A Distributed Cloud Architecture for Mobile Multimedia Services

Muhamad Felemban, Purdue University
Saleh Basalamah, Umm Al-Qura University
Arif Ghafoor, Purdue University

Abstract

Mobile cloud computing is emerging as a new paradigm for supporting a broad range of multimedia services. MCC alleviates the burden of storage and computation on mobile devices. In this article, we describe design requirements and an architecture for MCC. The novelty in this architecture is an integrated cloudlet and base station subsystem that can meet application-level quality of service requirements and allow mobile resource provisioning close to the user. We present a layered architecture for MCC that elucidates the required functions and protocols. We also propose a connection handoff mechanism among cloudlets and discuss related resource management challenges for MCC.

The rapid increase in the use of multimedia services and applications on mobile devices has led IT companies to evolve their technologies to cope with the multimedia requirements. Cloud computing, which is a new content-centric paradigm, can fulfill these requirements by providing data and computing resources on demand. It allows users to access infrastructure, platforms, and software at low cost. For example, Amazon provides its users personal storage spaces with Simple Storage Services (S3) and ability to perform extensive computation on the data using Elastic Compute Cloud (EC2). Likewise, Google's App Engine allows users to develop and deploy their applications on Google's platform.

On the user's side, the demand for mobile services is rapidly growing. It is expected that the number of mobile users will exceed 800 million by 2015 [1]. However, mobile devices have several limitations, such as short battery life, and limited storage and computation power. To address these limitations, mobile cloud computing (MCC) is presented as an integration of cloud computing and mobile technology. MCC is defined as the infrastructure where both data storage and processing are offloaded from mobile devices to the cloud, bringing mobile applications a much broader range of users [2]. MCC overcomes the limitations of mobile devices by moving the data processing and storage to the powerful platforms located in the cloud.

MCC can provide an infrastructure for various mobile applications such as emergency response management, large scale event planning such as Olympics, mobile gaming [3], and interactive video streaming [4]. It can support multimedia services in scalable mobile environments. For example, MCC can be used along with urban transportation systems [5] to provide updated traffic information for drivers. Traffic data and scenes are collected and processed on the cloud to make traffic decisions. Such information forms multimedia data that can be accessed by mobile users. In large-scale event planning, distributed cloud can provide a variety of multimedia data and services to fulfill the needs of tourists. Such information can include precomposed multimedia brochures, tour guide videos, and images. Such massive data can be archived and delivered by distributed clouds constituting the MCC architecture.

The objective of this article is to address the challenges of mobile services in terms of data management and networking, and develop an architecture that can lead to the design of MCC. In particular, our focus is on the retrieval and communication of preorchestrated multimedia data, which imposes several resource management challenges on designing an MCC architecture. The main technical challenges are highlighted as follows,

- *Heterogeneous networks and QoS requirements*: Multimedia services may span multiple heterogeneous network protocols, such as second generation (2G), 3G, and Long Term Evolution (LTE), with different quality of service (QoS) requirements. Dynamic resource allocation protocols are needed to meet these.
- *Heterogeneous multimedia data*: Distributed mobile services such as video over IP, multimedia streaming, and photo sharing can consist of various types of data including video, audio, and images. Such data may have different delivery requirements that need to be synchronized to provide coherent information to mobile users [6].

The proposed architecture entails multiple layers of functionality and addresses the QoS requirements and resource management challenges in terms of end-to-end delay, jitter, buffering, and bandwidth. A novel feature of this architecture is the integrated subsystem of cloudlet and base station, which provides a "close-to-the-user" proxy system functionality that ensures seamless delivery of data that meets QoS requirements. This functionality is achieved by dynamic allocation of resources, including buffers and radio frequency (RF) channels, synchronization of multiple streams, and seamless handoff of streams among base stations. In this article, we first present a cloud architecture for MMC and its components. Then we provide a layered architecture of the MCC and the handoff procedure. Finally, we present resource management challenges and performance assessment of the MCC architecture.

0890-8044/13/25.00 © 2013 IEEE

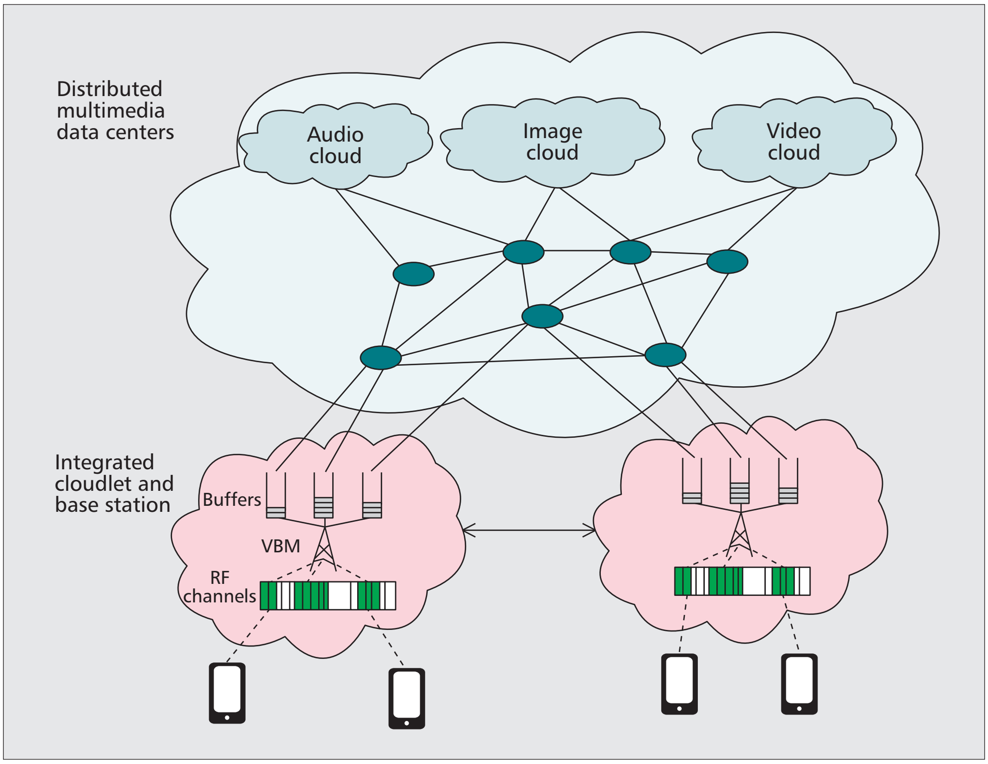


Figure 1. *Distributed cloud architecture for multimedia services.*

Cloud Architecture for Mobile Multimedia Users

In Fig. 1, we perceive a distributed multimedia cloud architecture for mobile users that consists of distributed multimedia data centers, and integrated cloudlet and base station. The proposed MCC supports this environment. It can be noted that MCC architecture has two major components; a set of distributed multimedia data centers and an integrated cloudlet and base station subsystem. We assume that the data centers act as repositories for multimedia information. A multimedia data center in the cloud retrieves the requested data from the database and communicates it to the cloudlet over the Internet. The cloudlet then transmits the multimedia information to mobile users on the RF channels. The MCC architecture overcomes the high latency that results from the direct communication between a large number of mobile users and multimedia data centers. The cloudlet ensures QoS to mobile users by managing the interface between the Internet and the mobile network. It coordinates with allocation of RF resources through its local base station. We now briefly discuss the two main components and accordingly discuss the layered architecture.

Virtual Multimedia Data Centers in MCC

Multimedia data is not monolithic in nature and can be composed of several objects that are stored in different multimedia cloud data centers, as depicted in Fig. 1. For example, real-time multimedia information can be streamed to users containing precomposed data including video, audio, and text. Multimedia information consisting of different data is represented as a multimedia document. Figure 2 depicts the composition of distributed multimedia objects into a single multimedia document. A document needs both spatial and temporal composition. Temporal composition refers to the process of synchronizing multiple streams of multimedia data, whereas spatial composition allows superposition and overlay of multimedia data.

Temporal composition of multimedia objects requires synchronization among data streams of a document. Temporal synchronization can be achieved at a level of fine-grained data unit referred to as a synchronization interval unit (SIU). A single multimedia object is transmitted as a stream of SIUs. Since the SIU is the basic unit of playout, it is essential that the SIU's playout deadline is met. Several document specification models that specify the temporal synchronization and quality of presentation (QoP) requirements exist in the literature. One such model is the object composition Petri-net (OCPN), which uses an augmented Petri-net model [7]. OCPN captures the synthetic relationships between the objects and identifies media synchronization points. For the OCPN model, a schema to maintain the temporal relationships between objects can be constructed. The schema is used for storing and retrieving the objects from the distributed data centers. OCPN also defines a specification of QoP requirements for multimedia communication that includes speed ratio, utilization, average delay, maximum jitter, maximum bit error rate, and maximum packet error rate. Multimedia applications might tolerate some of the QoP based on delay sensitivity and error tolerance requirements. Real-time video streaming, for example, requires a high data rate, and moderate delay and jitter. However, other multimedia applications cannot tolerate high delay and jitter, such as interactive multimedia applications.

From the data centers perspective, the aforementioned technical challenges of providing heterogeneous network and QoS requirements can be addressed by employing distributed multimedia data centers that store and deliver the required multimedia information. To increase the performance, multimedia information can be replicated at various data centers, which requires virtualized access and retrieval mechanisms. Korotich and Samaan [8] proposed a service virtualization architecture that hides the selection and configuration of the multimedia data center from end users. For the MCC architecture, a similar virtualization mechanism can be used that is composed of a virtual data center (VDC) mapped to a single

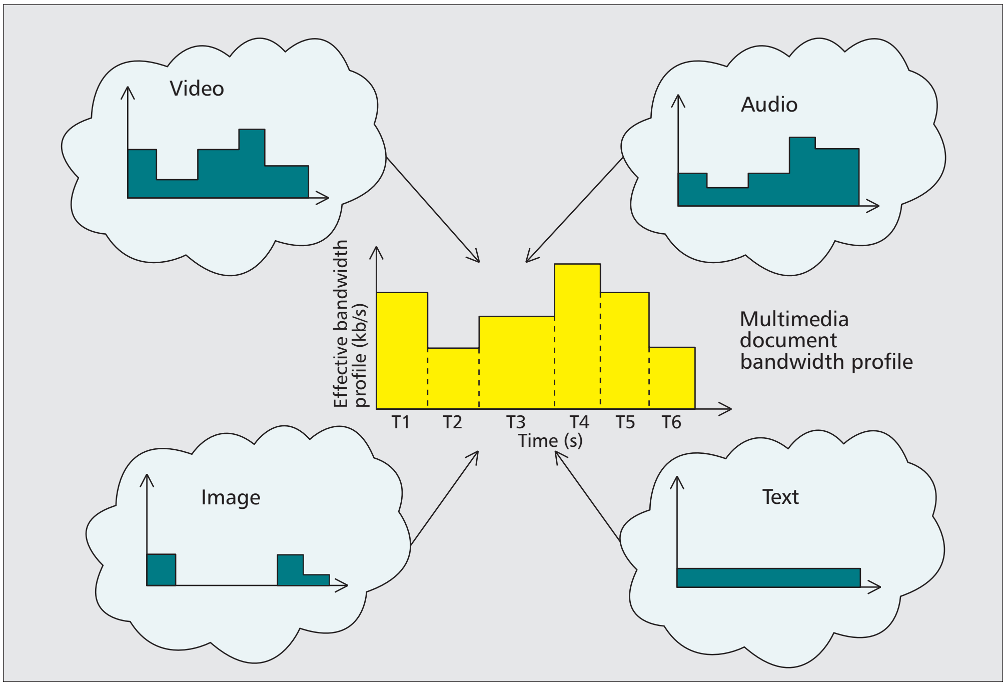


Figure 2. *Effective bandwidth requirements of a multimedia document.*

or group of physical data center(s) that are used to retrieve the required multimedia object. The main component of the architecture is the virtualization manager (VM), which maintains a hierarchy of VDCs and maps the end user's request to a VDC. The request mapping is achieved using a service broker that receives the request and locates the VDC where the object is stored based on the input and output parameters. Once the VDC is located, the VM delivers the multimedia object to the end user's cloudlet.

Integrated Cloudlet and Base Station Architecture of MCC

High latency of packet and jitter delay are fundamental obstacles to developing mobile multimedia services. Due to the limited bandwidth of 3G and 4G networks, widely dispersed, resource-rich, low-cost cloudlets can be deployed close to the mobile devices. Cloudlets can perform distributed synchronization and composition of multimedia objects, as shown in Fig. 3, to reduce the burden of computation from the mobile devices.

The main function of the cloudlets is to provide a seamless interface between two diverse networks: the Internet and mobile networks. For this purpose, cloudlets can be integrated with mobile base stations to form a logical entity that can provide seamless end-to-end synchronization of multimedia streams to users, as shown in Fig. 3. Accordingly, in coordination with its local cloudlet, a base station manages the outbound RF channel to support multimedia connection for its mobile users. For mobile multimedia services, the RF channel is a precious resource and therefore needs to be managed intelligently. The policy for channel allocation, however, can change dynamically due to various factors such as the number of users being served concurrently by the base station, the changing level of concurrency of multimedia objects, and managing migrated "calls" from neighboring base stations. The channel allocation policy can be designed based on the assumption that the requested multimedia data are delivered to the base station prior to the transmission to mobile users. However, this assumption may not be valid because of the non-deterministic delays data packets may encounter over the Internet. Therefore, buffering at the base station is required to compensate the jitter delays to avoid discontinuity of presentation at the mobile devices. The buffering requirement can be fulfilled by the cloudlet. The overall functionality of integrated cloudlet and base station is summarized as follows:

- Providing synchronization and composition functionality for multiple multimedia streams
- Handling speed mismatch between the Internet and the mobile network
- Managing handoff calls and multimedia sessions
- Dynamically allocating resources to mobile users

To manage the aforementioned functions, software-defined networking (SDN) technology can be utilized, which can allow separation of the control plane from the data plane [9]. In essence, the base station, as part of the control plane, implements session setup and teardown, paging, session handoff, and RF channel allocation protocols. On the other hand, the cloudlet manages data plane functions in terms of performing stream synchronization, buffering, and data forwarding to mobile devices. In this manner, SDN provides flexible management of the integrated cloudlet and base stations as a value-added service without interrupting basic operations of the base stations. Customization of the integrated cloudlet and base stations to support a wide range of mobile users, applications, and services can be realized through virtualization [10]. We elaborate on the use of SDN and virtualization of cloudlets/base stations in the following section, where we propose a functional layered architecture for MCC.

Functional Layered Architecture for MCC

In Fig. 4, the layered architecture of MCC is presented. Each layer includes a set of functions and protocols. The operation at each layer is performed in three phases: establishment, activation, and termination. Initial setup of a multimedia session between a user and MCC takes place during the establishment phase. After this setup, the multimedia data is transferred during the activation phase. Finally, the teardown of the session is done in the termination phase. In the following section, we discuss each layer in detail.

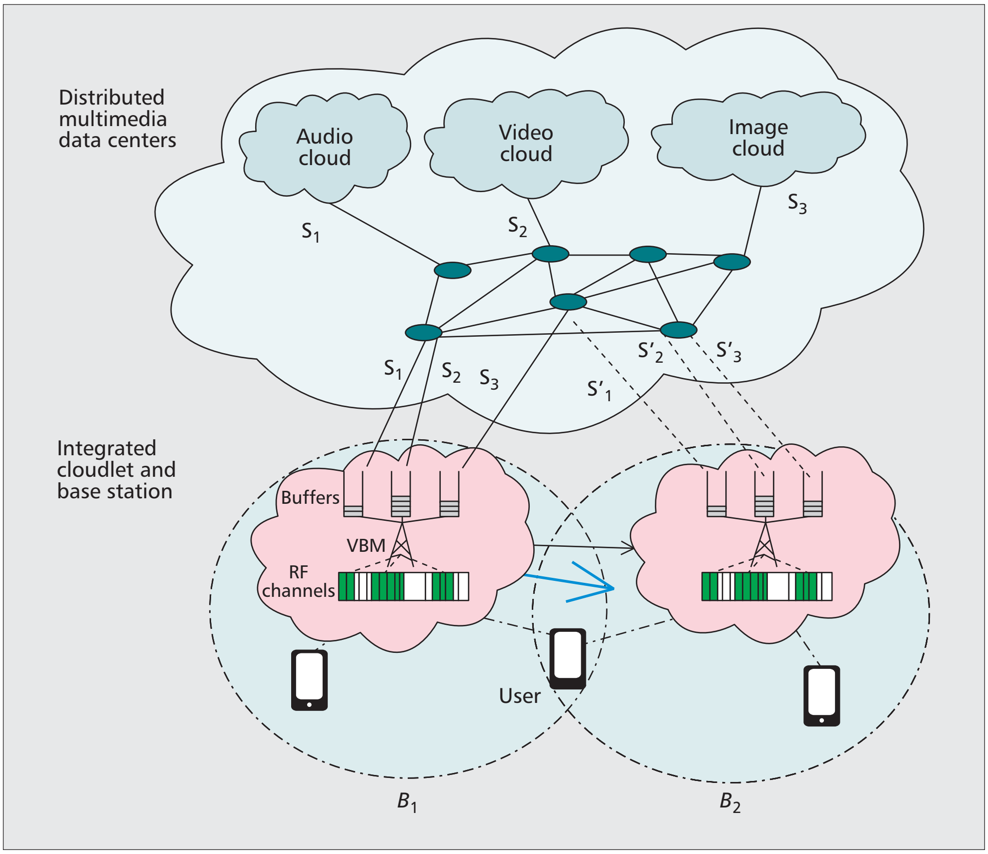


Figure 3. *Session handoff between cloudlets. The figure depicts three data centers, originally S_1, S_2, and S_3 in B_1 before migration. S'_1, S'_2, and S'_3 are the migrated sessions after the user migrates to* B_2.º

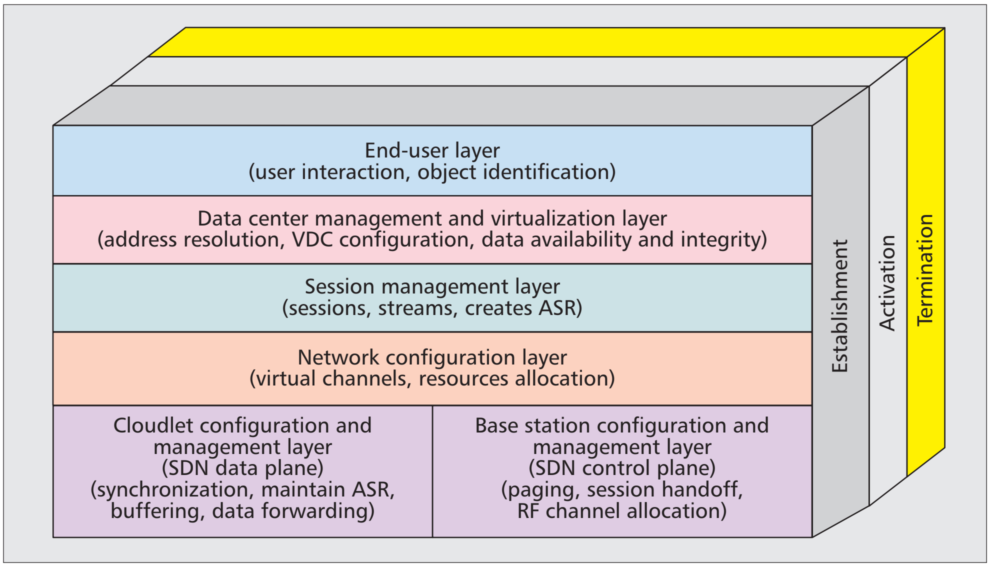
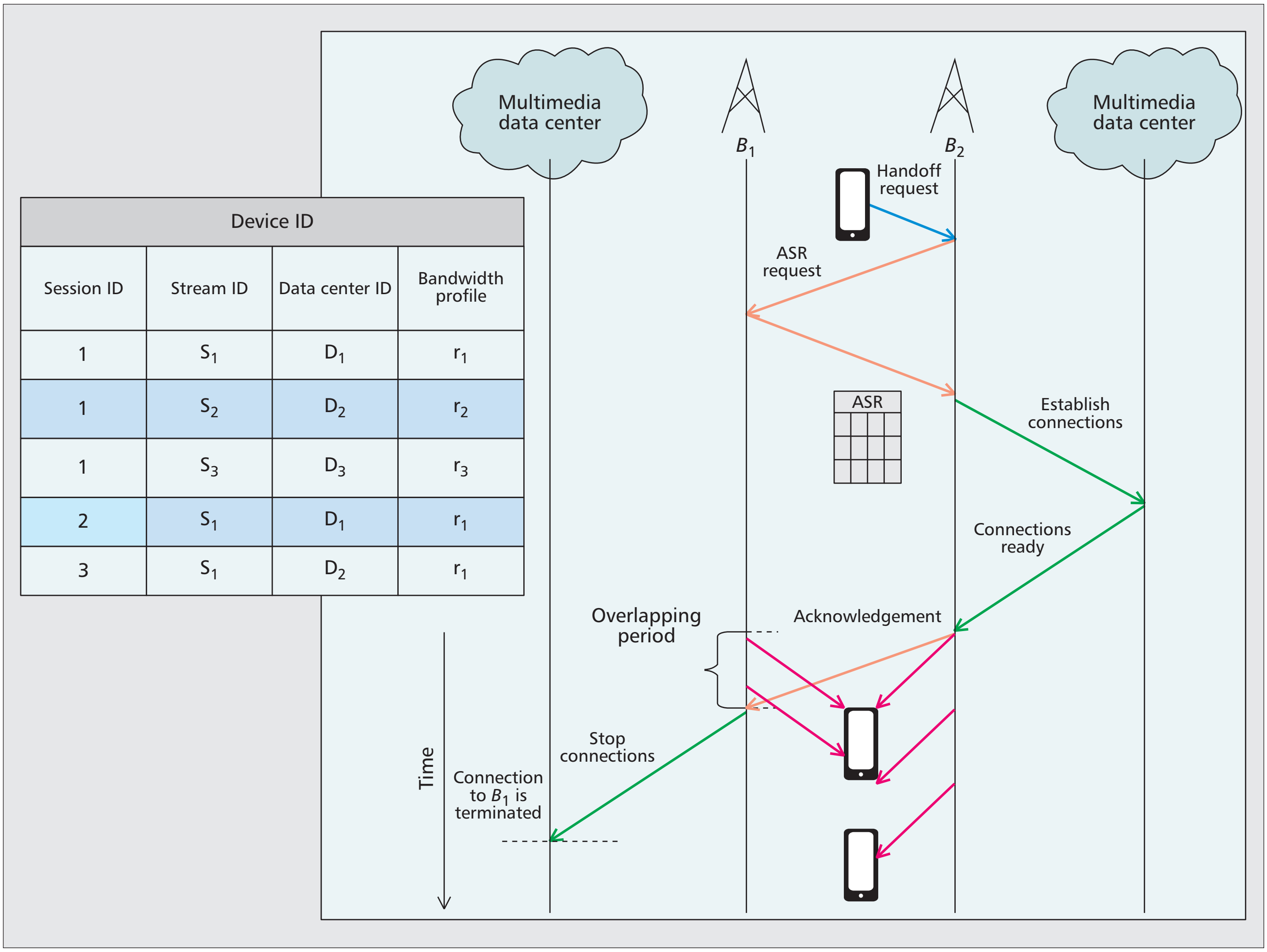


Figure 4. *Functional layered architecture for MCC.*

End User Layer

The end-user layer provides a graphical user interface (GUI) to facilitate direct user interaction with multimedia applications. The end-user layer identifies the objects and their QoP parameters by processing a user's requests. In addition, it allows end users to upload new multimedia data, and modify the relevant QoP parameters and authorization information. Such operations are managed through the cloudlet associated with the base station where the mobile user initiated the session.

In a multimedia document, multimedia objects may have varying bandwidth requirements, as shown in Fig. 2. It can be noticed that the overall bandwidth of the document and the resource requirements may change considerably over time depending on the concurrency level of the objects. In order to ensure QoP requirements, the underlying network, including the Internet and mobile networks, must dynamically allocate



Device ID			
Session ID	Stream ID	Data center ID	Bandwidth profile
1	S_1	D_1	r_1
1	S_2	D_2	r_2
1	S_3	D_3	r_3
2	S_1	D_1	r_1
3	S_1	D_2	r_1

Figure 5. *Handoff signaling procedure and active sessions record.*

sufficient resources. The requirements can be specified in terms of end-to-end peak or average bandwidth needed for transferring the objects. The document model can provide the bandwidth requirements of each object. It can be noticed that the profile of each object is maintained by its local data center. Accordingly, at the time of session establishment the bandwidth profile of the entire multimedia document becomes available to the cloudlet. In MCC, this profile is used by the cloudlet to allocate resources efficiently to ensure the desired QoP. An end user may establish multiple sessions with varying bandwidth profiles. Effectively, the integrated cloudlet and base station serves as an interface between the Internet and the mobile users. We assume that the Internet (the network between data centers and cloudlets) is resource-sufficient and has enough resources to guarantee the QoP required by the multimedia services.

Data Center Management and Virtualization Layer

Data center management and the virtualization layer provides the management functionality for the distributed objects and maintains their location in terms of data center IDs. As discussed earlier, virtualization allows resolution of logical addresses to physical locations where services are invoked by the user. In essence, the service broker performs object address resolution by identifying the specific VDC once the multimedia object is identified. The configuration, management, and mapping of VDCs to the physical data centers in the cloud is achieved in this layer. Moreover, VM manages data integrity and availability among VDCs.

Session Management Layer

Once the distributed objects are identified by the previous layer, session management layer establishes the corresponding streams from data centers. Each stream is identified by a unique stream ID. Multiple streams can form a multimedia session, which in turn is assigned with a distinct session ID. The session management layer creates a record of all active sessions. This record, called an active session record (ASR), is a table that has entries of active sessions' IDs, stream IDs that form the sessions, and the data centers' IDs of the streams, as shown in Fig. 5. In addition, an ASR contains an entry for the bandwidth profile of the objects in the session. The management of sessions is controlled by this layer. When session handoff takes place, the lower layer requests that the session management layer perform two functions:

- Terminate the streams and sessions supporting the migration process
- Reestablish sessions to the migrated cells and reroute data streams accordingly

The details of the handoff procedure are discussed later in this article.

Network Configuration Layer

The function of this layer is to establish and maintain virtual channels over the Internet. The establishment phase determines the routes for the virtual channels between data centers and cloudlets, and allocates sufficient resources based on the object's bandwidth profile and QoP requirements to ensure timely delivery of multimedia data to the cloudlets. Overlay networks are used in this layer to enhance multimedia QoP by:

- Discovering redundant paths between data centers and cloudlets
- Implementing routing policies that allow customized media delivery [11]

Cloudlet and Base Station Configuration Management Layer

The cloudlet and base station configuration and management layer has two components. The base station management layer works in the SDN's control plane by performing basic base station operations such as establishing and tearing down sessions with mobile devices, paging, allocating RF channels, and initiating a handoff procedure, as depicted in Fig. 5. On the other hand, cloudlet configuration and management layer handles inter-stream and intra-stream synchronizations to ensure continuity of presentation to the user. The function of this layer also includes aggregating the bandwidth profiles, allocating buffers, and forwarding data to mobile users. To manage the handoff process, the cloudlet configuration and management layer maintains the ASR of each mobile device. To allow seamless migration of sessions across cloudlets, the user can be provided with a virtualized service abstraction by the MCC. In the following section, we present the details of the handoff procedure and function of each layer in the handoff process as part of the virtualization service provided by MCC.

Session Handoff

A session is composed of one or more streams that may originate from different data centers. Routes between data centers and mobile users may change with the movement of users across multiple cells, as depicted in Fig. 3. To ensure smooth delivery of multimedia data to users, handoff and resource reservation mechanisms can be used to establish new sessions among data centers and a migrating user, as depicted by $S'1$, $S'2$, and $S'3$ sessions in Fig. 3. The handoff process is initiated when a mobile device moves out of reach of the base station into the coverage of the neighboring base station. There are two types of handoff procedures: hard handoff and soft handoff. In the case of hard handoff, there can be a short interruption time in the delivery of data streams during the migration. The interruption occurs when the first session terminates at, say, SIU_n, and the second session starts at SIU_{n+1}. However, any noticeable discontinuity is not favored by the user. On the other hand, soft handoff can avoid loss of data by allowing partial overlapping of the split SIUs. The overlapped data is then clipped according to SIUs from both streams.

A handoff procedure is initiated when the received signal level from the current base station, B_1, drops below a certain threshold. Subsequently, the mobile device identifies the base station, B_2, with the highest received signal level. The handoff procedure signaling is depicted in Fig. 5. The mobile device sends a handoff request to B_2 that includes B_1's ID. Once the request is received by B_2, it communicates with B_1 through the cloudlet configuration and management layer and sends an ASR request that contains the ID of the requesting device. B_1 replies back with the requested ASR to B_2, which in turn requests the session management layer to initiate the required stream connections to the multimedia data centers. In case of soft handoff, the original streams to B_1 are retained to avoid interruption. B_1 continues transmitting the streams to the mobile device until streams are established between B_2 and data centers. At that time, B_2 sends an acknowledgment to B_1. At the same time, B_2 starts transmitting data streams to the mobile device. When B_1 receives the acknowledgment, it stops transmitting data streams to the mobile device. B_1 requests the sessions and management layer to terminate the data streams of the migrated device. The overlapping period ensures a soft handoff and therefore continuity in the session.

Resource Management Challenges for Integrated Cloudlet and Base Station

As mentioned earlier, several resource management challenges need to be addressed while designing MCC architecture of Fig. 1. In this section, we present challenges related to managing two key resources, buffers and bandwidth at the integrated cloudlet and base station subsystem. Management of the resources to satisfy the QoP requirements can be formulated as an optimization problem, as illustrated in the following sections.

Virtualized Dynamic Buffer Allocation

Temporal intra-stream synchronization needs to be preserved in order to present the multimedia information correctly. For example, video objects require a certain playout rate to ensure continuity in the presentation. Jitter delays in multimedia streams occur when packets experience different delays while traversing from the multimedia cloud data centers to cloudlets over the Internet. To avoid discontinuity in presentation at the mobile devices, buffering by the cloudlet is used to compensate for jitter delays.

Inter-stream synchronization, on the other hand, preserves the timing relationship among multiple multimedia streams. Inter-stream synchronization is required to deliver a coherent multimedia document to the user. Multimedia streams flow along different routes over the Internet and experience different delays. Therefore, buffering is required to ensure inter-stream synchronization. However, buffer underflow and overflow can affect the QoP. Buffering underflow occurs when the session management layer transmits sessions in a just-in-time (JIT) manner. Streams in a JIT flow might experience unexpected network delay, leading to a playout deadline being missed, while buffering overflow occurs when the session management layer dumps sessions at full speed. One of the challenging issues in this regard is to provide an upper and lower buffering bound to support a large number of sessions.

In order to achieve inter-stream and intra-stream synchronization, cloudlets maintain virtual buffers for active sessions managed by a virtual buffer manager (VBM), as shown in Fig. 1. Buffers provide temporary storage for multimedia objects communicated over the Internet in each session in order to smooth jitter delays and facilitate inter-stream synchronization of multimedia data. As multimedia objects can arrive at a cloudlet ahead of their playout deadlines, they are buffered until playout time. In this case, buffer underflow is prevented. The VBM assigns free RF channels to the SIUs with looming deadlines. The assigned RF channel capacity may not be sufficient due to resource constraints, causing a rate mismatch between the arrival rate of SIUs from the servers and the outbound transmission rate onto the mobile networks. The VBM can dynamically allocate buffer to compensate for the rate difference between the Internet and the mobile network in order to avoid buffer overflow, which can lead to loss of data.

Dynamic RF Channel Capacity Allocation

Multimedia information, such as high-definition video, is characterized by high-bandwidth data transfers. Consequently, the management of RF channels in mobile networks is a significant challenge. This resource management problem can be

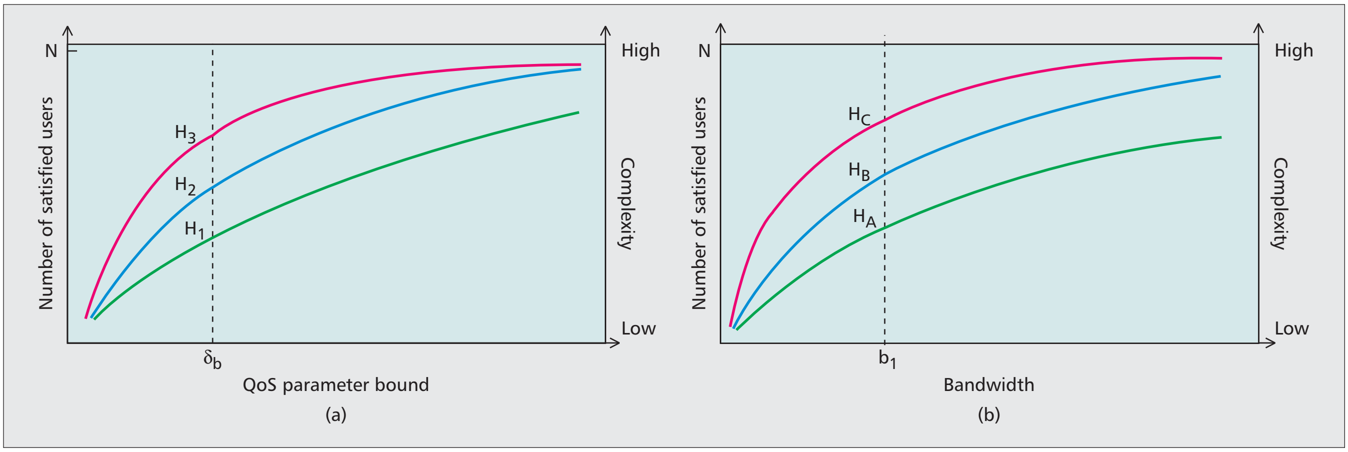


Figure 6. *Performance assessment of MCC architecture.*

posed as an optimization problem. For example, in RF channel allocation, let us assume that O_1^{Si}, O_2^{Si}, …, O_n^{Si} represent the concurrently transmitted objects in a document within session S_i. Let the corresponding bandwidth requirements be r_1^{Si}, r_2^{Si}, …, r_n^{Si}. Figure 2 shows that the aggregate bandwidth requirements changes with time at random transition points T_1, T_2, and so on. These transition points are called resource allocation decision points (RADPs) [7]. The cloudlet determines the resource requirements at these transition points.

Let I be the time interval between two consecutive RADPs, j and $j + 1$. The aggregate bandwidth requirement R^{Si} of the objects in interval I for session S_i is given by

$$R^{S_i} = \sum_{x=1}^{n} r_x^{S_i} \qquad (1)$$

Assuming multiple sessions are initiated by user, the aggregate bandwidth requirement R for that user is given by

$$\begin{aligned} R &= \sum_{\forall S_i} R^{S_i} \\ &= \sum_{\forall S_i} \sum_{x=1}^{n} r_x^{S_i} \end{aligned} \qquad (2)$$

If the channel capacity C in the base station is greater than R at a given time interval, the bandwidth requirements of individual session are guaranteed. However, the base stations might not be able to satisfy the bandwidth requirements due to the establishment of new connections. Then at least $(R - C)$. $|I|$ amount of information is dropped. Let δ_i^{Si} denote the dropping ratio of object O_i^{Si} in a session S_i, given by

$$\delta_i^{S_i} = \frac{\text{number of SIUs dropped in } O_i^{S_i}}{\text{total number of SIUs in } O_i^{S_i}} \qquad (3)$$

Shafiq *et al.* present a fair channel allocation policy for a single session and formalize it as a nonlinear programming problem [7]. However, the optimization problem dealing with multiple sessions for all the users can be solved in order to allocate RF channel resources to individual sessions and users.

Resource Management for Migratory Sessions

Mobile traffic load at base stations may vary dynamically according to a number of factors such as the varying aggregated bandwidth requirements of a session, the number of incoming sessions, and the number of migrated sessions from neighboring base stations. New sessions might be rejected if the channel capacity in the base station is not sufficient to accommodate more sessions. A migrated session is treated as a new session request that invokes the RF channel assignment procedure at the base station. To avoid dropping migrated sessions, channels are reserved in advance at all prospective base stations the user is expected to visit during the lifetime of the session. This can be done by delivering the bandwidth profile of a multimedia document to base stations in advance.

A user's mobility profile can be used to estimate the arrival and departure time at each base station using information such as size of base station, geographic location of each cell, and maximum speed of mobile users. The probability density function of the residency time T_R is given in [12] under the assumption that the session's duration is greater than the residency time in the base station's cell with radius R. Moreover, the mobile user is assumed to be traveling at a constant speed in the interval [0, Vmax]. The density function of the residency time T_R in the base station in which the session is initiated is given by [12]

$$f_{T_R(t)} = \begin{cases} \dfrac{8R}{3V_{\max}\pi t^2}\left[1-\sqrt{\left\{1-\left(\dfrac{tV_{\max}}{2R}\right)^2\right\}^3}\,\right] & \text{if } 0 \le t \le \dfrac{2R}{V_{\max}} \\ \dfrac{8R}{3V_{\max}\pi t^2} & \text{if } t \ge \dfrac{2R}{V_{\max}} \end{cases} \qquad (4)$$

In addition, the probability density function of the T_R in a base station where a handoff occurs is given by [11]

$$f_{T_R}(t)_{(forhandoff)} = \frac{3}{2} f_{T_R}(t) \qquad (5)$$

Using this function, the arrival and departure time of mobile users within a base station can be estimated based on the residency time T_R. Accordingly, a set of tuples consisting of the estimated residency time and the expected visited base station IDs can be maintained for each mobile user during the lifetime of the session.

Performance Assessment of MCC Architecture

For management of resources in the MCC architecture, various techniques can be implemented with varying degrees of performance [6, 7]. Several criteria can be used for assessment of such techniques. For example, one key performance criteria, given in the previous section (Eqs. 2 and 3), is to estimate the number of users (N) whose QoS requirements are satisfied by the heuristics employed at the bottom layer of the MCC architecture in Fig. 4. The general performance behavior of the architecture for such criteria is depicted in Fig. 6. As shown in Fig. 6a, for a given heuristic, the parameter N tends to increase as the QoS threshold increases. Such a threshold can be specified by the application. Here we use the data dropping ratio δ_b to illustrate this point. However, heuristics yielding higher performance generally entail higher complexity (e.g., H1 vs. H3). Figure 6b depicts another performance assessment of various heuristics in terms of change in N with varying degrees of availability of resources. It is intuitive that with the increase in the amount of resources (e.g. RF bandwidth), N also increases. Again, heuristics yielding high performance tend to have a high complexity. Note that high-complexity heuristics may not be desirable for real-time multimedia applications, resulting in a trade-off between the guaranteed QoS and the real-time performance of the related heuristics.

Conclusion

In this article, we have proposed a novel mobile cloud computing architecture for supporting mobile multimedia applications and services in mobile networks. The key part of this architecture is the integrated cloudlet and base station subsystem that provides a "close-to-the-user" proxy functionality and performs dynamic allocation of resources. In addition, we have presented a functional layered architecture that includes a set of functions and protocols to support multimedia applications and services. We have also presented the connection handoff mechanism among cloudlets and its related challenges. In addition, we have discussed prospective challenges in managing resources including buffer and RF channels.

References

[1] S. Zeadally, H. Moustafa, and F. Siddiqui, "Internet Protocol Television (IPTV): Architecture, Trends, and Challenges," *IEEE Sys. J.*, vol. 5, no. 4, 2011, pp. 518–27.

[2] H. Dinh *et al.*, "A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches," *Wireless Commun. and Mobile Computing*, 2011.

[3] L. Garber, "GPUs Go Mobile," *Computer*, vol. 46, no. 2, Feb. 2013, pp. 16–19.

[4] G. Lawton, "Cloud Streaming Brings Video to Mobile Devices," *Computer*, vol. 45, no. 2, Feb. 2012, pp. 14–16.

[5] R. Xue, Z.-S. Wu, and A.-N. Bai, "Application of Cloud Storage in Traffic Video Detection," *7th Int'l. Conf. Computational Intelligence and Security*, 2011, pp. 1294–97.

[6] T. D. C. Little and A. Ghafoor, "Spatio-Temporal Composition of Distributed Multimedia Objects for Value-Added Networks," *Computer*, vol. 24, no. 10, 1991, pp. 42–50.

[7] B. Shafiq *et al.*, "Wireless Network Resource Management for Web-Based Multimedia Document Services," *IEEE Commun. Mag.*, vol. 41, no. 3, 2003, pp. 138–45.

[8] E. Korotich and N. Samaan, "A Novel Architecture for Efficient Management of Multimedia-Service Clouds," *IEEE GLOBECOM Wksps.*, 2011, pp. 723–27.

[9] L. E. Li, Z. M. Mao, and J. Rexford, "Toward Software-Defined Cellular Networks," *Proc. 2012 Euro. Wksp. Software Defined Networking*, 2012, pp. 7–12.

[10] M. Satyanarayanan *et al.*, "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, 2009, pp. 14–23.

[11] M. Venkataraman and M. Chatterjee, "Quantifying Video-QoE Degradations of Internet Links," *IEEE/ACM Trans. Net.*, vol. 20, no. 2, 2012, pp. 396–407.

[12] D. Hong and S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-Prioritized Handoff Procedures," *IEEE Trans. Vehic. Tech.*, vol. 35, no. 3, 1986, pp. 77–92.

Biographies

MUHAMAD A. FELEMBAN (mfelemban@gmail.com) received a B.S degree in computer engineering from King Fahd University of Petroleum and Minerals, Saudi Arabia, in 2008, and an M.S degree in computer science from King Abdullah University of Science and Technology, Saudi Arabia, in 2011. He is currently working toward a Ph.D. degree in the School of Electrical and Computer Engineering at Purdue University. His research interests include data streams management and underwater acoustic networks.

SALEH BASALAMAH is an associate professor at Umm Al-Qura University. He has an M.Sc. from the University of Bristol and a Ph.D. from Imperial College London. His research interests include computer vision and multimedia.

ARIF GHAFOOR [F] is a professor in the School of Electrical and Computer Engineering at Purdue University. His research interests include multimedia information systems, database security, and distributed computing.