A Survey to Enable Rich Mobile Multimedia Applications

Ashwini Prabhu
2nd SEM, Dept. of Computer Science & Engineering
SJB Institute of Technology, Bangalore-60
Karnataka, India
ashprabhu27@gmail.com

Abstract— **Mobile/cloud computing is the combination of cloud computing and mobile networks to bring benefits for mobile users, network operators, as well as cloud computing providers.A new Cloud Mobile Gaming (CMG) approach, where the responsibility of executing the gaming engines, including the most compute intensive tasks of graphic rendering, is put on cloud servers instead of the mobile devices, has the potential for enabling mobile users to play the same rich Internet games available to PC users. However, the mobile gaming user experience may be limited by the communication constraint imposed by available wireless networks and computation constraint imposed by the cost and availability of cloud servers. Mobile cloud computing can help bridge this gap, providing mobile applications the capabilities of cloud servers and storage together with the benefits of mobile devices and mobile connectivity.**

We analyze the challenges imposed by mobile cloud computing that need to be addressed to make CMM applications viable, including response time, user experience, cloud computing cost, mobile network bandwidth, and scalability to large number of CMM users, besides other important cloud computing issues like energy consumption, privacy, and security. We illustrate the challenges using Cloud Mobile Gaming (CMG), an approach that enables rich multiplayer Internet games on mobile devices, where compute intensive tasks like graphic rendering are executed on cloud servers in response to gaming commands on a mobile device, and the resulting video has to be streamed back to the mobile device in near real time, making it one of the most challenging CMM applications.

Keywords— Mobile Computing, Cloud Computing, Cloud Gaming, Multimedia Applications.

I. INTRODUCTION

Over the last few years, there has been an increased number of applications that have "migrated to the cloud", and new cloud-based applications that have become popular. Most of the early adopters of cloud have been enterprise applications and IT departments. Similar motivations that have driven mobile enterprise cloud services are also driving adoption of mobile consumer cloud services: the ability to access media from anywhere: any device, platform, and network.

In this paper, we focus on Cloud Mobile Media (CMM) applications and services, which will enable mobile users to not only access rich media from any mobile device and platform, but even more importantly, which will enable mobile users to engage in new, rich media experiences, through the use of mobile cloud computing, that are not possible otherwise from their mobile devices. CMM will also enable service providers and network operators to offer services much more efficiently, with lower cost and better user experience. As more consumers adopt smartphones and tablets as one of their primary media experience platforms, CMM has the potential of significantly boosting the revenue of cloud Software-as-a-Service (SaaS) providers.

Some of the media rich CMM services will require new and richer platform and infrastructure capabilities as explained in the next sections, thereby providing a new set of revenue opportunities for cloud platform and infrastructure providers. And finally, CMM offers new opportunities for mobile network operators to close the growing gap between growth in data usage and data revenue by offering innovative CMM services and experiences, outside of conventional application stores where their participation has not been strong so far.

We illustrate the challenges using Cloud Mobile Gaming (CMG), one of the most compute and mobile bandwidth intensive CMM applications. Subsequently in Section IV, we propose an adaptive mobile cloud computing technique to address the challenges associated with CMG. We show that it is possible to dynamically vary the richness and complexity of graphic rendering in the cloud servers, depending on the mobile network and cloud computing constraints, thereby impacting both the bit rate of the rendered video that needs to be transmitted from the cloud server to the mobile device, and the computation load on the CMG servers [6]. An adaptation process whose offline steps determine in advance complexity and rendering models, which are used by the online step to adapt rendering settings in real time to meet changing communication and computation constraints.

We present experimental results demonstrating the ability of the proposed approach to dynamically address changing network conditions to ensure user experience, as well as ensure scalability by significantly reducing cloud computing costs and network bandwidth needed. We conclude in Section V with suggestions for future research for mobile cloud computing to efficiently enable future Cloud Mobile Media applications.

II. SYSTEM ARCHITECTURE

Utilizing available cloud computing and storage resources, we expect a heterogeneous set of Cloud Mobile Media services and applications to emerge, with different types of consumer experiences and advantages enabled. We first describe the typical end-to-end control and data flow architecture of CMM applications. Next, we categorize the existing and expected CMM applications, and analyze for each category the cloud infrastructure and platform needs, advantages and user experiences enabled, and challenges to make the applications successful.

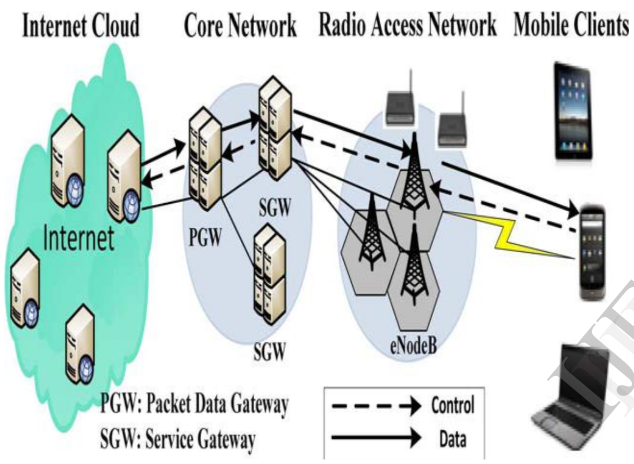

Fig.1.: Cloud Mobile Media architecture, showing control and data flows.

Fig.1 shows the overall architecture, including end-to-end flow of control and data between the mobile devices and the Internet cloud servers, for a typical CMM application. Though a CMM application may utilize the native resources of the mobile device, like GPS and sensors, it primarily relies on cloud computing Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) resources, like elastic computing resources and storage resources, located in Internet public, private, or federated (hybrid) clouds. A typical CMM application has a small footprint client on the mobile device, which provides the appropriate user interfaces (gesture, touchscreen, voice, text based). to enable the user to interact with the application. The resulting control commands are transmitted uplink through cellular Radio Access Networks (RAN) or WiFi Access Points to appropriate gateways located in operator Core Networks (CN), and finally to the Internet cloud.

Subsequently, the multimedia data produced by the cloud, either as a result of processing using the cloud computing resources, and/or retrieval from cloud storage resources, is transmitted downlink through the CN and RAN back to the mobile device. The CMM client then decodes and displays the results on the mobile device display. *Mobile Cloud Storage* is the most commonly used category of CMM application/service today, with offerings from Amazon, Apple, Dropbox, Funambol, and Google, among others. These services provide diverse capabilities, including storing documents, photos, music and video in the cloud, accessing media from any device anywhere irrespective of the source of the media and/or the device/platform used to generate the media, and synchronizing data/media across multiple devices a typical user owns mobile devices grow, and camera resolutions of mobile devices continue to increase.

To enable mass adoption of such services, the PaaS providers will need to ensure high availability and integrity of data, and the SaaS provider will need to ensure content security and user privacy. However, a major impact of the mass adoption of this category of CMM service will be significant increase in mobile data traffic, and potentially larger data bills for mobile subscribers, issues that will need to be addressed for these services to be used ubiquitously, including using cellular networks.

III. RENDERING ADAPTATION APPROACH

To address the challenges of ensuring high user experience, low cloud cost and network bandwidth, and high scalability for cloud based mobile rendering applications, in this section we propose an innovative rendering adaptation approach, which can dynamically vary the richness and complexity of graphic rendering depending on the network and server conditions, thereby impacting both the bit rate of the rendered video that needs to be transmitted from the cloud server to the mobile device, and the computation load on the cloud servers. While the proposed rendering adaptation approach can be widely used for any cloud based rendering applications, we use Cloud Mobile Gaming (CMG) as a running example to introduce and validate our approach in details.

A. Overview of Proposed Rendering Adaptation Approach

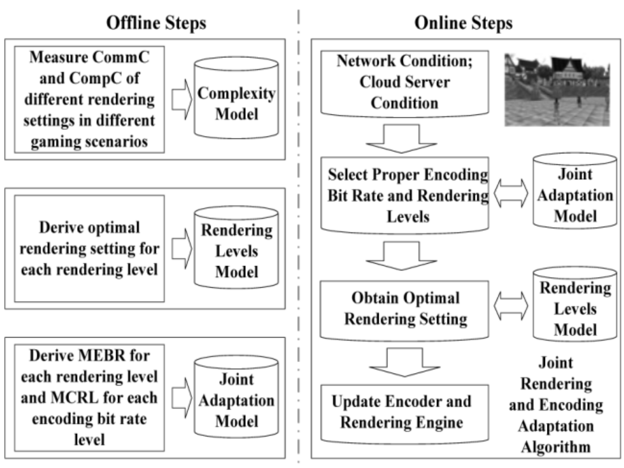

Fig.2: Proposed rendering adaptation methodology.

ISSN: 2278-0181

Graphic rendering is the process of generating an image from a graphic scene file, which usually contains geometry, viewpoint, texture, lighting, and shading information as a description of the virtual scene. It is configurable by a set of rendering parameters. The term "rendering setting" usually denotes a setting which consists of different values of these rendering parameters.

We next introduce the Communication Complexity (CommC) and Computation Complexity (CompC) associated with each rendering setting in the CMG approach. The Communication Complexity (CommC) of a rendering setting denotes the level of how much bit rate is needed to deliver CMG video with this rendering setting. While the video bit rate is determined by the video compression ratio used, it is largely affected by the video content complexity.

B. *Adaptive Rendering Parameters and Settings*

The first step in enabling dynamic game rendering adaptation in the CMG approach is to identify the adaptive rendering parameters and adaptive rendering settings. A game may have many different rendering parameters, but only a few of them have obvious impacts on CommC or CompC. An "adaptive rendering parameter" must be able to adapt at least one of CommC or CompC.

Based on the above principles, we identify four common parameters which we believe are suitable for rendering adaptation in most 3D games: 1) *Realistic effect*: Realistic effect basically includes four parameters: color depth, anti-aliasing, texture filtering, and lighting mode. Each of the four parameters only affects part of graphic rendering.

Varying any one of them may not reduce the graphic rendering load. Thus when we reduce/increase the realistic effect, we vary all four parameters. 2) *Texture detail*: This is also known as Level of Detail (LOD). It refers to how large and how many textures are used to present objects. 3) *View distance*: This parameter determines which objects in the camera view will be included in the resulting frame, and thereby should be sent to the display list for graphic rendering. 4) *Environment detail*: Many objects and effects (grass, flowers, and weather) are applied in modern games, to make the virtual world look more realistic. However they are not really necessary for users playing the game. Therefore, we could eliminate some of these objects or effects to reduce CommC and CompC if needed.

IV. OPTIMAL ADAPTIVE RENDERING SETTINGS AND LEVEL-SELECTION ALGORITHM

We present a 3D rendering adaptation scheme for Cloud Mobile Gaming, which includes: 1) an off-line step of identifying the optimal rendering settings for different adaptation levels, where each adaptation level represents a certain communication and computation cost; 2) a run-time level-selection algorithm, that automatically adapts the adaptation levels, and thereby the rendering settings, such that the rendering cost will satisfy the communication and computation constraints imposed by the fluctuating network bandwidth and server available capacity.

A. *Optimal adaptive rendering settings*

In our rendering adaptation technique, we adapt application cost by adjusting the adaptation level. The higher adaptation level, the higher CommC and CompC will be. Since there are too many possible rendering settings and different rendering parameters have different impacts on CommC and CompC, it is important to identify the optimal rendering setting for each adaptation level in advance.

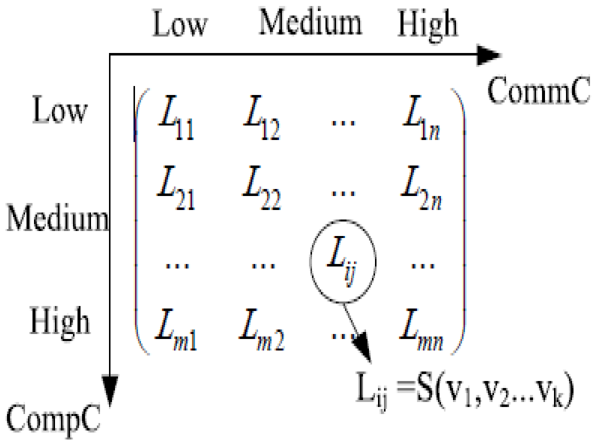

Fig 3: Adaptatation levels and matrix L

As shown in Fig.3 computation cost at level *i* and communication cost at level *j*. In each adaptation level, there is a k-dimensional (k is the number of rendering parameters) rendering setting S. The elements of S indicate the values of the rendering parameters. All the adaptation levels in L. should be able to provide acceptable gaming quality to the user. The more resource cost, the better gaming quality will be.

A. *Level-selection algorithm*

Having identified the optimal rendering settings for different adaptation levels, we next develop a level-selection algorithm, which can decide when and how to switch the rendering settings during a gaming session, in response to the current network conditions and server utilization, so as to satisfy the network and server computation constraints. The level-selection algorithm should first be able to realize when the network is constrained or the server is over utilized. We use both network delay and packet loss as indicators to detect constrained network. In the best network conditions (not overloaded), packet loss rate is 0, but there is a certain minimum round-trip delay, denoted by *MinDelay*, due to the time needed to transmit a packet through the core network and the RF link.

We empirically use 1.2 times of *MinDelay* as the threshold for round-trip delay together with the packet loss to estimate the constrained network. To detect server over-utilization, we empirically use 90% as the threshold of GPU utilization. If GPU utilization is above this threshold we will identify the GPU is getting over utilized. When either a constrained network or an over utilized server is detected, the level selection algorithm will select a lower adaptation level (with lower cost). When the network condition and/or server utilization improves, the algorithm will select a higher level (with higher cost, and thereby higher user experience).

However, to avoid oscillations between adaptation levels, we also have set conditions for increasing the adaptation level. For communication cost level, we increase it if the network has stayed in the good condition (no packet loss, delay less than threshold) for 2 minutes. For computation cost level, we increase it when the server utilization is below 80%.

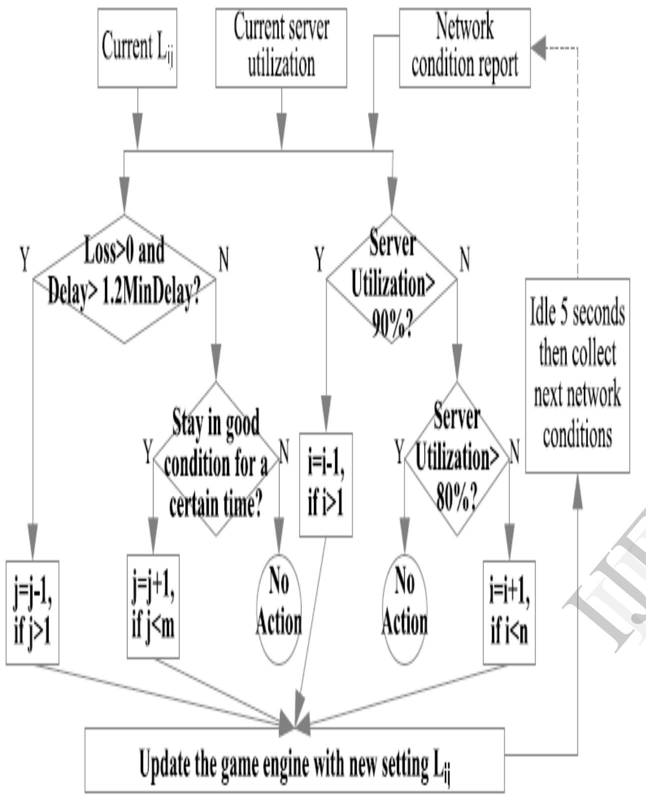

Fig 4: Work flow of level-selection algorithm

Fig.4 shows the proposed level-selection algorithm. Depending on the network conditions (delay and packet loss),and server utilization, the level-selection algorithm decides to select a lower or higher adaptation level, or keep the level the same as the current level Lij. Its mechanism is as follows:

If there is packet loss and the network round-trip delay is greater than the threshold (1.2**MinDelay*), we lower the CommC level j by 1. On the other hand, if the network has been in good condition (no loss and delay less than the threshold) for sufficient time (we use 2 minuted empercially in this paper), , we increase j to the next higher level, unless j is already at level n. For other situations, the algorithm keeps the same level.

If server utilization is over 90%, we decrease the computation cost level (index i) by 1. Otherwise if server utilization is not over 80%, we increase i by 1 unless it is already at level m. For the new adaptation level selected by the rate-selection algorithm above, the optimal rendering settings are selected from the adaptation matrix, and updated into the game engine.

V. CONCLUSION

We can conclude in this paper with two additional approaches which influences the user experience, cost and scalability challenges associated with CMM applications. We presented a rendering adaptation technique to address the constraints of wireless network communication bandwidth and server computation capacity in a Cloud Mobile Gaming approach.A critical challenge for CMM applications is the latency and jitter associated with the uplink and downlink transmissions between the mobile device and the Internet cloud servers. Transmission of large amount of content between cloud servers and mobile devices, inherent in CMM applications, poses a major concern for the capacity of the networks to enable CMM applications.

Cloud computing to the edge of the mobile network, supplementing gateway nodes in the mobile Core Network (CN), and edge nodes like base stations in Radio Access Networks (RAN), and Femto and WiFi access points, with computing and storage resources, to form a true *Mobile Cloud*. With a Mobile Cloud architecture, content processing (like graphic rendering or video encoding) and retrieval can be performed at the edge of the mobile networks, as opposed to in Internet clouds, thereby reducing round trip network latency, as well as reducing congestion in the mobile CN and RAN.

REFERENCES

[1] S.Wang and S. Dey, "Cloud mobile gaming: Modeling and measuring user experience in mobile wireless networks," in *Proc. ACM SIGMOBILE MC2R*, 2012, vol. 16, no. 1, pp. 10–21.
[2] S. Wang and S. Dey, "Modeling and characterizing user experience in a cloud server based mobile gaming approach," in *Proc. IEEE GLOBECOM*, Honolulu, HI, USA, Dec. 2009.
[3] S.Wang and S. Dey, "Rendering adaptation to address communication and computation constraints in cloud mobile gaming," in *Proc. IEEE GLOBECOM*, Miami, FL, USA, Dec. 2010.
[4] S. Wang and S. Dey, "Addressing response time and video quality in remote server based internet mobile gaming," in *Proc. IEEE WCNC*, Sydney, Australia, Mar. 2010.
[5] S.Wang, Y. Liu, and S. Dey, "Wireless network aware cloud scheduler for scalable cloud mobile gaming," in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012.
[6] M. van der Schaar and D. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Transactions on Multimedia*, 9(1):185–197,Jan.2007.

.

[7]. S. Mukhopadhyay, C. Schurgers, S. Dey, "Joint Computation and Communication Scheduling to Enable Rich Mobile Applications, in *Proc.of IEEE GLOBECOM,* Washington D.C., Nov. 2007.
[8]. A. Shieh, et al, "Sharing the data center network", in *Proceedings of the 8th USENIX conference on Networked systems design and implementation*,2011.
[9]. R. Nathuji, et al, "Q-clouds: managing performance interference effects for QoS-aware clouds", *in Proc. 5th European conference on computer systems,* pp. 237–250. ACM, New York, 2010.
[10]. Luis M. Vaquero et al., "A break in the clouds: towards a cloud definition,"ACM SIGCOMM Computer Communication Review, Jan. 2009.