PS Now service indicates that cloud gaming has great potential to become the next generation, universal computer gaming platform.

Offering high-quality cloud gaming experience with the remote cloud through the public Internet, however, is no easy task. Current cloud gaming services have leveraged real-time video streaming technologies that have a long history of development. They share certain common characteristics including high bandwidth demand and real-time constraints. Yet the demands of gaming are often higher than live or on-demand video streaming. Certain cloud gaming services dictate a network bandwidth of about 5 Mbps for smooth gaming experience, which is unavailable in many residential networks. The literature has also indicated that gamers of first-person shooting

games start noticing lack of responsiveness if the interaction delay is longer to 100 ms. Depending on the game genre, gamers may prefer to have either shorter interaction latency or higher graphics quality. Striving to achieve a balanced tradeoff between latency and quality is therefore quite tricky.

Although some of the requirements have been met by carefully optimizing the components in state-of-the-art cloud gaming platforms, we still need to think outside the box for revolutionary cloud gaming solutions. Let's briefly look at past, present, and future cloud gaming.
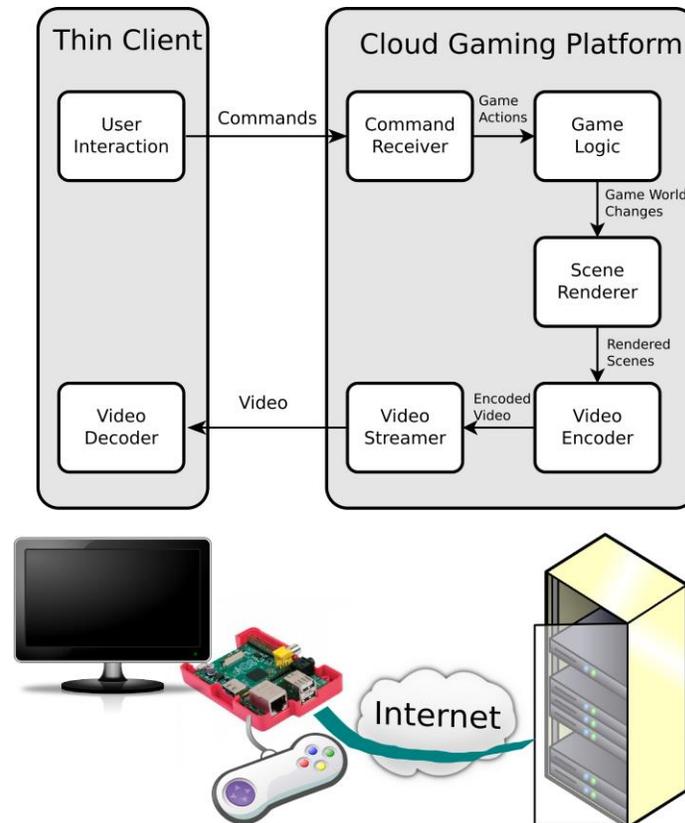

Figure 1 An abstract view of cloud gaming.

I.   Commercial Cloud Gaming Services

The first cloud gaming service, built by G-Cluster[7], appeared in the early 2000's. G-cluster's service was tightly coupled with several third-party companies, including game developers, network operators, and game portals. This may be attributed to the less mature Internet connectivity and data centers, which forced G-cluster to rely on network QoS (Quality of Service) support from network operators. In the late 2000's, emerging cloud computing companies started offering Over-The-Top (OTT) cloud gaming services, represented by OnLive [8] and GaiKai [9]. OTT refers to delivering multimedia content over the Internet to end users above the internet

service providers (ISPs), which trades QoS supports for ubiquitous access to cloud games. OnLive [8] launched subscription based service in 2009 but ceased operations in 2015 after selling their patents to Sony. Another cloud gaming service, Gaikai [9], was founded in 2008 and offered cloud gaming service which allowed gamers to try new games without purchasing and installing them. GaiKai was acquired by Sony in 2012, leading to PS NOW's launch in 2014.

These commercial cloud gaming services are either integrated with underlying networks or provided as OTT (Over-the-Top) services. Tighter integration provides better QoS guarantees, which potentially lead to better user experience, while OTT reduces the expenses on cloud gaming services at a possible risk of unstable user experience. Gamers pay for services by per-game or monthly subscription, or such third-party companies as game developers and network operators may cover the cost, offering free services for end users. At this moment, it is still not clear which integration approach and charging model will work the best in the future. More importantly, these commercial cloud gaming platforms are mostly blackbox-based [1, 2]. That is, unmodified games run inside cloud gaming servers, and the rendered audio and video are captured, compressed, and streamed from these cloud gaming servers. This blackbox approach is widely adopted so far, thanks to its simplicity and short time-to-market. It however leaves limited room for optimization.
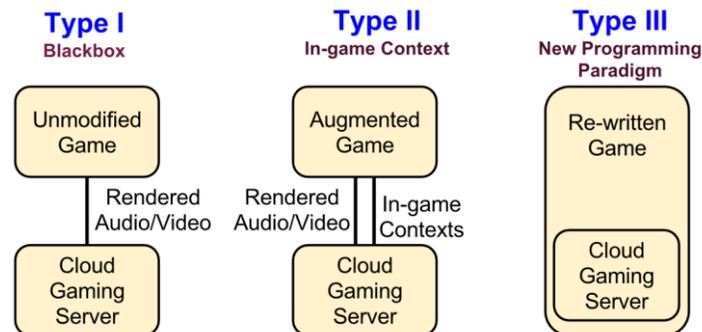

Figure 2 Cloud gaming platforms are classified into three types

II.   From Blackbox to Deep Integration

As shown in Figure 2, advanced cloud gaming should go beyond the blackbox solution (type I), and seamlessly integrate the game software with the cloud gaming servers from the very beginning (Type II and III). In Type II cloud gaming, the gaming software are augmented to expose in-game contexts such as camera location and orientation, for better interaction latency and graphics quality [3, 4]. Using the in-game contexts, cloud gaming servers can perform more efficient video compression through accelerated motion estimation and wrapped video frames. Type III cloud gaming further dictates new programming paradigms [5, 6], where games are written using advanced SDKs (Software Development Kits) that are specialized for cloud gaming.

Although these new programming paradigms offer a larger room for optimization, they impose steeper learning curves for game developers.
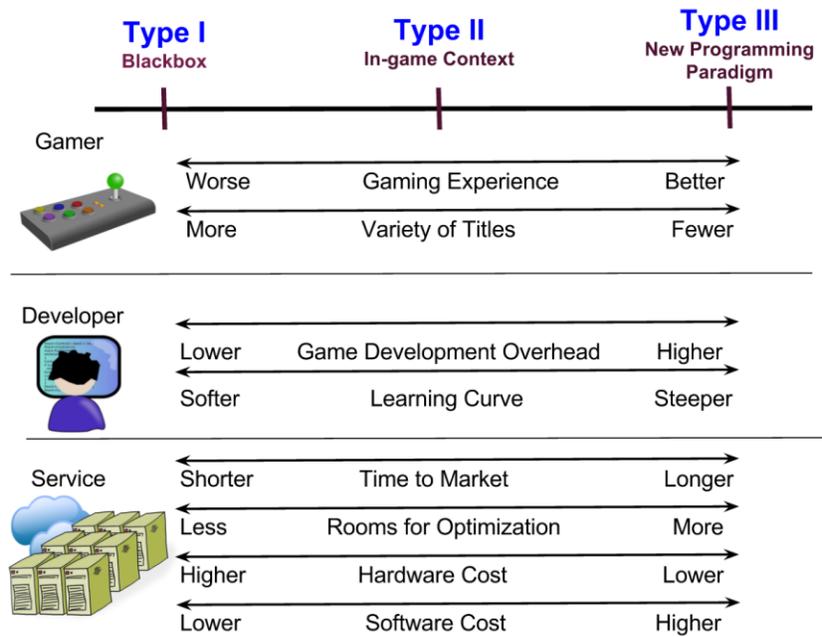

Figure 3 Comparisons among the three types of cloud gaming platforms.

Figure 3 compares these cloud gaming platforms from the perspective of gamers, developers, and of their service requirements. For gamers, the blackbox approach offers more titles but at lower quality. For game developers, new programming paradigm leads to higher development overhead and steeper learning curves. For cloud gaming services, there are two tradeoffs between: (i) time-to-market and room for optimization and (ii) hardware and software costs. In general, the new programming paradigm approach appears to offer more optimization opportunities at the expense of higher implementation complexity. It is therefore interesting to see if the gamers' demand for a high quality gaming experience justifies the additional cost of new programming paradigm due to the implementation complexity. These complicated tradeoffs ensure that the future development of cloud gaming services will remain very intriguing.

III.   The Future of Cloud Gaming

Cloud gaming services have gone through their initial growing pains, and it is now the key moment for cloud gaming to be deployed in living rooms everywhere. Several new cloud gaming services have been launched, including the OTT service offered by SONY PS Now and the integrated service offered by Comcast XFINITY Games. With such breakthroughs, we believe cloud gaming will undergo a series of dramatic upgrades in all aspects, and we present some of our *forecasts* on the future of cloud gaming below.

Forecast #1: Technologies will shift towards a new programming paradigm. We believe commercial cloud gaming services will start to implement Type II cloud gaming services, which leverage in-game contexts to either optimize the gaming experience or reduce the hardware cost. The multimedia research communities have proposed context-aware optimization algorithms for cloud gaming videos [3, 4], which can be readily deployed on commercial cloud gaming platforms. Specifically, approaches can be summarized as following directions: video compression, graphic compression, video-graphic hybrid transmissions, adaptive rendering, and adaptive video encoding. Moving from Type II to Type III may happen later, due to the high implementation complexity. Web games, facilitated by centralized remote server processing and ubiquitous local browser rendering, can be considered as the primary step toward this direction. As a further step, decomposed cloud gaming architecture [5] enables dynamic partitioning for game programs, which learns from system environment and players' behaviors to perform cognitive optimizations. This novel paradigm brings in more technical issues to cloud gaming, such as decomposition methodology, effective system measurement, user behavior identification, and dynamic partitioning algorithms. How soon this will happen totally depends on the evolvement of relevant technologies and whether there are strong demands on high-fidelity cloud gaming.

Forecast #2: Innovative charging models may take time to develop. None of the three current charging models are perfect in market practice. Under the subscription model, OnLive set the monthly fee to about 15 USD, and had a hard time remaining financially sustainable due to insufficient number of subscribers. Sony recently set the PS Now monthly fee to about 20 USD, even higher than that of OnLive. The per-game model requires gamers to purchase the right to play a specific game title for a certain amount of time. Unfortunately, the cost of purchasing (actually renting) a game title for 7 days would already be very close to buying the game title in retail stores. The free model allows gamers to try new games, and hopefully pursues them to buy the actual games, where the revenue comes from game developers who pay for the advertisement of the game titles. We argue that the market of free model may not be large. Allowing gamers to play a new game might introduce negative effects as well: players may notice the limitations of the game, thus, decide not to buy it. According to these, we believe a more comprehensive charging model that involves cloud providers, game developers, network operators, and even gamers is the key to the larger success of cloud gaming services in the future. For instance, a game theory based pricing research should be conducted on the purpose of welfare maximization for all parties involved. Furthermore, due to the heavy rendering workload on cloud server at peak hours, a dynamic instant spot pricing methodology might optimize the server load balancing and also attracts more players with relatively lower price.

Forecast #3: Diverse Internet services are to be integrated. Even before a sophisticated charging model for a public cloud gaming service is developed, we expect to see cloud gaming being utilized as part of other innovative Internet services. For example, a new type of *instant game play* has been provided by several startups, which allows users to try out any mobile app without downloading and installation. The demand for such services becomes more evident when we consider the huge number of mobile apps on the market, and that installing (and probably removing, if users are dissatisfied after trials) each mobile app one is interested in becomes a tedious job, even if each action only involves a few touches. Moreover, the installation of mobile

apps may introduce some security and privacy leaks on mobile devices. A lightweight mobile app trial service without installation overhead would eliminate the aforementioned troubles and potential risks. The value of such services would become more obvious when they are part of Internet ads: When users see the advertisement of a mobile app, they can now immediately and safely give the mobile app a try with a single touch and no extra effort and risk. Providing try-out services for mobile apps to all the users is costly due to the computational and networking overhead, which may impose negative impacts on the profits of service providers. *Targeted* try-out services, in which only the mobile apps that will attract specific users would be shown based on sophisticated recommender systems, are therefore crucial for the financial success to these services. Predicting what are the mobile apps a specific user would eventually purchase is no easy task, but is certainly possible after carefully analyzing the historical preferences of individual users, the relationships between users with similar interests, and the content of individual mobile apps. Analyzing the user preferences and behaviors may bring additional benefits for the try-out services providers, e.g., they may provide large-scale and timely inputs, such as bug reports and user experience, to the mobile app developers, which may in turn enhance their apps in the next release sooner.

Forecast #4: More multiplayer games will be engaged. The gaming industry has seen a shift towards games with multiplayer facilities. Cloud gaming has huge potential in multiplayer games, which involve more than one player in the same game environment at the same time. Applying cloud gaming to multiplayer games introduces additional benefits as follows. i) The nature of connectivity: the indispensability of network connectivity is a critical drawback of the cloud gaming system which may have kept some customers away. However, this worry is unlikely to impact the decisions of gamers when it comes to multiplayer games, since network access is already a mandatory for such games. ii) Temporary engagement: An important feature of cloud gaming is to enable game play without download and installation. This becomes more attractive in a multiplayer scenario where people in proximity are engaged to play the same game within a short period. In this case, the benefit of click-and-play cloud gaming becomes self-evident. iii) Competition fairness: To achieve fairness among multiplayers is a crucial issue in designing online games. Players may suffer from unfairness, especially when the QoS (e.g., latency, packet loss rate) of their network connections varies. In cloud gaming paradigm, players' gaming instances are merely identical in the cloud, and the message exchanges among game instances are internal data transmissions, whose QoS is easier to be guaranteed. The cloud gaming system may be able to adapt itself to a terminal's network to provide better fairness. For example, by reducing the video quality, gamers with less capable devices or poor network conditions can be treated more fairly in multiplayer games. From the technical perspective, existing studies have already demonstrated potentials in optimizing cloud gaming system. Utilizing peer-to-peer (P2P) caching and sharing for gaming data will significantly reduce the download burdens, while adopting cooperative video encoding scheme explores the similarities among peer players' frames, in order to minimize bandwidth consumption for real-time gaming video transmissions.

Forecast #5: Richer interactions will emerge among gamers and observers. The abundant and elastic resources from the cloud infrastructure allow cloud games to host more observers, compared to traditional games. In particular, with cloud gaming, observers do *not* need to

purchase the game software to watch matches. This will attract many more observers (since it's now free and instant) and more gamers (so that their matches can be watched by more people), which lead to new business models, like those adopted by *Twitch*, a popular live game streaming service. E-sport events, including major tournaments are streamed over Twitch, while pro-gamers make a living from sponsorships and advertisements, with fans tuning in everyday. In 2014, Twitch is the 4th largest video content provider across the world; while the number of unique Twitch viewers reaches 100 million viewers per month. The cloud infrastructure also enables multi-modal interactions among gamers and observers. For example, Twitch offers text-based chat rooms; while webcam/microphone streams can also be exchanged. We believe these interactions further encourage more gamers and observers to become cloud gaming users. Yet given the realtime nature of gaming, cross-user synchronization will become necessary. Otherwise the observers with shorter broadcast latency may act as spoilers, while those with longer broadcast latency may post comments on the content already watched by others a while ago, both significantly affecting user experience. Considering the heterogeneous network conditions of individual observers, their end-to-end delays would inevitably differ to a high degree, leading to highly unsynchronized views. As such, smart scheduling in the cloud and adaptive streaming rate control for individual observers are necessary to ensure both timeliness and fairness among them.

Forecast #6: Novel gaming paradigms will converge in cloud. Cloud computing provides additional opportunities for novel gaming paradigms, such as Virtual Reality (VR) games, Augmented-Reality (AR) games, and context-aware games. VR technology creates real-time omnipresent 3D scenarios for immersive players, which relies on high performance hardware. Cloud computing provides a potential opportunity in boosting the popularity of VR games, as it provides rich rendering resources remotely in the cloud, e.g., NVIDIA GRID. Moreover, the high volume of memory and computational power make the infinite virtual world possible for gamers However, due to the extremely high requirement on responsiveness (no more than 50 ms delay to achieve a tolerate gaming experience), it is still an open issue to overcome the network latency overhead for accessing the cloud. On the other hand, AR and context-aware games both accept real-time environmental inputs, perform fast learning and then generate according dynamic virtual contents. These processes involve many high-workload technologies, including video processing, sensor data transmission, pattern recognition, and graphics rendering, which can also be facilitated by cloud resources.

IV.   Conclusion

In this article, we have classified cloud gaming platforms into three types based on how games are integrated with platforms. We have also reviewed the history of cloud gaming services, and noted that it is a key moment for cloud gaming services to increase their penetration rates. Last, built upon our extensive research experience in cloud gaming, we share several of our visions into future cloud gaming technologies, business models, and social impacts, in the format of forecasts. While our forecasts may not be an exhausted list, we firmly believe this article will

stimulate more discussions among cloud gaming researchers and practitioners, resulting in a sustainable cloud gaming eco-system.

V. References

1. C. Huang, K. Chen, D. Chen, H. Hsu, and C. Hsu, "GamingAnywhere: The first open source cloud gaming system," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 10, no. 1s, pp. 10:1–10:25, January 2014.
2. R. Shea, D. Fu, and J. Liu, "Rhizome: utilizing the public cloud to provide 3D gaming infrastructure," in Proc. of ACM SIGMM Conference on Multimedia Systems (MMsys'15), Portland, OR, March 2015, pp.97-100.
3. M. Semsarzadeh, M. Hemmati, A. Javadtalab, A. Yassine, and S. Shirmohammadi, "Video encoding speed-up architecture for cloud gaming," in Proc. of IEEE International Workshop on Cloud Gaming Systems and Networks (C-Game'14), Chengdu, China, July 2014, pp. 1–6.
4. S. Shu, C. Hsu, K. Nahrstedt, and R. Campbell, "Using graphics rendering contexts to enhance the real-time video coding for mobile cloud gaming," in Proc. of ACM International Conference on Multimedia (Multimedia'11), Scottsdale, AZ, November 2011, pp. 103–112.
5. W. Cai, H. Chan, X. Wang, V. Leung, "Cognitive Resource Optimization for the Decomposed Cloud Gaming Platform," in IEEE Transactions on Circuits and Systems for Video Technology, vol.25, no.12, pp. 2038-2051, Dec. 2015
6. Amazon AppStream, https://aws.amazon.com/appstream/.
7. OnLive, http://onlive.com.
8. GaiKai, https://www.gaikai.com.