# Cloud Gaming:

# Architecture and Performance

## 一、 動機

隨著雲端技術的進步與資料中心的布建，cloud computing 不再僅限於簡單的運算工作。得利於 cloud 的計算卸載(computational offload)技術，能顯著提升系統效率，並藉由策略性部署 cloud data center 以減少使用者於互動上的延遲，使 cloud gaming 得以實現。

玩家可透過網路將指令傳送至遊戲平台上的主程式，經平台運算後再將處理後的影像與回應傳回至玩家螢幕中。Cloud gaming 的架構能讓性能較弱的設備，如：智慧型手機、平板電腦等，透過 thin client 方式與應用程式互動，顯示運算部分只需要交由 cloud rendering sever 即可，據此來運行高規格需求的遊戲，同時也可降低玩家須自行購買硬體的設備成本。

在遊戲開發商方面，由於遊戲的程式碼並不直接在玩家的 local machine 上執行，且 cloud computing 的硬體由 cloud gaming 提供商所控制，因此能擁有更好的數位版權管理，降低管理成本與盜版所帶來的損失，同時也能減少 customer support cost。

## 二、 議題

Cloud gaming system 必須收集使用者的動作並傳送到 cloud server 進行處理，再根據結果爲遊戲世界因動作而造成的改變進行編碼、壓縮後，以影片串流送回給玩家。這一系列的步驟由於必須確保使用者的互動性，因此每個環節都必須在毫秒內完成。而由於使用者對於互動上的延遲容忍度低，系統於進行上述如影像壓縮等步驟時的時間需要越短越好。此外，若網路上的延遲越高，對於玩家互動體驗產生的負面影響也越高，以下將就兩點進行討論：

A. Interaction delay tolerance:

對於 cloud gaming 而言，無論是多人連線遊戲或是單人遊戲，由於所有遊戲的執行都是在遠端上進行，再將執行結果以串流傳回使用者的 thin client。然而，網路傳輸資料多少皆有延遲的問題，因此需將這些延遲控管在不影響玩家遊戲體驗範圍內，互動的延遲性也就成爲首要的關注重點。

下方 Table 1 列出目前傳統常見遊戲類型與其所能接受的延遲，而 Table 2 爲作者利用 Onlive 平台進行測試的結果，在 Onlive base processing time 僅有 36.7(ms)的情況下，cloud 產生的基本 overhead 就達到 100 ms，意味著在 Onlive 上進行需要低延遲的即時射擊遊戲時，玩家

遊玩體驗必定受到影響。而在 RPG 類與 RTS 遊戲上則相對是叫有機會於 Onlive 上順暢運行的，因此解決延遲問題的關鍵在於 cloud 處理能力，必須降低 cloud overhead 才得以將更多低延遲容忍度的遊戲整合至雲端上執行。

| Example game type | Perspective | Delay threshold |
|---|---|---|
| First person shooter (FPS)<br>Role playing game (RPG)<br>Real-time strategy (RTS) | First person<br>Third person<br>Omnipresent | 100 ms<br>500 ms<br>1000 ms |

Table 1. *Delay tolerance in traditional gaming.*

| Measurement | Processing time (ms) | Cloud overhead (ms) |
|---|---|---|
| Local render | 36.7 | N/A |
| Onlive base | 136.7 | 100.0 |
| Onlive (+10 ms) | 143.3 | 106.7 |
| Onlive (+20 ms) | 160.0 | 123.3 |
| Onlive (+50 ms) | 160.0 | 123.3 |
| Onlive (+75 ms) | 151.7 | 115.0 |

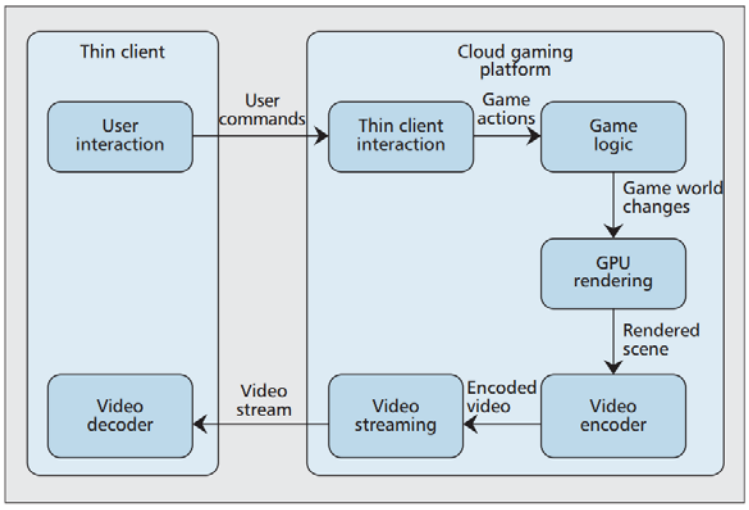
Table 2. *Processing time and cloud overhead.*

B. Video streaming and encoding:

Cloud gaming 中的影音串流與傳統串流機制不同，Cloud gaming 架構必須快速的對影片進行編碼(如:H.264)、渲染以及壓縮並將該影片回傳給使用者，且由於在雲端中的動作會在短時間內(100 ~ 200 ms)完成，因此在 client 端的緩衝並沒有太大的發揮空間而無法預先暫存資料。因而 cloud gaming 中的 sensitive real-time 編碼需求對於 cloud gaming 的提供商至關重要。

# 三、 研發方向

在本文中作者提出一套通用的 cloud gaming 架構，並解釋 cloud gaming 所需的功能和模組，其架構如下圖：



i. User Interaction: 使用者對於應用程式所要求執行的命令，例如：在遊戲世界中飛行、施行法術，藉由網際網路傳送至 cloud gaming 平台。

ii. Thin client interaction: 在 cloud gaming 平台上，負責接收由 thin client 傳送來的使用者命令並轉換成遊戲中的動作(Game action)。

iii. Game logic: 根據遊戲中的動作以遊戲邏輯解釋後得到遊戲世界中應產生的變化。

iv. GPU rendering: 由 cloud system 中的圖像處理單元(GPU)將應產生的變化處理成爲要回傳的 scene。

v. Video encoder: 對 scene 進行編碼壓縮後傳送給影片串流模組。

vi. Video streaming: 將 scene 以影片串流傳送回給 thin client。

vii. Video decoder: thin client 將影片串流解碼後，以播放器顯示。

viii. 在短時間內完成並重複(i) ~ (vii)，達成使用者一連串的遊戲互動。

以上的架構將所有需要計算的部分全部皆交由 cloud 處理(computational offloading)，讓使用者端的 thin client 部分僅需要專注於呈現遊戲影片。

# 四、 未來預測

本文深入檢視了 cloud gaming 的設計架構，同時也測量了具代表性的 cloud gaming platform 性能。根據在不同遊戲、電腦設備以及網路配置下的互動延遲以及串流品質的結果，顯示 cloud gaming 具有未來發展的潛力，同時也揭示了廣泛部署上的挑戰。

隨著 AR 與 VR 遊戲開發越趨蓬勃，所需最基本硬體規格也不斷向上調整，但倘若相關遊戲欲放上雲端則必須要有極短的延遲，才能確保玩家的遊戲體驗。此外智慧型手機的普及，使手遊市場也愈趨壯大，但受限於智慧型手機硬體能力，手由規模始終無法如傳統電腦遊戲一般，若是藉助 cloud gaming 或將可突破如此障礙。除了專注於遊戲互動延遲的議題上，thin client 的進步也可能帶來一些改變。當 thin client 可以執行簡單的遊戲邏輯和遊戲場景變化，能夠藉此隱藏互動上的延遲，或者將遊戲的執行分散到多個各司其職的 virtual machine 上(distributed game execution)，例如：當遊戲角色在施行法術的動作時，遊戲角色將進行一連串的動作，而這些動作可以由 thin client 完成，這過程應足夠掩蓋過互動上的延遲。

在硬體方面，cloud gaming 的發展也讓硬體商開始設計解決 cloud gaming 問題的硬體解決方案，例如：NVIDIA 推出同時 GeForce grid graphical processor，該處理器爲具有 encoding solution 的圖像處理器，有能力同時 render 和 encode 四個遊戲，並宣稱能夠顯著減輕目前 cloud gaming 的延遲問題。配合軟體的優化再加上硬體的輔助，目前 NVIDIA 甚至能同時 render 四位玩家的畫面，且爲可接受的延遲。此外，在網路方面目前不易找到能將延遲控制在 200 毫秒以內的解決方案，但或許 LTE 或 5G 技術的發展能帶來相關突破。