2013 IEEE International Conference on Cloud Engineering

Energy Saving in Mobile Cloud Computing

Mazedur Rahman
San Jose State University, USA
mazedur.rahman.liton@gmail.com

Jerry Gao
San Jose State University, USA
jerry.gao@sjsu.edu

Wei-Tek Tsai
Arizona State University, USA
wtsai@asu.edu

Abstract – Green computing and energy-saving have been important and hot research topics in mobile computing and cloud computing. In the recent years, many existing research results have been published to address energy saving issues and challenges in cloud computing or mobile computing. Today, the significant increase of mobile users and strong business interests in cloud computing and services drive an emergent opportunities and needs in mobile cloud computing. This brings a new attention to energy saving in mobile cloud computing and services. This paper first presents the research scope and classified issues in energy saving in mobile clouds. Then, it reviews the existing research results and techniques, and examines their strengths and weaknesses. Finally, the paper offers observations, and identifies the open issues and needs for future research.

Keywords – *Mobile Cloud, Energy Efficient, Green Cloud, Energy Saving, Cloud Computing*

1. INTRODUCTION

Green computing has been a popular topic. A new research report, titled as *"The Future of Virtualization, Cloud Computing and Green IT – Global Technologies & Markets Outlook – 2011-2016*", concludes that total green IT market forecast is going to grow by 40.5% CAGR from 2010 to 2016, with an aggregate $16 billion over the same period. A recent report from Forrester predicted that the market for green IT services could grow from about $450 million in 2008 to $4.8 billion in 2013 before leveling off in 2014. Visiongain's latest management research report on Mobile Cloud Computing Industry Outlook Report (2011-2016) examines the market, and discusses about how to leverage mobile cloud solutions to enable enterprise business strategies and models. According to a new white paper by FeedHenry and the Yankee Group, 42 percent of businesses are looking to mobilize their field and sale operations, while 27 percent of organizations want to mobilize internal processes. According to TechNavio's recent forecast, the Enterprise Mobile Cloud Computing market in North America will grow at a CAGR of 18.12 percent over the period of 2011–2015.

Mobile cloud computing (MCC) is an emerging paradigm that encompasses mobile computing; cloud computing, networking and virtualization. MCC provides us new opportunities and challenges in energy saving by leveraging computing resources, network resources, and mobile resources to increase resource utilization, sharing, and virtualization to achieve cost reduction. Today, the fast increase of the number of mobile device users drives a strong demand on new mobile cloud infrastructures, services, and technologies. Current and future mobile cloud infrastructure and service vendors must address energy saving issues and challenges. Therefore, it is necessary for us to study, understand, and review the current state-of-the-art research results on energy saving in mobile cloud computing. Although numerous published papers discussed energy saving issues, most of them are addressed from one of three scopes: a) mobile computing, b) cloud computing, and c) network computing. This paper is written to have an intention to presents a snapshot of the existing research results on energy saving in the scope of mobile cloud computing. Moreover, the paper examines their limitations, discusses the future needs and issues.

This paper is organized as follows. Section 2 provides backgrounds on green computing and mobile cloud computing. Section 3 reviews the energy saving issues of mobile devices and discusses the existing results. Section 4 presents our review on energy-saving issues and solutions in wireless networking and communications. Section 5 covers detailed energy saving strategies, issues, and solutions on cloud infrastructures and servers. Section 6 discusses the major issues and needs. Finally, the conclusion remarks are given in Section 7.

2. BACKGROUND

2.1. Mobile Cloud Computing

According to [1], Mobile Cloud Computing (MCC) is "a rich mobile computing technology that leverages unified elastic resources of varied cloud and network technologies toward unrestricted functionality, storage, and mobility. It serves a multitude of mobile devices anywhere and anytime through the channel of Ethernet or Internet regardless of heterogeneous environments and platforms based on the pay-as-you-use principle." Experts believe MCC is an amalgam of three foundations, namely cloud computing, mobile computing, and networking [2]. The recent fast increase of mobile users is one of the strong driving forces for cloud computing service delivered to mobile users over the wireless Internet [3, 4, 5].

Here is the list of major motivations for mobile cloud computing.

- To address the never ending need of increasing the processing power and battery life time of mobile devices;
- To cope with the increasing need on diverse application services for mobile users with low-end devices;
- To maximize resource sharing and usage of existing cloud infrastructures and resources by providing efficient mobile connectivity to mobile users;
- To eliminate existing resource shortage issues and limitations of current mobile devices by connecting to unlimited cloud application services; and
- To provide virtualizations of diverse resources in MCC, including mobile devices, cloud computing servers and infrastructures, and wireless networks and Internet.

2.2. Energy Saving in Mobile Clouds

K. Murugesan [6] defines green computing as the "the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems — such as monitors, printers, storage devices, and networking and communications systems — efficiently and effectively with minimal or no impact on the environment". One of the primary objectives is to reduce energy usage and increase energy efficiency in computing devices, applications as well as in their executions.

In general, energy saving refers to efforts made to reduce energy consumption, that can be achieved through increasing energy efficiency and decreasing energy consumption and/or reducing consumption from conventional energy sources. Intuitively, its major benefits include cost reduction, energy conservation, and environment protection. Because mobile clouds include computing server infrastructures, networks, and mobile devices, researchers must pay a special attention to energy

978-0-7695-4945-3/13 $26.00 © 2013 IEEE
DOI 10.1109/IC2E.2013.37

conservation in MCC considering underlying large-scale computing and network infrastructures and the significant body size of mobile device users.

In mobile cloud computing, energy conservation issues can be classified in the following aspects (see Figure 1).

- Issues in network infrastructures and communications;
- Issues in cloud computing infrastructures and servers;
- Issues in mobile cloud service applications; and
- Issues in mobile devices and computing.

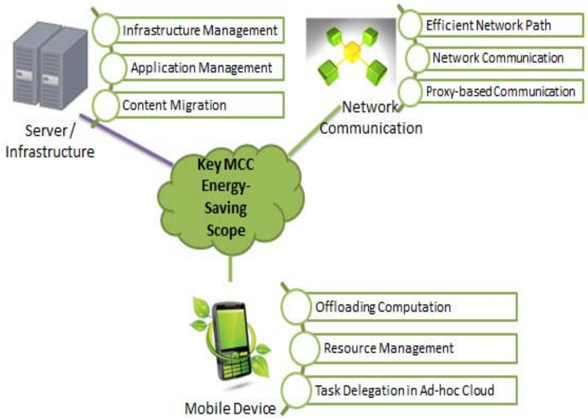

Figure 1. Energy-Saving Research Scope in MCC

3. Energy Saving Solutions for Mobile Devices

3.1. Computation Offloading to the Cloud

The idea is to offload computation intensive tasks from mobile devices to cloud servers in a mobile cloud to conserve energy on mobile devices and improve mobile user experience. This approach has two challenges listed below:

- What is the optimum condition when offloading computation from mobile clients to the cloud?
- What are the factors that need to be addressed while offloading computation to the cloud?

K. Kumar [7] addresses these issues based on the analysis of the energy consumption for a computation task on mobile devices versus on the server in a cloud. They proposed the following formula (1) to associate an offloading computation from mobile devices to the cloud. With this formula, they derive the amount of energy saved during offloading:

$$P_c \times \frac{C}{M} - P_i \times \frac{C}{S} - P_{tr} \times \frac{D}{B} \quad (1)$$

Where, C is the number of computing instructions to be offloaded,
M is the speed of mobile device (instructions/second),
S is the speed of the cloud server (instructions/second),
P_c is the mobile device power consumption (watts),
P_i is the mobile device idle power consumption (watts),
P_{tr} is the mobile device transmission power consumption (watts),
D is the bytes of data to be exchanged and
B is the network bandwidth.

With P_c, P_i, P_{tr} being constant, if the above formula produces a positive number, offloading reduces energy consumption, otherwise not. In addition, offloading can save more energy if the required data are stored on the cloud because there is no need to transfer data from mobile devices to the cloud over the wireless network. However, if the data needs to be transferred over the wireless networks with security requirements i.e., an encryption processing, then the overall energy consumption, when server is F times faster, can be expressed by formula (2) below:

$$\frac{C}{M} \times \left(P_c - \frac{P_i}{F}\right) - P_{tr} \times \frac{D}{B} - P_c \times \frac{C_p}{M} \quad (2)$$

Where, C_p is the additional number of instructions needed for the added computation for security (i.e. encryption). This extra energy consumption can be avoided when C_p become zero if encryption operation is performed on the mobile device. Clearly, computation offloading leads to other issues:

- When offloading involves data transfer over networks (either wireless networks or wireless Internet) to a cloud server, how can one minimize the power consumption by wireless communications and servers? The result given by C. Pluntke [8] has addressed this issue in section 4.1.
- Offloading requires mobile devices to scan for available wireless interfaces whenever data transfer is needed. This suggests that too many offloads also increase power consumption by mobile devices to discover a network. Hence, it becomes a concern to manage the number of offloads.

3.2. Mobile Device Resource Management

Current mobile platforms and operating systems support the executions of mobile client applications without paying much attention to efficient mobile resource management. Typical examples include applications running in the background, scanning for Wi-Fi network and synchronizing email account during night when mobile users are sleeping. Many ad-hoc techniques are available to minimize draining of power by these applications. Yet, they are insufficient to save enough energy on mobile devices.

Vallina-Rodriguez [9] proposed ErdOS, an extension of Android OS that extends the battery life of mobile devices by integrating two approaches below:

- An user-centered proactive resources management system that forecasts the demands for resources based on periodically monitored users' context, habits, and actions, such as stationary at home or work, commuting toward workplace, and charging the phone.
- Opportunistic accesses to resources where the mobile device can access resources from nearby mobile devices through local wireless networks and social networks. The authors suggested Bluetooth as the most suitable wireless interface because of its prevalence and low energy consumption.

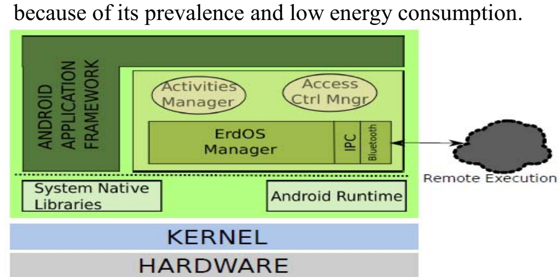

Figure 2. Architecture of ErdOS as an Android OS extension [9]

ErdOS, shown in Figure 2, includes three major components: *Activity Manager*, *Access Control Manager*, and *ErdOS Manager*. *Activity Manager* is responsible for monitoring and profiling user activities, run-time resource usage of Android and its application framework. Meanwhile, it also predicts resource demands using machine learning techniques. ***Access Control Manager*** contains all the necessary information such as - user profiles, public keys, policies, and required IPC mechanisms. ***ErdOS manager*** is the core component that is responsible for managing local resources, discovery of nearby devices, and deciding access patterns based on forecasting algorithms.

The authors also noted that the decision making process may lead to an increased computation on mobile devices.

3.3. Task Delegation in Ad-Hoc Mobile Cloud

G. Huerta-Canepa [10] proposed an ad-hoc mobile cloud framework where a small portion of the task is executed in the mobile device locally and the remaining portion is delegated to the nearby mobile devices that are already running the same task. The architecture of the ad-hoc mobile cloud in Figure 3 consists of five functional components: ***application manager***, ***resource manager***, ***context manager***, ***P2P component*** and ***offloading manager***.

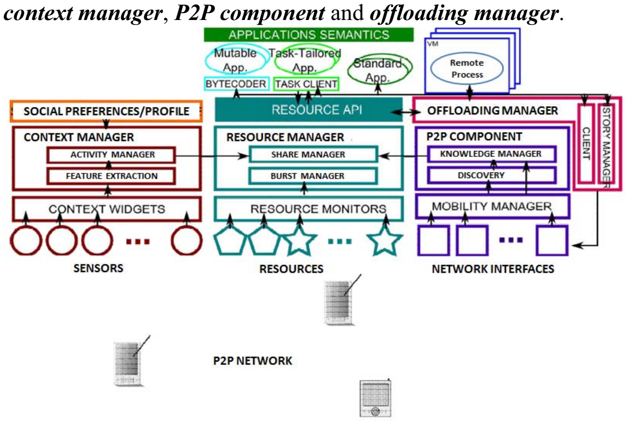

Figure 3. The Architecture of Ad-hoc Mobile Cloud [10]

The application manager is responsible for initiating and intercepting an application at loading time and changing the application to include features needed for offloading such as proxy creating and RPC support. The main function of the ***resource manager*** is to monitor the resources on a local device. For each of the application, a profile is created to keep track of the number of remote devices required to develop a virtual cloud. This profile is periodically monitored by the ***application manager*** to ascertain whether an instance of the virtual provider needs to be created or not. The ***context manager*** helps to synchronize the contextual information from the context widgets, and makes it accessible by other processes. The ***context manager*** is composed of three sub-components: a) context widgets that handle communication with the sources of information; b) context manager that extracts new contexts from the information; and c) social manager the keeps track of various types of relationships among users. The ***P2P*** component sends notifications to the ***context manager*** if a new device is included in the environment or if an existing device gets away of from the environment. The ***offloading manager*** sends out and manages tasks from the current mobile device to the neighboring mobile devices, accepts and processes tasks from the remote devices.

While the result indicates this framework saves energy, it needs to address decision making to identify stable nodes, adopting access mechanisms from neighbors, and dealing with energy consumed by additional computation to make decisions.

4. Energy Saving in Wireless Communications

4.1. Energy-Saving by Efficient Wireless Interface Discovery

Today high-end mobile devices (such as smartphones) are equipped with multiple air interfaces like Wi-Fi, 3G, and 4G. Although these different interfaces provide flexible communication and energy-saving options, existing communication protocols like TCP and UDP cannot take advantage of these options. For example, a communication session established with a 3G interface will try to maintain with the same interface until the session is finished or the interface is unavailable, even when the device is close to other better interfaces in terms of communication speed and energy-saving. To support multiple network paths concurrently and to shift an active session to the most energy efficient air interface, C. Pluntke [8] proposed two multipath scheduling algorithms using MPTCP. They are known as ***MDP scheduler*** and ***Omniscient ORACLE*** based on ***Markov decision processes*** (MDPs). The scheduler architecture of MPTCP consists of the following components.

a) MPTCP enabled multipath transfers;
b) MPTCP-capable proxies that assist the mobile devices to connect to standard TCP servers over the Internet; and
c) MPTCP scheduler (in mobile devices) that controls the switching among different air interfaces.

Using *MDP Scheduler*, scheduling is performed dynamically based on incomplete information i.e. partial information in every Δt time interval using a ***Markov decision process***, expressed as:

$$MDP = (S, A, P, C)$$

Where S denotes a set of states of different interfaces, A denotes a set of actions for selecting next interface after Δt time interval, P is the transition probabilities derived from the application and energy model while switching one state to the next, and C denotes a cost function for making a transition between states under a specific action. Two finite state machines (FSMs), shown in Figure 4, are used to derive the energy model for 3G and Wi-Fi respectively. Here, DCH state denotes the dedicated channel mode, and FACH state denotes the forward access channel mode.

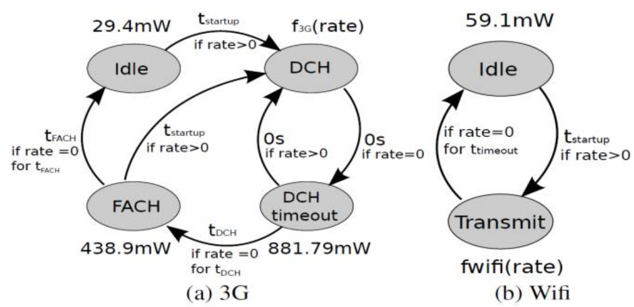

Figure 4. Sample FSMs for 3G and Wi-Fi Interfaces [8]

Using *Omniscient ORACLE,* scheduling is performed offline based on complete information. It serves as a lower bound as well as the guidance for a comparison, as in (3):

$$E_{ORACLE} \leq E_{MDP} \leq min(E_{3G}, E_{WIFI}) \quad (3)$$

Where, E_{ORACLE} and E_{MDP} denote energy consumption for offline and online scheduling in MPTCP respectively, and E_{3G} and E_{WIFI} denote energy consumed by traditional 3G and Wi-Fi interfaces respectively. The experiment results indicate that MDP scheduler and ORACLE scheduler are the most energy efficient among them. For example, the multipath scheduler is about 9.7% overall energy efficient compared to the traditional energy models. However, scheduler derivation requires intensive computation. The authors in [7] also pointed out this issue and suggested offloading computation to cloud servers.

4.2. Energy-Saving through Wireless Communication

Smartphones scan for nearby Wi-Fi interfaces very frequently that drains a significant amount of device energy. To minimize the scan for Wi-Fi networks, F. Xia [11] proposed assisted GPS (A-GPS) scheme where a smartphone scans for Wi-Fi networks periodically, stores networks information (such as SSID/BSSID, timestamp, location, and range) into a log and uses it to provide mobile users with the option of switching to the nearest Wi-Fi interface. They suggested avoiding the usage of power hungry GPSs and performing assisted or collaborative discovery of Wi-Fi networks using peer-to-peer, global and cloud sharing.

The proposed system, as in Figure 5(a), depicts the key components and the workflow. The cellular data rates will be periodically monitored at a chosen interval on the mobile device. The gathered data of cellular data rates are useful to decide whether to switch to a nearby Wi-Fi network. Whenever a switch decision is made, it triggers the switching model to find the nearest available Wi-Fi network access point for network connections and communication. Figure 5(b) shows the flow chart of the proposed switching module, which comprises of four steps.

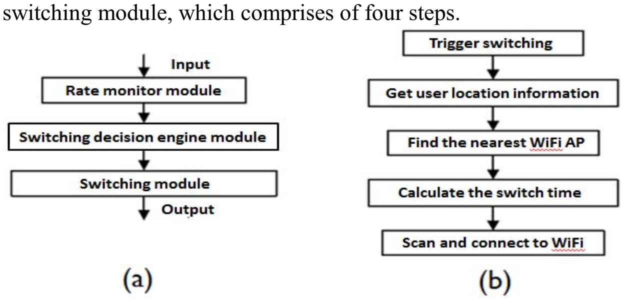

Figure 5 (a) Block diagram and (b) Switching module flow chart of A-GPS scheme [11]

The authors in [11] also evaluated the A-GPS scheme by conducting experiments using other schemes such as - non-switching schemes like GSM assisted network switching, scanning assisted network switching, GPRS non-switching, and Wi-Fi non-switching. The experiments are conducted and reported based on different data rate thresholds in different usage categories, such as text messaging, web browsing, video streaming, and file download. The experimental results showed that A-GPS switching scheme is overall more energy efficient compared to all other schemes. Because switching decisions are made only based on data rates, the performance results may not be consistent when considering other factors, such as network conditions, and selected network quality.

4.3. Energy-Saving using Proxy-based Architecture

The battery life of mobile devices such as smartphones, tablets is heavily impacted by data transfers over the networks for various applications like web surfing, file transfers, streaming multimedia and maps. F. R. Dogar in [12] introduces an energy saving strategy called Catnap that saves device energy when data transfer takes place in a hybrid network that uses combination of high bandwidth access link such as 802.11 home networks and slow link such as DSL network. The data delivery rate between slow and fast link results in delay between the packets. The slow link slows down the overall delivery time. But the mobile devices need to stay awake because TCP needs the both ends active during the entire transfer time. Catnap system, shown in Figure 6, accumulates the small intervals between packets to form a larger interval when the device is sent to sleep mode to save power without little or no change in packets' actual delivery time. It has three major components:

- **Catnap Proxy** acts as a translator and decouples wired and wireless segments.
- **Workload Hints** provides information about the content type of a transfer. Usually hints are of two categories: i) content type (e.g. web, ftp, streaming) when sleep mode applicable; and ii) content type (e.g. VoIP) when sleep mode is not applicable.
- **Scheduler** is the brain of Catnap system. It makes decisions on possible sleep time based on transfer size, cost of sleep mode, wired segment bandwidth and wireless segment bandwidth. The decisions are of two types: i) 802.11 power save mode (PSM) when device switches between Wi-Fi radio on and off; ii) deep sleep mode (S3) when device is sent to standby or suspended to RAM mode.

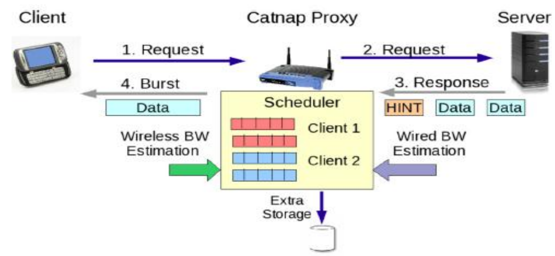

Figure 6. Catnap system overview [12]

The experiments based on Catnap system showed up to 2 – 5x improvements of battery life for Nokia N810 and Thinkpad T60. The 802.11 PSM modes showed up to 2 x improvements in battery life while S3 mode showed up to 5 x improvements. However, S3 mode is applicable only when user does nothing during these lager data transfers. But current smartphone devices provide the capabilities of doing multiple tasks at the same time which limits the application of S3 mode. In addition, Catnap system may cause a little delay in data delivery downgrading user experience which is very crucial for mobile devices. Proxy-based architecture was also proposed by some other prior research [13, 14, 15] to save mobile device energy related to wireless network communication. Their experiment also showed significant energy saving but with the expense of delay in data transfer or some other consequences.

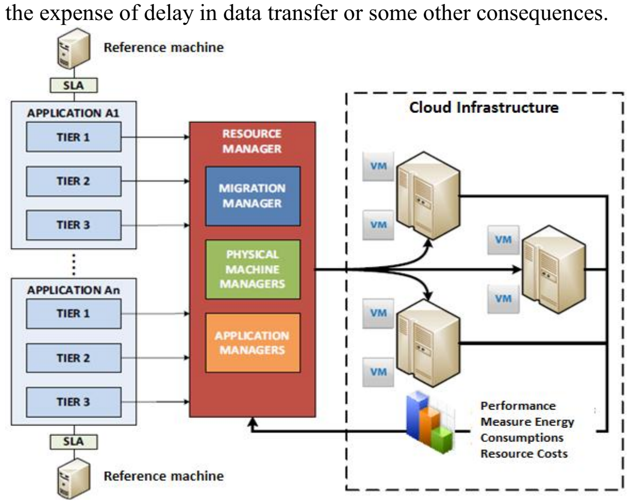

Figure 7. Architecture of the resource management framework [17]

5. ENERGY SAVING IN CLOUD INFRASTRUCTURES

5.1. Cloud Server Infrastructure Management

Cloud servers in data center infrastructures accounted for 1.5% of total U.S. energy consumption in 2006 with an annual growth 18% [16]. A large chunk of consumed energy is wasted by idle servers due to inefficient management of these resources. M. Guazzone in [17] proposed a dynamic resource management framework, as in Fig. 7, that requires no prior knowledge of the system, and maximizes the quality of service to maintain service-level agreement (SLA) while minimizing energy consumption by physical resources.

According to this scheme, multi-tiered user applications are deployed on the cloud infrastructure where each tier of the application is deployed onto a separate VM on available physical machines. The core of the framework is the resource manager, and it contains ***migration manager***, ***physical machine managers***, and ***application managers***. The main objective of this ***resource manager*** is to continuously monitor the performance and efficiency of each deployed application instance on VMs so that it can react accordingly to achieve the targeted performance goals while minimizing energy conservation. A ***migration manager*** is responsible to make migration decisions by closely monitoring the energy consumption status of involved VMs. Its primary goal is to achieve higher energy efficiency while increasing computer resource sharing and utilization. An ***application manager*** provides the application with the required amount of resource capacity for satisfying the service level objectives. The ***physical machine manager*** has to meet CPU demands received from ***application managers***. The proportional policy is used to calculate the CPU share demand,

$$\hat{d}_i = \frac{d_i}{\sum_{j=1}^{n} d_j} D$$

Where, d_i denotes the incoming share demand and D denotes the capacity of a physical machine to host n VMs and $0 < D \leq 1$.

The authors experimentally compared their dynamic approach with other two approaches listed below:

- *Static Service-Level Objective (SLO)* approach in which service satisfaction has higher priority than minimization of energy consumption.
- *Static Energy* approach in which energy consumption reduction is the highest priority.

In their case study, they used different workload types, such as daily-cycles of activity, self-similar activity and temporal burst. The results showed slightly better performance than other approaches in SLA satisfaction and energy consumption. However, the authors also mentioned that they are yet to implement the *migration manager* that may add additional energy overhead. To minimize the number of migration for energy conservation, B. Li in [18] proposed a resource provision scheme called over-provision approach where resource allocation is adjusted dynamically based on over-provision ration, α (0 ≤ α ≤ 1). The initial resource allocation denoted by *size'(x) = (1 + α) size(x),* where *size(x)* denotes preset resource demand of *x*. Moreover, they also provide an algorithm to support resource provisions.

5.2. Server Application Management for Energy Saving

Lack of efficient placement and migration of dynamic application workload may result in energy inefficient allocations of server machines. Figure 8(a) shows one example. As shown in Figure 8(b*)*, rearranging workloads appropriately at the time of arrival and departure can reduce the number of running server nodes [18].

B. Li in [18] presented an approach (called EnaCloud) for energy conservation in data centers. EnaCloud uses an energy-aware heuristic algorithm to generate the application placement and scheduling schemes in response to the workload (or task) arrival, departure, and resizing events. The heuristic algorithm addresses:

- *Workload arrival event*: A newly arrived workload recursively replaces the existing smaller workloads.
- *Workload departure event*: The released resource by a departure workload is rearranged to reduce the number of running nodes (servers).
- *Workload resizing event*: The workload-resizing event is equivalent to the combination of arrival and departure events.

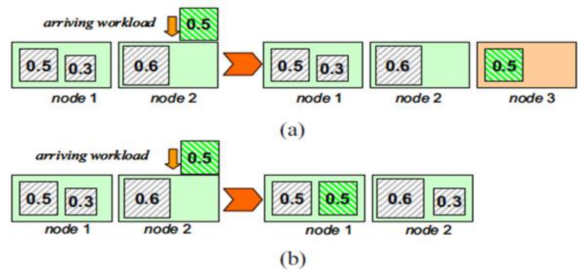

Figure 8. (a) 3 nodes are needed without migration
(b) Only 2 nodes are needed with migration [18]

As shown in Figure 9, the EnaCloud framework consists of several functional components. They are: ***global controller***, ***VM controller***, ***resource provision manager,*** and ***performance monitor***. **Global controller** consists of concentration manager, job scheduler, and they are deployed globally while the other components are installed in each node. ***Concentration manager*** generates application placement schemes (that dispatched by the job scheduler) to the VM controller to start, stop or migrate workload actions. The ***resource provision manager*** optimizes this approach further to avoid the frequent application migration.

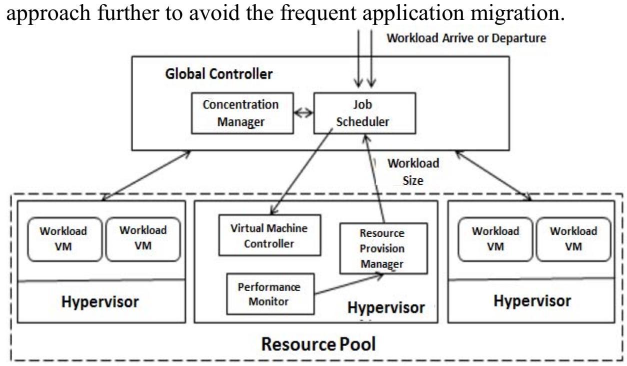

Figure 9. EnaCloud Architecture [18]

The authors compared their approach to *First Fit* and *Best Fit* algorithms in terms of number of active nodes, energy consumption, and pool utilization with different workload types. Their approach reduced energy consumption about 10% with 10 fewer nodes compared to *First Fit* algorithm, and 13% with 6 fewer node compared to *Best Fit* algorithm. It also showed superior pool utilization compared to two other approaches in maintaining 90% utilization on average. However, factors like CPU and memory should be considered for real life application.

5.3. Content Migration for Energy Saving

The content migration, from the main cloud computing (CC) data center to local cloud data centers or MCC data centers, as in Figure 10, is another approach to reduce energy consumption. This technique is useful to cope with many content-based applications, such as distance learning for academic lectures, conferences, museums. M. Altamimi in [19] had systematically compared the energy consumption in a cloud infrastructure for content migration in two different scenarios:

- Download the data from CC data center that uses smartphone network, Internet network and CC data center.
- Download the data from MCC data center that uses smartphone network, MCC data center, Internet network and CC data center.

A *smartphone network* is comprised of either a Wi-Fi based WLAN access network or a 3G Cellular/WiMAX based broadband access network. An *Internet network* is comprised of broadband gateways, data center gateways, Ethernet switches, provider edge routers, core routers, and wavelength division multiplexed (WDM) fiber links. A *CC data center* is comprised of content servers, and hard-disk array. An *MCC data center* is comprised of all the elements of a CC data center but to a smaller extent. Energy consumed by MCC data center will be small if MCC data center is carefully designed according to the needs.

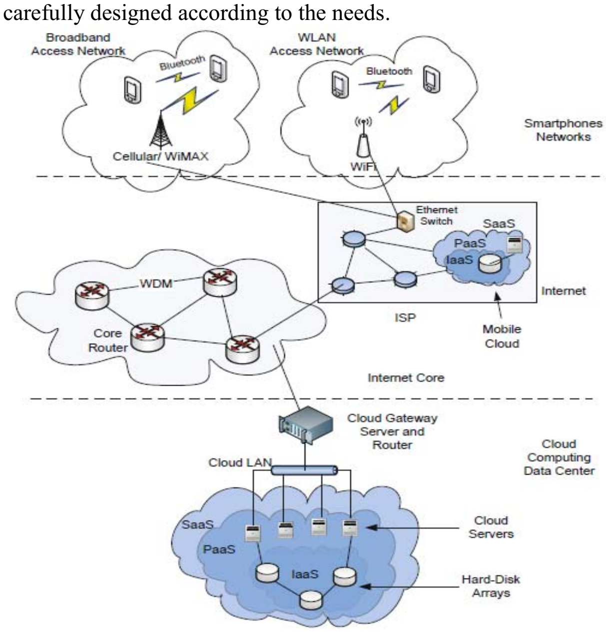

Figure 10. Temporary Content Migration System Model [19]

The experiment results in [19] support this. Using this approach, the energy consumption can be reduced by 63% to 70%. However, the authors found that MCC data center becomes inefficient when the ratio of power consumption of the MCC to the CC is greater than or equal to 8. While the MCC data center based approach can play a huge role in green cloud technology by saving energy when content is accessed locally, the adoption of this approach may not be feasible if the cost becomes higher for purchasing and maintaining same sets of equipment for many MCC data centers. In addition, more research is needed to find the effectiveness of this solution in other cloud services like software-as-a-service (SaaS), and platform-as-a-service (PaaS).

6. Energy Saving Issues and Needs in Mobile Clouds

6.1. Open Energy Saving Issues in Mobile Cloud Computing

Here, we identify and summarize several important research issues and challenges below.

- *Issue #1: How to engineer and deliver mobile cloud services with minimum mobile device energy consumption without affecting mobile user experience?*

Similar to today's mobile application users, future mobile device users expect highly efficient application services in mobile clouds on wireless Internet. Although today's market still shows an increasing need on high-end mobile devices such as smartphones, mobile cloud computing will still be highly dependent on low-end mobile devices because of its low cost and energy saving. This brings out an important issue in engineering and delivering mobile cloud services to users with rich mobile experience using low-end devices.

As saving device energy and improving user experience conflict with each other, reducing energy consumption without affecting user experience is a challenge.

- *Issue #2: How to reduce cloud data center energy consumption without violating service-level agreements with the cloud clients?*

To save energy consumed by cloud servers, the server administrators usually turn off under used servers during the period of low demand. This poses a risk of affecting customer needs during peak time if enough servers are available to serve increased demand. Hence, saving energy without violating service agreement is a challenge.

- *Issue #3: How to design a computation offloading mechanism that will learn computation needs and make decisions to reduce energy consumption?*

Due to the diverse nature of smartphone users and their usage style, making a prediction on usage pattern becomes difficult. Because tracking all the activities to make decisions for smartphone devices may require significant computation in the smartphone device. This may further increase energy consumption. Thus, designing an energy-aware computation offloading scheme is another challenge considering these constraints.

6.2. Emergent Needs on Energy Saving in Mobile Clouds

Clearly, one needs new mobile devices, computer servers and storage serves that address green computing issues. Meanwhile, there are a number of important needs to fulfill in mobile cloud computing to achieve energy saving. They are listed below.

- *Need #1: Innovative energy-aware mobile cloud infrastructures*

Often sparing management of resources is responsible for excessive energy consumption. To overcome this, one needs an improved resource management system for both cloud servers and mobile devices.

- *Need #2: Intelligent offloading management algorithms*

The availability of desktop based rich applications to smartphones in the manner of Web-based cloud services leads to increased computation and resource need in smartphones. This in turn increases energy consumption in smartphone devices. To overcome limitations of smartphone devices, an intelligent offloading management algorithm is highly desirable. This algorithm needs to be smart enough to learn all the computation need in smartphones and make energy optimal decision based on its gained knowledge. This algorithm will initiate computation offloading to the cloud server from mobile devices only when necessary. This algorithm should be smart enough to consider all the possible user contexts and circumstances.

- *Need #3: Energy saving communications and standard solutions*

Widely used wireless communication protocol UDP is not known for energy efficiency. Thus, new wireless communication protocol, that will consider energy consumption as one of its parameters during wireless communication, is a huge need.

- *Need #4: New architecture addressing energy issues*

The mobile cloud computing is a new concept that needs further research including its architecture. Thus a new architecture that addresses energy issues for mobile cloud computing that comprises of cloud computing, mobile computing and wireless networks.

- *Need #5: Intelligent and dynamic application migration schemes for data center servers*

The arrival of workload to cloud data center server is unpredictable. Thus to save energy consumed by running servers, one needs an intelligent and dynamic application migration scheme that will learn workload arrival nature and migrate applications dynamically to an energy efficient node intuitively.

- *Need #6: Mechanisms for discovering energy efficient wireless network profiles*

As smartphones devices are equipped with multiple wireless network profiles, selection of energy efficient network profile is very crucial in saving smartphone energy. To achieve this, one needs a mechanism to discover energy efficient wireless network profiles intelligently.

6. CONCLUSIONS

Mobile cloud computing is becoming an emerging research subject due to its strong business market growth. Green computing for mobile clouds and mobile computing services has become a hot research topic because mobile cloud vendors are looking for efficient energy solutions to achieve cost reduction. This paper discusses and reviews existing energy saving strategies and solutions in three perspectives: a) mobile devices, b) network infrastructures and communications, and c) cloud infrastructures and computing software. In addition, the paper also examines the limitations and future needs in energy saving for mobile cloud computing. Clearly, current research results and technologies only focus on specific issues from one perspective of mobile cloud computing. Therefore, there is a lack of a comprehensive energy saving infrastructure with well-defined frameworks and solutions to address diverse green computing issues and needs in mobile clouds. Our current and future research focus is developing an innovative energy-saving infrastructure based on real-time analysis and intelligent decision-making.

REFERENCES

[1] Z. Sanaei, S. Abolfazli, A. Gani, and M. Shiraz, “Service-Based Arbitrated Multi-Tier Infrastructure for Mobile Cloud Computing,” IEEE Workshop on Mobile Cloud Computing, Beijing, China, 2012.

[2] Z. Sanaei, *et al.*, “Tripod of requirements in horizontal heterogeneous mobile cloud computing,” *Proc. of 1st Int. Conf. on Computing, Information Systems, and Communications*, 2012.

[3] W. Song and X. Su, “Review of Mobile cloud Computing,” in *Proc. of IEEE 3rd Int. Communication Software and Networks (ICCSN) Conf.*, 2011, pp. 1–4.

[4] S. Chetan, *et al.*, (2010). *Cloud Computing for Mobile World* [Online]. Available: chetan.ueuo.com/projects/CCMW.pdf.

[5] A. Berl, *et al.*, "Energy-Efficient Cloud Computing," The Computer Journal, vol. 53 (7), pp. 1045–1051, 2009.

[6] S. Murugesan, “Harnessing Green IT: Principles and Practices,” IEEE *IT Professional*, Jan–Feb 2008, pp. 24–33.

[7] K. Kumar and Y.-H. Lu, “Cloud computing for mobile users: Can offloading computation save energy?” *IEEE Computer*, vol. 43, pp. 51–56, 2010.

[8] C. Pluntke, L. Eggert, and N. Kiukkonen, “Saving Mobile Device Energy with Multipath TCP”. In *ACM Workshop on MobiArch*, 2011, pp. 1–6.

[9] N. Vallina-Rodriguez and J. Crowcroft, “ErdOS: achieving energy savings in mobile OS,” in *Proceedings of the sixth international workshop on MobiArch*, ser. MobiArch ’11. New York, NY, USA: ACM, 2011, pp. 37–42.

[10] G. Huerta-Canepa and D. Lee, “A virtual cloud computing provider for mobile devices,” in *1st ACM Workshop on Mobile Cloud Computing and Services: Social Networks and Beyond(MCS10)*, San Francisco, California, 2010, pp. 1–5.

[11] F. Xia, W. Zhang, F. Ding, and R. Hao, “A-GPS Assisted Wi-Fi Access Point Discovery on Mobile Devices for Energy Saving,” in *Proc. of the IEEE GIIS 2011*, Da Nang, Vietnam, 2011.

[12] F. R. Dogar, P. Steenkiste, and K. Papagiannaki, “Catnap: Exploiting high bandwidth wireless interfaces to save energy for mobile devices,” in *Proc. ACM MobiSys’10*, San Francisco, California, 2010, pp. 107–122.

[13] G. Anastasi, *et al.*, “An energy-efficient protocol for multimedia streaming in a mobile environment.” *Journal of Pervasive Computing and Communications*, vol. 1, no. 4, pp. 301–312, Dec 2005.

[14] T. Armstrong, O. Trescases, C. Amza, and E. Lara, “Efficient and transparent dynamic content updates for mobile clients”, in *ACM MobiSys: Proceedings of the 4th Int. Conf. on Mobile Systems, Applications and Services*, New York, NY, 2006, pp. 56–58.

[15] M. Gundlach, *et al.*, “Dynamic, power-aware scheduling for mobile clients using a transparent proxy,” in *Proc. of the Int. Conf. on Parallel Processing, 2004,* vol. 1, pp. 557–565, 2004.

[16] Report to Congress on Server and Data Center Energy Efficiency, U.S. Environmental Protection Agency ENERGY STAR Program, July 15, 2012.

[17] M. Guazzone, C. Anglano, and M. Canonico, “Energy-Efficient Resource Management for Cloud Computing Infrastructures,” presented at the Cloud Computing Technology and Science (CloudCom), IEEE 3rd Int. Conf., 2011, pp. 424–431.

[18] B. Li, *et al.*, “Enacloud: An energy-saving application live placement approach for cloud computing environments,” in *IEEE Int. Conf. on Cloud Computing*. IEEE Computer Society, Bangalore, India, 2009, pp. 17–24.

[19] M. Altamimi and K. Naik, “The Concept of a Mobile Cloud Computing to Reduce Energy Cost of Smartphones and ICT Systems,” in *Proc. of the 1st Intl. Conf. on Information and Comm. on Tech. for the fight against Global Warming*’11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 79–86.