A JOURNEY TOWARD A COMPUTE CONTINUUM:
CLIENT-AWARE CLOUD SERVICES FOR SMART CLIENTS

One of the primary objectives of mobile Internet services is to deliver a compelling user experience. The key is to deliver the services that best fit the devices that serve the end users and make the users have a consistent continuum of experiences across these devices. As part of the much grander vision of a compute continuum, client-awareness and HTML5 are critical architecture building blocks. This article focuses on the architectural essence of the compute continuum and how client-aware cloud services for smart devices fit into the grand vision. Since HTML5 is a key technology for cross-platform content delivery from PC to tablet, TV, and smart phone, this article also focuses on key technology challenges of cloud-awareness using HTML5 technology.

Introduction

The compute continuum is Intel's vision of cloud computing to enable the seamless computing to cross different device form factors, for both enterprises and consumers. In this article, we introduce the emergence of client-aware cloud to fulfill the vision of the compute continuum, the technical challenges of the client-aware cloud, and Intel's initiative to explore and solve those challenges. Combining with the HTML5 initiative from W3C, we also introduce how client-aware technology will be shaped in near future to present the seamless Compute Continuum.

The Road to the Compute Continuum–Architecture Tenants

We constantly understand how people are currently using connected devices and looking for ways to make connected devices work better, and work better together.

The Cross-Device Continuum Experience

The compute continuum is an Intel-wide initiative aimed at delivering a range of seamless, personalized, and protected cross-device experiences for end users using their preferred devices. Intel is laying the groundwork in 2012 for bringing exciting and new ways for users to enjoy connected device experiences.

There are two different experiences on the compute continuum: enterprise IT and consumers.

From the enterprise IT perspective, by taking advantage of a combination of technology trends and emerging computer models, such as ubiquitous
Internet connectivity, virtualization, and cloud computing, there is an opportunity to proactively address changing user requirements and redefine the way IT provides service. This represents the next major change in the way that employees will use technologies. Key aspects include the following:

- Users have access to corporate information and IT services securely from anywhere on any device at any time, no matter whether it is personal or corporate-owned.
- Multiple personal and corporate devices work together seamlessly.
- Corporate information and service are delivered across these devices while the information continues to be protected.

For consumers, the compute continuum as one of the smart technologies and services is becoming the industry trend. Consumers are looking for these to blend into their everyday work, lives, and activities, while maintaining privacy and security. Users can experience the compute continuum during their time of entertainment, work, or learning, by enabling simple, secure interactions between varieties of devices across a broad spectrum of such usages. These interactions enable the users’ devices to collaborate, to overcome the limitations of a single computing platform, and to enjoy a familiar and consistent experience.

There is no singular usage that compute continuum delivers, but it should include but not be limited to:

- **Device Pairing and Composition.** Every device can seamlessly connect to other devices and share resources, securely.
- **Context Awareness.** Every device is aware of the user and is able to do actions on his/her behalf.
- **My Media, Anywhere.** Every device gives easy means to purchase, access, and enjoy premium content.
- **Identity and Security.** Every device understands and protects the user’s identity.
- **Natural UIs.** Every device provides intuitive modes of interaction.

**Intelligent Content Delivery**

Currently, most users have more than one end-point device, such as a PC, tablet, smart phone, or smart TV. How to share, sync, or pass control among these devices is becoming more and more important. The essentials of intelligent content delivery include:

- Allowing access to content from another device for sharing/viewing.
- Controlling an application on another device.
- Consuming an application from another device.

Furthermore, a smart enough mechanism may process the data before it is delivered. For instance, based on its knowledge of the client, the cloud could
Eventually, we should have an elastic and on-demand IT environment that supports all kinds of devices and various computing models, and it will be client-aware. That will be the Compute Continuum!

**Why a Client-Aware Cloud Is Important**

A client-aware cloud is an intelligent cloud infrastructure that can deliver cloud-based applications taking into account the capabilities of the device at hand.

Today’s users are increasingly likely to utilize multiple devices, including smart phones, tablets, and PCs, to access information. They embrace new
applications and devices in their daily life and expect those same capabilities to be available at work. Yet when it comes to the ability to access, display, manipulate, or secure data, some devices are clearly more capable than others.

The reality is that different devices may have different capabilities, such as different screen size, graphics capacity, and security level. Unfortunately, most Internet services are dumbed down. They probably can recognize the screen size but may be unable to determine and then take advantage of other local capabilities.

End users and cloud providers each have unique needs. Rather than relying solely on the data center to drive cloud capabilities, Intel believes that there’s a better and more balanced approach. By taking advantage of the client-aware cloud, service providers have greater flexibility to optimize application delivery. End users can also benefit from improved experience.

**User Benefits**
The user benefits of a client-aware cloud are:

- **Responsiveness**: for cloud-based applications, response time depends on multiple factors including network, performance, application performance, and cloud infrastructure performance. Enabling client-aware or even client-side execution for cloud-based applications can help improve end-user experience, in particular, for applications that are compute-intensive or bandwidth-constrained.

- **Productivity**: client-aware cloud services may allow users to access data either online or offline. Furthermore, once reconnected, applications can synchronize the change since last connection.

**IT Benefits**
The IT benefits of a client-aware cloud are:

- **Application delivery**: the ability to execute a given application or portions of it on the end point provides IT with an additional option to improve application delivery.

- **Flexible architecture**: with profiles exposed to cloud-aware services, deployment becomes easier because of the reduced dependence on the architecture. For example, organizations that have complex, sensitive, or network-constrained applications may find it hard to take advantage of the cloud because of the need for such things as a high bandwidth network connection, a secure communication tunnel, or dedicated client resources. With a client-aware cloud, there may not be so much dependency on the architecture, as the intelligent cloud can detect a client’s capability and then determine how the application should run accordingly, such as providing a lower resolution picture over a poor network connection or providing a sensitive application with additional authentication and authorization in an untrusted network.
Technical Challenges

Intel defines client awareness\(^2\) in terms of:

- **Compute performance**, which comprises the local performance capabilities in the device, mainly the processor and graphics
- **The device context**, which refers to the ability of the cloud to detect the end-user’s compute environment (that is, network connectivity and bandwidth, power source and battery level, and so on)
- **Security capabilities**, which is being aware of client security features and determining what applications can be delivered

To make the cloud-aware clients’ performance, context, and capabilities exposed to the cloud may bring some challenges:

- What’s the programming environment inside the client and how can the cloud discover its information?
- How does the cloud know the client’s information?
- How to make sure the client security (protecting virtualized applications, managing systems with full disk encryption) and how to make sure processor-intensive encryption and decryption does not impact responsiveness too much.

As the industry leading provider of silicon building blocks that go into virtually every element of cloud, including client, server, storage and networking, Intel is investing in developing application programming interfaces (APIs) that allow developers to enable cloud-based applications to detect the capabilities of the local client device. Since the Web is the major container of clients of the cloud, Intel has unveiled the Intel Web APIs Connector to enable Web applications to detect endpoint information via JavaScript inside the browser, which contains the following APIs\(^3\):

- Network bandwidth API
- Processor performance API
- Battery life API

The Intel® Core™-based vPro™ PC has hardware-based manageability and security, which can comply with the security requirements to run critical applications locally. Table 1 shows a high level overview of the capabilities\(^2\).

Based on these Intel Web APIs and the Intel Core-based vPro PC, there are two great pilot projects happening in the industry\(^4\):

- Gproxy’s Client Device Score (CLIDES) can rank a client device’s capabilities including CPU, CPU load, connection type, bandwidth, and screen resolution. The score can assist the client web application to optimize the user experience
- NetSuite Ecommerce Cloud Platform offers enhanced web UI optimization, shopping cart optimization, preserving user-state and improving Web analytics via Intel Web API.
These pilot projects are excellent examples of how to address the technical challenges mentioned previously. These challenges can definitely be solved by innovations based on existing techniques:

- **Client programming environment**: Web environment and JavaScript is a proven way to establish the client programming environment
- **How the cloud knows the client’s information**: Http (Ajax) is a convenient way to bridge cloud and client
- **Client security**: The CPU API can determine whether the client processor is an Intel® Core™ vPro™ processor, which enables more processing to be done locally, and speeds up encryption and decryption via the Intel AES instruction set.

Another non-technical challenge is how the main Web browsers should implement similar APIs to benefit enterprise IT and the consumer broadly. Since the World Wide Web Consortium (W3C) is the main international standards organization for the Web, we’d like to introduce W3C’s standardization roadmap to the client-aware cloud, combining with HTML5 evolution.

### The Role of HTML5 for the Client-Aware Cloud

Web technology has gone through numerous evolutions since the early beginnings of the 1980s. Its Episode 5 starts at 2007 when the HTML5 specification was adopted as the starting point of the new HTML working group of the W3C.
The emergence of HTML5 opens tremendous opportunities to the Web, which is no longer limited to only make up the content delivered from server, but can also access and fully utilize the client’s capabilities inside web runtime.

HTML5 is not a single technology, but a collection of improvements to HTML and keeps expanding and migrating. Currently HTML5 already covers canvas, video/audio playback, offline capacity, drag-and-drop, cross-document messaging, history management, web storage, geolocation, Web SQL, index DB, file and file system, web socket, and so on. W3C is actively driving and standardizing more specifications to expand broader device capacities. For example, W3C WebRTC (Web-based Real Time Communication) API is a natural expansion from HTML5 video/audio playback to a bidirectional video/audio chat. Another greater example comes from W3C Device APIs Working Group’s DAP API serial, which covers battery status, contacts, media capture, network information, sensor, vibration, web intents, calendar, and permissions APIs.

Therefore, as illustrated by Figure 2, HTML5 would play a significant role in achieving a compute continuum experience by exposing the client context information to client-aware cloud services, consuming the customized contents delivered from remote peers, and generating suitable contents as needed to end users or cloud. In other words, HTML5 is the essential technology to enable smart clients to address the challenges raised to client side due to client-aware services.

According to W3C’s announcement that the major standards of HTML5 will be ready in July 2014, W3C believes that the Open Web Platform including HTML5 and various APIs will be more mature and have greater interoperability in the near future. Industries are turning to the Web as the platform of choice for integrating diverse devices, services, and business models. There will be an explosion of HTML5 feature adoption in upcoming years. HTML5 definitely is on the road to playing a more and more important role for client-aware cloud for the following reasons:

- HTML5 provides standardized client status description
- HTML5 provides standardized client’s web runtime capacity description
• HTML5 provides standardized network information description
• HTML5 provides standardized service and application description

HTML5 for Device, Service, and Application Discovery
W3C Web and TV Interest Group[8] is the pioneer W3C group that was formed in 2010 with one of its missions being to address and standardize the discovery gap and API for audio-visual content.

W3C Web Intents Task Force[9] is a joint task force of the W3C Device APIs Working Group and the W3C Web Applications Working Group[10] that focuses on service discovery and lightweight RPC mechanism for web applications. First web intents edit draft was published in 2012 and has enormous industry impact to drive the browser vendors’ implementations.

HTML5 for Customized Content Delivery
HTML5 makes web-based bidirectional content delivery more realistic.

HTML5 Client as a Content Consumer
A Web-based customized content delivery capacity and usage model is greatly enlarged by HTML5 in various ways:

• The W3C Network Information API[11] enables the adaptive content delivery model according to different network environment.
• The W3C Geolocation API[12] can identify the location of client, which makes location-based content delivery a reality.
• The W3C Battery Status API[13] exposes the battery status of a client. Based on the client’s power information, content delivery and consumption can be more power efficient.
• HTML5 Web Sockets[14] and Web-based Peer-to-Peer connection API take bidirectional communication between client and cloud to the next level.
• HTML5 video and audio[15] gracefully enables multimedia content consumption. With the help of canvas, more usage models can be implemented in a much richer way than ever before.
• HTML5’s offline capacity, the Indexed Database API[16] significantly enriches the local cache capability of web applications, which is key to a flexible and efficient content delivery to reduce the dependency on a network connection.
• The Web Audio API[17] makes it possible to offload more server-side audio processing to the client side.
• The WebGL API[18] makes it possible to interpret 3D semantic content and render it inside an HTML5 client.

The HTML5 Client as a Content Producer
Benefitting from HTML5, the Web-based client has also evolved to be a powerful content producer.
The WebRTC API utilizes the client’s microphone and camera to capture audio and video streams, sends out captured media stream, and presents the received stream at the same time.

W3C GetUserMedia and HTML media capture API standardizes the HTML5 client’s audio and video capture capacity with pre- and post-media-content-processing capacity.

Conclusion

In this paper, we first introduced Intel’s vision of the compute continuum for the enterprise and consumer. We then discussed why the client-aware cloud is important for the compute continuum and the technical challenges ahead. We illustrated Intel’s exploration and initiatives to help address these roadblocks. Finally, we briefly described W3C’s HTML5 and WebAPI efforts, which as the future client technology, would deliver the compute continuum by naturally collaborating with client-aware cloud.

References


[16] Indexed Database API, W3C Working Draft 06 December 2011 http://www.w3.org/TR/IndexedDB


Author Biographies

Jennifer Jin is the project owner for the Intel DCSG/CIG/IASI Compute Continuum PRC. She set up the Compute Continuum Engineering Experience Lab in Folsom in 2011. She supports the product group on ecosystem solution assessments to give the architects input on Intel’s Compute Continuum project design.

Jonathan Ding (Junyong, jonathan.ding@intel.com) is a software engineer in the Software and Services Group at Intel. His expertise covers browser, Web runtime, HTML5, and related service frameworks.

Samuel Xu (Shu) is a staff software engineer at the System Optimization Technology Center in the Intel Software and Service Group. Samuel focuses on HTML5 optimization and WebAPI development. He holds an MS degree in Computer Science from the Chinese Academy of Sciences. His email is samuel.xu@intel.com.