CONSIDERATIONS FOR A REAL TIME BROADBAND DELIVERED INTERACTIVE GAMING SERVICE

Charles H. Jablonski
VP Operations, OnLive Inc

Abstract

Real Time graphics intensive applications, such has "first person shooter games" have presented challenges in being server/cloud based because of the inherent roundtrip latencies.

This paper will describe an approach that provides the ability not only to deliver real time games but also that is able to deliver any graphics intensive application over existing broadband infrastructure.

The approaches taken to provide the necessary low latency compression, server and data center architecture, integrating into exiting carrier and ISP data infrastructures as well as the impacts of "last mile" ISP (DSL and Cable) considerations as well as the required performance of the date delivery, packet loss and jitter specifications of the infrastructure.

An approach using existing PC/Mac client for delivery of the service to end consumers will be also be described as well as a small stand alone low cost (<$30) terminal adapter that connects a consumer television directly to a internet connection to enable game play.

The system allows a variety game content, media and enterprise application to also be provisioned and delivered over the same infrastructure.

.

<u>Introduction</u>

The march of enabling technology and the continued improvement in broadband connectivity and availability have enabled many services and products to be delivered to end consumers and enterprises as various compressed data and media files. The success of services such as video on demand and "ultra" high speed broadband are well known and their effects have caused consumer tastes and expectation to evolve accordingly.

In few short years linear packaged media has evolved from a shrink wrap retail point of sale product to a anywhere, anytime, almost any device availability. At times the business issues have been far more daunting that the technological, but one undisputed fact is that the consumer continues to be more demanding and the expectations of availability, selection, schedule and independence increases.

Further the same technologies that have enabled these distribution advances have lead to a far greater availability of consumer unilateral and multilateral consumptions platforms (PCs, Net books, Tablets, Portable Devices) and the ability to imbed the functionality into broader consumer platforms, and inexpensive stand alone devices.

One of the last frontiers remaining is true real time interactive consumer experiences in the home, using inexpensive and available end devices, that are provisions by centralized "cloud based" systems. In these applications, such as gaming, the consumer experience cannot be fulfilled unless the challenges of latency, quality and availability using cloud based systems and broadband networks are met.

OnLive is one such system, and it is believed the first among many that will be taking advantage of the broadband infrastructure to widely deliver these interactive real time applications.

<u>Business Issues and Consumer Expectations</u>

As with any good technological innovation, there must be an underlying business reason to support the development and implementation and the implementation must be compelling enough for all parties to embrace it. Science experiments, although intellectually fulfilling, must have the support and resources of a real business problem to evolve into broad and accepted use.

The gaming industry, although a large mature business (>$50B annually worldwide), the distribution of the end product to consumers has been dominated by retail point of sale. Even though the economics of game development resemble other media (high development costs, intensive consumer promotion and advertising, short selling season and short initial release life) the actual end distribution is the last unfulfilled digital delivery for modern media.

Typically many scores of millions are spend developing a game, a few score spent in advertising its release, and then the selling season is typically a few weeks before the holidays. This is all delivered through shrink wrap product that is sold through retail distribution channels. These channels suffer from various impeded costs (duplication, distribution, platform fees, retailer margins) as well as theft, returns and “spoilage”. Also since the end product is in the consumers hands piracy is a constant battle, and >30% of the games are resold through the used market where the initial developer and published receiving none of that incremental revenue. Some high end titles are resold and used 4-5 times (as evidenced by registrations to the publishers).

This “quantized” method of sale also makes incremental releases, episodic, upgrades extremely problematical in the games market. Although there is a strong online community and infrastructure its broad use is limited to small add-ons, consumer social features test and chat. What few online sales mechanisms in existence they either suffer from extremely long and large downloads, or they are very limited in their catalog of current first release titles.

Typically software publishers receive less than 50% of the MSRP at the retailer point of sale. Clearly there is room for improvement in the economics of the existing system. Now, when one examines the exiting consumer experience for high end interactive games, that picture has room for improvement as well.

Typically a consumer has to spend $500 (in early cycle) on a proprietary closed gaming console, that by definition and practice will require replacement in 3-5 years, purchase games at retail, sometimes having to wait in line or just wait for availability of a hot game, spend additional monies to participate in a social network for that console. The other choice is to purchase an extremely high end PC (>$5000) and suffer the same issues acquiring the games and software.

Naturally with retail purchase model the end consumer may not have immediate access to new “hot” games, and be in the position on not knowing the game experience until it has been purchased brought home, installed and used.

As with the publisher side of the equation, this systems is ripe for a disruptive change and consumers continue to expect to enjoy real time availability and experience of all other digital media, an online, cloud based gaming system that addresses these business and consumer needs can be a commercial success.

The Consumer Experience

In order for a server based, cloud served system to be successful; it must meet or exceed the existing experience expectations for current consoles and PCs. What does that incorporate?

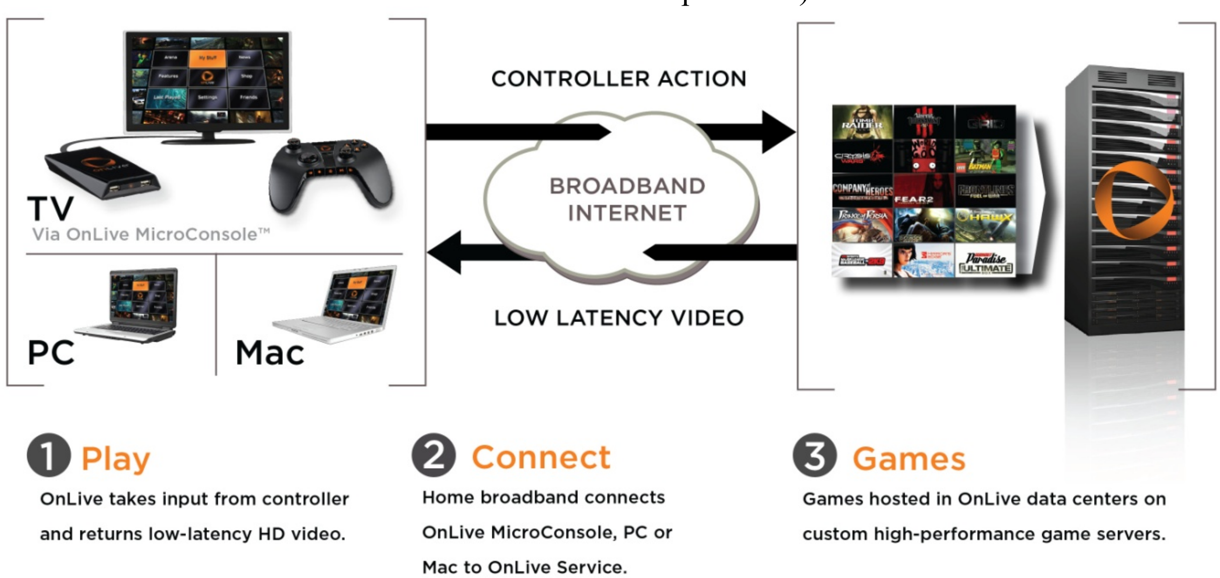
One the image quality must be the same. That requires the ability to render, in real time, high definition, high frame rate images. The response time for the system, using readily available controllers, must be fast enough not to limit the game play because of lag or latency, a large amount of amount devices must be available and supported, be deployable on existing PCs or Macs, it must be extendable to a large screen display device (i.e. TV), be enabled and installed simply over exiting broadband infrastructure, afford social infrastructure and features and have wide availability of new release games from various publishers.

Up until now, the key limitation in existing technologies has been the latency of real time video compression systems. These existing technologies (JPEG, MPEG et.al.) have either required bandwidth only available on large corporate LANs, or required many milliseconds to compress real time images. (Typically game play requires <100ms of response time to be effective in competitive game play)

The OnLive Platform

The following is a brief description of the OnLive platform:

When the user performs an action on a computer or TV connected to OnLive (e.g. presses a button on controller or moves a mouse) that action is sent up through the internet to an OnLive data center and routed to a server that is running the game the user is playing (or the application the user is using—since the interactive demands for video games are generally higher, remote video game operations will be primarily described in the following paragraphs, but these discussions are entirely applicable to remote application operations).



OnLive Platform

The game computes the next video frame based on that action, then a proprietary chip compresses the video from the server very quickly, and the user's PC, Mac or OnLive MicroConsole™ decompresses the video and displays the new frame of video on the user's computer display or TV set. The entire round trip, from the point the button is pressed to the point the display or TV is updated is so fast that, perceptually, it appears that the screen is updated instantly and that the game is actually running locally.

When the user performs an action on a computer or TV connected to OnLive (e.g. presses a button on controller or moves a mouse) that action is sent up through the internet to an OnLive data center and routed to a server that is running the game the user is playing (or the application the user is using—since the interactive demands for video games are generally higher, remote video game operations will be primarily described in the following paragraphs, but these discussions are entirely applicable to remote application operations). The game computes the next video frame based on that action, then a proprietary chip compresses the video from the server very quickly, and the user's PC, Mac or OnLive MicroConsole™ decompresses the video and displays the new frame of video on the user's computer display or TV set. The entire round trip, from the point the button is pressed to the point the display or TV is updated is so fast that, perceptually, it appears that the screen is updated instantly and that the game is actually running locally.

The key challenge in any cloud system is to minimize and mitigate the issue of perceived latency to the end user.

Latency Perception

Every interactive computer system that is used, whether it is a game console, a PC, a Mac, a cell phone, or a cable TV set-top box, introduces a certain amount of latency (i.e. lag) from the point you perform an action and you see the result of that action on the screen. Sometimes the lag is very noticeable (e.g. on some TV set-top boxes it takes over a second to move a selection box in a program guide). Sometimes it isn't noticeable (e.g. if you have a well-designed game running on fast hardware, and pressing the fire button results in what appears to an instantaneous display on your screen of the your gun firing).

But, it's important to note that, even when your brain perceives game response to be "instantaneous", there is always a certain amount of latency from the point you perform an action and your display shows the result of that action. There are several reasons for this. To start with, when you press a button, it takes a certain amount of time for that button press to be transmitted to the computer or game console (it may be less than a millisecond (ms) with a wired controller or as much as 10-20 ms when some wireless controllers are used, or if several are in use at once). Next, the game needs time to process the button press. Games typically run between 30 and 60 frames per second (fps), so that means they only generate a new frame every 1/30th to 1/60th of a second (33ms to 17ms). (Further, when games are generating complex scenes, sometimes they take longer.) So, even if the game responds right away to a button action, it may not generate a frame for 17-33ms or more that reflects the result of the action. And, then finally, there is a certain amount of time from the point the game completes generating the frame until the frame appears on your display. Depending on the game, the graphics hardware, and the particular monitor you are using, there may be almost no delay, to several frame times of delay. And, if your game is an online game, there typically will be some delay to send a message reflecting your action through the internet to other game players, and the game may (or may not) delay the action occurring in your game so as to match your screen action to that of screen action of players who are playing the game remotely. So, in summary, even when you are running a game on a local machine there is always latency.

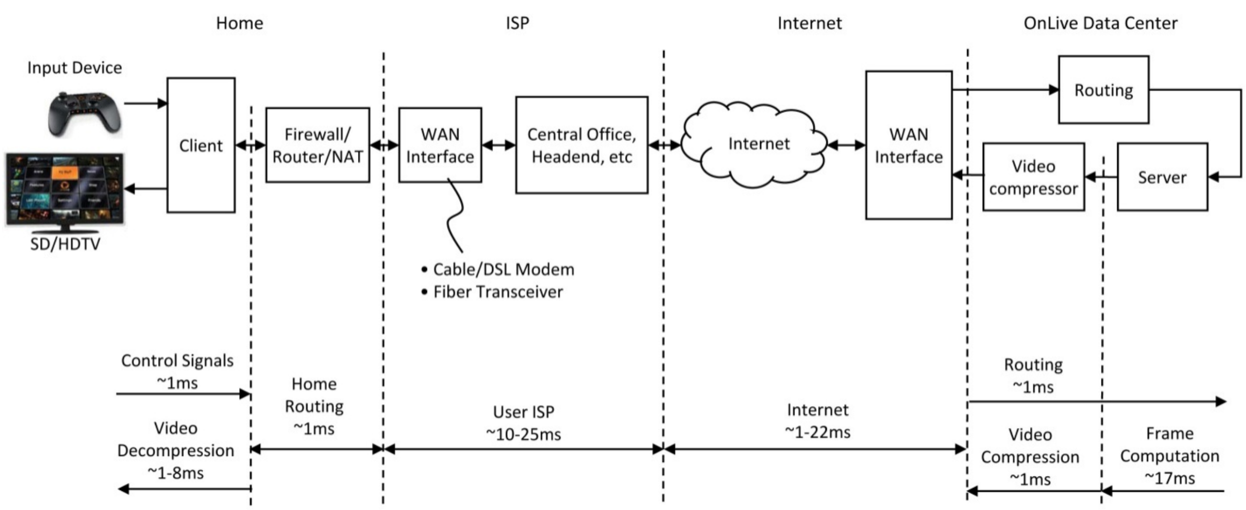
The question is simply how much latency. So, while there certainly are more subtleties to the perception of latency, as a general rule of thumb, if a player sees a fast-action game respond within 80ms of an action, not only will the player perceive the game as responding instantaneously, but the player's performance will just as good as if the latency was shorter.

And, as a result, 80ms is the "latency budget" needed to meet for the OnLive system to be practical. That is to say, OnLive has up to 80ms to: send a controller action from the player's home through the internet to an OnLive data center, route the message to the OnLive server running the game, have the game calculate the next frame and output the video, compress the video, route the compressed video out of the data center, send the compressed video to the player's home through the internet, decompress the video on the players computer or and output the video to the player's display. And, of course, OnLive has to do this at rate of 60fps with HDTV resolution video over a consumer internet connection, running through consumer internet gear in the home.

Over Cable and DSL connections, OnLive is able to achieve this if the user's home is within about 1000 miles of the OnLive data center. So, through OnLive, a user who is 1000 miles away from a data center can play a video game running on a server in the data center with the perception (and the game play score) as if the game is running locally.

<ins>OnLive's Latency Calculations</ins>

The simplified diagram below shows the latencies encountered after a user's action in the home makes it way to an OnLive data center, which then generates a new frame of the video game and sends it back to the user's home for display. Single-headed arrows show latencies measured in a single direction. Double-headed arrows show latencies measured roundtrip.



OnLive System Latency

There latency numbers shown here are numbers that OnLive has seen in practice, given the way the OnLive system was architected and optimized, and reflect what has been measured after using OnLive in various locations over the years. If you add up all of the worst-case numbers, it shows the latency can be as high as 80ms. That said, it is highly unlikely that every segment will be worst case so the total latency will likely be much less (and indeed, that is what we see in practice).

ISP latency

Potentially, the largest source of latency is the "last mile" latency through the user's Internet Service Provider (ISP). This latency can be mitigated (or exacerbated) by the design and implementation of an ISP's network. Typical wired consumer networks in the US incur 10-25ms of latency in the last mile, based on OnLive's measurements. Wireless cellular networks typically incur much higher last mile latency, potentially over 150-200ms, although certain planned 4G network technologies are expected to decrease latency.

Within the internet, assuming a relatively direct route can be obtained, latency is largely proportional to distance, and the roughly 22ms worst case round-trip latency is based on about 1000 miles of distance (taking into account the speed of light through fiber, plus the typical delays OnLive has seen due to switching and routing through the internet.

Consequently, OnLive will be locating its data centers such that the distance to most of the US population is less than 1000 miles
The compressed video, along with other data required by the OnLive client to keep it tightly sync'd with the OnLive service, is then sent through the internet back to the user's home. Notably, the data generated by the video compressor is carefully managed to not exceed the data rate of the user's internet connection because if it did, that might result in queuing of packets (incurring latency) or dropped packets. Since the user's home data rate is constantly changing, the OnLive service is constantly monitoring the available data rate, and constantly adapting the video compression (and if necessary, dropping the video resolution) to stay below the available data rate.

One common misconception about home broadband connections are that the latency is directly tied to data rate (i.e. the effective connection speed) and/or data throughput (i.e. the data rate available to a particular user). Latency is actually largely independent of data rate, so long as the data throughput demands are less than the capacity of the broadband connection.

OnLive video decompression latency

Once the compressed video data and other data is received by the OnLive client (i.e. the OnLive application running as a plug-in or standalone in your PC or Mac, or the OnLive MicroConsole attached to your TV), then it is decompressed. The time needed for decompression depends on the performance of your PC or Mac (CPU and frame buffer bandwidth…no GPU is needed), and may vary from about 1 to 8ms. If your computer's CPU and/or memory bus is tied up doing another processing-intensive task or if you have an extremely low performance computer, OnLive may find it is unable to decompress video at full screen resolution. If so, then it will scale down the video window accordingly. But, we have found most computers made in the last few years work fine up to their screen resolutions so long as they are not tied down running some other intense application at the same time. In any case, even if you are in a processing-constrained situation, OnLive will select a video frame size which will maintain low latency.

OnLive round-trip latency

As mentioned before, while there is a certain amount of latency variability in each leg of the journey, it is rare that a given user will end up in a worst-case scenario with each leg. Consequently, what we typically see in practice are latencies on the order of 40 to 60ms. Sometimes we see latencies that are higher and sometimes we see latencies that are shorter. And, we expect latencies to continue to decline as “last mile” infrastructure is upgraded, both for wired and wireless networks.

Video Quality

As previously discussed video quality is paramount to successful consumer experience. Not only does the image quality have to be acceptable, the rendered frame rate must also be high to provide the gaming experience

During high action game segments, stereo audio

Resolution (p60)	Aspect Ratio	Peak Per Stream[1] Data Rate (Mbps)
480 x 270 (cell)	16:9	0.6
640 x 360 (SDTV[2])	16:9	1.2
1280 x 720 (HDTV)	16:9	4.0
1920 x 1080[3] (HDTV)	16:9	8.0

OnLive Data Rates

The 1280 x 720 resolution currently provided, provides adequate visual experience, and with a peak data rate at 5Mps allows the service to be widely available to consumers through existing broadband infrastructure.

Consumer Infrastructure

As described earlier the OnLive system has been engineered to be deployable on a wide variety of PCs and Macs using a downloadable software client. As for the large screen device, OnLive has developed its own consumer TV adapter, the MicroConsole™.

MicroConsole TV Adapter

The OnLive MicroConsole™, which will be generally available later this year, has been developed to allow simple integration into a home consumer display. The device has an standard NIC connector for access to the broadband service, an HDMI adapter to connect directly to the large screen device to provide high image quality as well as multichannel audio. (A high quality audio experience is another "must have" for a successful consumer gaming experience.)

There are USB connectors to allow integration of standard controllers and devices, wireless radios for headsets, and an optical audio connector. The cost of this device is minimal to allow wide deployment and availability to the consumer.

OnLive Game Service Portal

Consumer and Social Features

As with any current consumer offering, ease of use and simplicity of navigation is paramount. Social features must also be available to support competitive game play.

In addition to providing access to games, OnLive includes a multitude of social features such as social groups (friends) text and voice chat. One of the unique offerings in the OnLive service is the feature of brag clips and spectating. As all the games sessions are running in OnLive datacenters, the games sessions are available and can be seen by other OnLive users in real time. Using multicast distribution any of these streams can be distributed to any OnLive user so your friends can watch, comment and cheer your game play.

Further this system has a continuous recording buffer that allows the immediate capture of the preceding game play as a "brag clip" that can captured stores and downloaded for use in other social media and social networks.

As this system is completely cloud based, all user information is safely stored, include state of game play, so the user can pause, log out and resume from the existing place in the game and maintain all their statistics, rankings and data from their previous session. This information can be accessed on the OnLive system anywhere. This allows a user to log off and resume their play across town or across the country. This allows game playing and spectating by individuals throughout the system.

Product (Games) Availability

The OnLive system and its servers utilize standard PC games with minimal modifications. A SDK is freely available for any game publisher or author to provision their games for the OL platform. As virtually all games are developed on PC development systems virtually all the currently available and developed games can be easily provisioned and provided for the platform.

OnLive has secured long term arrangements with virtually every game publisher that allows their products to be provided to OnLive at launch in June of 2010. These agreements provide for the same day and date release of titles that will be available at retail stores.

Because of the nature of the OnLive systems publishers can now produce and distribute not only primary releases, but provide episodes, mini release and features for their games after the initial release. The OnLive system supports purchase, rental, subscription and demo distribution of product. Try before you buy now becomes a reality for the consumer.

OnLive Status

OnLive has been in development for several years. It was publicly announced in March 2009 at the Games Developer Conference, and beta began shortly after that. The Beta trials have been extensive and provided valuable insight and detailed information on the OnLive Technology, service and network performance. All the Beta participants have been helpful and the ISP and carriers have been extremely supportive in the efforts to bring this service to market. This information has enabled OnLive to improve its underlying technology and systems.

It also enabled the improved of the real time adaptive aspects of the system and provide valuable feedback to the carriers, games providers and ISPs.

In March of this year OnLive announced it will launch it service in June 2010.

Other Applications and Enterprise Solutions

As the OnLive platform has been initially targeted for the consumer gaming experience, other applications such as media distribution and playback are easily incorporated and blend well with the cost and performance targets of the consumer platforms.

Further this technology (low latency, high quality, cloud based) that enable high end gaming applications to be provisioned over broadband networks to inexpensive widely available PC and Mac platforms also enables cloud based "SAS" type implementation of high end, expensive, piracy prone design, graphics, engineering and architectural applications. These applications being cloud based can leverage off the OnLive infrastructure, and reduce the end users cost of high end PCs and support. Work continues by OnLive in this area.

Summary

OnLive has developed and implemented the technology that enables cloud based real time interactive gaming. This service can be widely available via exiting broadband connections, and has been optimized for minimal latency, high quality over these networks.

Further the system is easily integrated into the consumer environment and provides unique features and availability of games, products and social interaction. The business issues, games availability and features all mesh well and represent an marked improvement for both the consumer and the games providers.

OnLive is among the first of such services that will become widely available taking advantage of broadband and will help drive the uptake and demand of the consumer for improved bandwidth and quality of broadband. The system allows has unique advantages for business and enterprise applications.