Abstract -- We are particularly aware that Green Cloud Computing (GCC) is a broad range and a hot field. The distinction between “consumer of” and “provider of” cloud-based energy resources may very important in creating a world-wide ecosystem of GCC. A user simply submits its service request to the cloud service provider with the connection of Internet or wired/wireless networks. The result of the requested service is delivered back to the user in time, while the information storage and process, interoperating protocols, service composition, communications, and distributed computing are all smoothly interactive by the networks. In this paper, this is a survey on green cloud computing schemes based on networks. We first introduce the concept and history of Green computing, and then focus on the challenge and requirement of Cloud computing. Cloud computing needs to become green, which means provisioning cloud service while considering energy consumption under a set of energy consumption criterions and it is called GCC. Furthermore, we introduce recent work done in GCC based on networks, including microprocessors, task scheduling algorithms, virtualization technology, cooling systems, networks and disk storage. After that, we present the works on GCC from our research group in Georgia State University. Finally, we give the conclusion and some future works.

I. INTRODUCTION

Here we are particularly aware that Green Cloud Computing (GCC) is a broad range and a hot field. The distinction between “consumer of” and “provider of” cloud-based energy resources may very important in creating a world-wide ecosystem of GCC. A user simply submits its service request to the cloud service provider with the connection of Internet or wired/wireless networks. The result of the requested service is delivered back to the user in time, while the information storage and process, interoperating protocols, service composition, communications, and distributed computing are all smoothly interactive by the networks. In this section, we first introduce the concept and history of Green computing, and then focus on the challenge and requirement of Cloud computing. Cloud computing needs to become green, which means provisioning cloud service while considering energy consumption under a set of energy consumption criterions and it is called GCC.

II. GREEN COMPUTING: CONCEPT AND

HISTORY

Green computing or green IT, refers to environmentally sustainable computing or IT. San Murugesan defines the field of green computing as "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 [1]."

Modern IT systems are complicated because all of them rely on so many factors such as applications or software, people, networks and hardware. A solution may also need to address end user satisfaction, management restructuring, regulatory compliance, and return on investment (ROI). To reduce the use of hazardous materials, maximize energy efficiency during the product's lifetime, and promote the recyclability or biodegradability of defunct products and factory waste are the main goals of Green computing, which can be attained by making the use of computers as energy-efficient as possible and designing algorithms and systems for efficiency-related computer technologies.

In 1992, the U.S. Environmental Protection Agency (EPA) launched a voluntary labeling program named Energy Star, which is designed to promote energy-efficiency technologies in monitors, climate control equipment and so on. The term "green computing" was probably coined shortly after the Energy Star program began. Concurrently, the Swedish organization TCO Development launched the TCO Certification program. At first, this program was to promote low magnetic and electrical emissions from CRT-based computer displays, and was later expanded to include criteria on energy consumption, ergonomics, and the use of hazardous materials in construction.

III. CLOUD COMPUTING: CHALLENGES AND REQUIREMENTS

“If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. The computer utility could become the basis of a new and important industry.” John McCarthy, who received the Turing Award in 1971 for his major contributions to
the field of Artificial Intelligence (AI), said at the MIT Centennial in 1961. This vision of computing utilities based on a service provisioning model that computing services will be readily available on demand, like other utility services (e.g. electricity) available in today’s society. Users pay only when they access the computing services, and no longer need to build and maintain complex IT infrastructure.

Fig. 1 show cloud computing based on networks is as a service. This a model is called utility computing, or recently as Cloud computing or Clouds [3]. Cloud computing delivers platform, and software as services under the pay-as-you-go model, which provide both physical and virtualized cloud resources and are referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) respectively. The cloud computing has three kinds of conditions: Public clouds, Hybrid clouds, and Private clouds, which are connected by networks. In Clouds, businesses and users access services based on their requirements from anywhere in the world without regard to where the services are hosted.

Many computing service providers, including Microsoft, Yahoo, Google and IBM are rapidly deploying data centers in various locations around the world to deliver Cloud computing services. These data centers host a variety of applications on shared hardware platforms. And the applications include serving requests of web applications that only run for a few seconds, large data set processing that run for longer periods of time, distributed databases that need real-time response and internet banking that requires security guarantees. The need to manage multiple applications in a data center creates the challenge of on-demand resource provisioning and allocation in response to time-varying workloads. In order to maintain isolation and provide performance guarantees, high performance becomes the sole concern in data center deployments, and data center resources are statically allocated to applications based on peak load without paying much attention to energy consumption. A large data center may require many megawatts of electricity, enough to power thousands of homes [4]. Organizations such as Google, Microsoft, Amazon, Yahoo!, and many other operators of large networked systems cannot ignore their energy costs. A back-of-the-envelope calculation for Google suggests it consumes more than $38M worth of electricity annually. A modest 3% reduction would therefore exceed a million dollars every year [27]. Cloud service providers are taking measures to ensure that their profit margin is not dramatically reduced due to high energy costs. For instance, Google, Microsoft, and Yahoo are building large data centers in barren desert land surrounding the Columbia River, USA to exploit cheap and reliable hydroelectric power [5]. As energy costs are increasing while availability dwindles, we cannot only focus on optimizing data center resource management for pure performance but also optimizing energy efficiency. Besides the expensive maintaining cost, data centers are unfriendly to the environment. Data centers are made of hundreds of thousands of servers, which lead to high energy costs and huge carbon footprint for powering and cooling. Data centers now drive more in carbon emissions than both Argentina and the Netherlands [4]. As we all know, carbon footprints result in a significant impact on climate change, it raises the concern from Governments worldwide to reduce carbon footprints. For example, the Japanese government has established the Japan Data Center Council to address the soaring energy consumption of data centers [6]. Leading computing service providers have also recently formed a global consortium known as The Green Grid [7] to promote energy efficiency for data centers and minimize their environmental impact.

The pervasive demands of users are growing so quickly that larger servers and disks, more powerful chips are needed to process them fast enough within the required time period. Cloud computing with increasingly interacting between front-end client devices and back-end data centers will cause an enormous escalation of energy usage. To lower the energy usage of data centers while meeting service provision guarantees is the big challenge and requirement that Cloud computing faces. This is essential to ensure that the future growth of Cloud computing is sustainable. The future data center resources need to be managed in an energy-efficient manner, that means Cloud resources need to be allocated not only to satisfy Quality of Service (QoS) requirements specified by users via Service Level Agreements (SLA), but
also to meet a criterion of energy usage.

IV. GREEN CLOUD COMPUTING
The future of Cloud computing must be green, here it is called Green Cloud computing. Green Cloud computing is envisioned to achieve not only high performance processing and utilization of computing infrastructure, but also has a set of criterions to limit energy consumption. The Green Grid is developing metrics to measure data center productivity as well as efficiency metrics for all major power-consuming subsystems in the datacenter. In 2007, the Green Grid proposed the use of Power Usage Effectiveness (PUE) and its reciprocal, Datacenter Efficiency (DCE) metrics, which enable datacenter operators to quickly estimate the energy efficiency of their datacenters, compare the results against other datacenters, and determine if any energy efficiency improvements need to be made [29]. And later in 2008, they redefine PUE as datacenter infrastructure efficiency (DCiE) [30]. Now PUE and DCiE have received broad adoption in the industry, including companies like AMD, APC, Dell, HP, IBM, Intel and Microsoft, to name a few.

It is reported that the most part of power consumption in datacenters comes from computation processing, disk storage, network and cooling systems. Nowadays, there are new technologies and methods proposed to reduce energy cost in datacenters. The following subsections will introduce recent work done in these fields.

V. COMPUTATION PROCESSING, MICROPROCESSORS
It is well-known that microprocessors play a key role in computation processing. With shrinking dimensions of the transistors in the microprocessors, leakage currents consume more power than the actual computational processes. In the late 2000s, new materials were introduced to reduce this burden so as to cut down the energy consumption of datacenters. Most notably, the replacement of the SiO2 gate oxide, which is only a few atomic layers thick, with a physically thicker layer of a hafnium-based oxide enabled an appreciable reduction of the gate tunneling currents while maintaining the electrical performance of the transistor [8-9].

For the newest members of microprocessor families, sophisticated circuit architectures have been introduced, which allow the power associated with computational processes and also the leakage power to be adapted [10-11]. These innovations, that the microprocessor frequency can be adjusted and circuit blocks can be temporarily turned down completely when not in use, lead to energy savings for a computational load that comes in bursts or that is bound to memory latency or I/O operations. G. Semeraro et al. [31] describe a Multiple Clock Domain (MCD) processor, in which the chip is divided into four (coarse-grained) clock domains, corresponding to the front end (including L1 instruction cache), integer units, floating point units, and load-store units (including L1 data cache and L2 cache), within which independent voltage and frequency scaling can be performed. Against traditional singly-clocked globally synchronous systems, the multiple Clock Domain (MCD) processor with DVFS can lower energy consumption obviously.

VI. COMPUTATION PROCESSING, TASK SCHEDULING
In order to reduce the power consumption of computation processing, much work has been done in task scheduling of data centers. Thermal aware resource management for data centers has recently attracted much research interest from high performance computing communities. The computational fluid dynamics (CFD) [12] models may be the most elaborate thermal aware schedule algorithms for tasks in data centers, they presents a detailed 3-dimensional Computational Fluid Dynamics based thermal modeling tool, called ThermoStat, for rack-mounted server systems. Because the CFD based model is too complex and is not suitable for online scheduling, some researchers develop several less complex online scheduling algorithms. For example, Q. Tang et al. [14] show through formalization that minimizing the peak inlet temperature allows for the lowest cooling power needs. Using a low-complexity, linear heat recirculation model, they define the problem of minimizing the peak inlet temperature within a data center through task assignment (MPIT-TA), consequently leading to minimal cooling requirement. J. D. Moore et al. [15] develop an alternate approach, which leverages the non-intuitive observation that the source of cooling inefficiencies can often be in locations spatially uncorrelated with its manifested consequences, providing additional energy savings. And there are other more examples like sensor-based fast thermal evaluation model [16-17], Generic Algorithm & Quadratic Programming [14, 18], and the Weatherman – an automated online predictive thermal mapping [19].
VII. COMPUTATION PROCESSING, VIRTUALIZATION

Virtualization of computer resources [20] is one of the key technologies in Cloud Computing. Traditionally, an organization purchases its own computing resources and deals with maintenance and upgrade of the outdated hardware, resulting in additional expenses. Virtualization technology allows one to create several Virtual Machines (VMs) on a physical server and, therefore, reduces amount of hardware in use and improves the utilization of resources. Organizations can outsource their computation needs to the Cloud, thereby eliminating the necessity to maintain own computing infrastructure.

A Virtual Power approach has been proposed by Nathuji and Schwan [21] for online power management to support the isolated and independent operation assumed by guest virtual machines (VMs) running on virtualized platforms and to make it possible to control and globally coordinate the effects of the diverse power management policies applied by these VMs to virtualized resources. The resource management is divided into local and global policies. On the local level, the system leverages guest operating system’s power management strategies. Consolidation of VMs is handled by global policies applying live migration to reallocate VMs.

Kusic et al. [22] implement and validate a dynamic resource provisioning framework for virtualized server environments wherein the provisioning problem is posed as one of sequential optimization under uncertainty and solved using a look-ahead control (LLC) scheme. The proposed model requires simulation-based learning for the application-specific adjustments.

Consolidation of applications in cloud computing environments presents a significant opportunity for energy optimization. After study the inter-relationships between energy consumption, resource utilization, and performance of consolidated workloads, Srikantiah et al. [23] have proposed a heuristic for multidimensional bin-packing problem as an algorithm to handle problem of requests scheduling for multi-tiered web-applications in virtualized heterogeneous systems in order to minimize energy consumption, while meeting performance requirements.

Song et al. [24] propose a multi-tiered resource scheduling scheme which automatically provides on-demand capacities to the hosted services via resources flowing among VMs. The resource allocation to applications accords to their priorities in multi-application virtualized cluster to optimize resource allocation among services in data center.

VIII. COOLING SYSTEM

Data center power consumption and cooling are two of the biggest energy issues that confront IT organizations today. Cooling systems consume nearly half of the electricity energy of data centers [25]. Traditionally, the cooling infrastructure of data centers is to remove heat by forced circulation of large amounts of chilled air. Now, it is reported that scientists are trying to use chilled-liquid cooling in high-end mainframes and densely packed servers to cope with the high heat fluxes. Microchannel heat sinks can be designed such that the thermal resistance between the transistor and the fluid is reduced to the extent that even cooling water temperatures of 60°C to 70°C ensure no overheating of the microprocessors [26]. Using this hot water cooling, chillers are no longer required year-round, that means the data-center energy consumption can be reduced by up to 50%. And more attractively, direct utilization of the collected thermal energy becomes feasible, either using synergies with district heating or specific industrial applications. With such an appealing waste-heat recovery system, the green diligence of data centers would be upped substantially [30].

IX. NETWORK

Over the past years, both energy cost and network electrical requirements show a continuous growth, besides a more widespread sensitivity to ecological issues, which arouse the interest on energy efficient networking. Traditionally, networks, links and devices are provisioned for peak load in order to meet high performance requirements, which typically exceed their average utilization by a wide margin, resulting in flat energy wastes. Nowadays, telecoms, network equipment manufacturers and the networking research community mainly regard the innovative criteria and technologies, which are able to dynamically adapt network capacities and resources to current traffic loads and requirements to save energy.

In order to develop today’s network equipments, Raffaele Bolla et al. [28] explore and try to evaluate the feasibility and the impact of power management policies that can well suit a heterogeneous set of highly modular architectures. The proposed policies aim at optimizing the power consumption of each device component with respect to its expected network performance.

The conventional approach to reduce energy costs has been to reduce the amount of energy consumed. Asfandyar Qureshi et al. [27] analyze a new method to reduce the energy costs of running large Internet-scale systems. They found that electricity prices exhibit both temporal and geographic variation, due to regional demand differences, transmission inefficiencies, and generation diversity. And
they characterize the variation due to fluctuating electricity prices and argue that existing distributed systems should be able to exploit this variation for significant economic gains.

X. CONCLUSION

This is a survey on GCC based on networks. We first introduce the recent research schemes on GCC: (I) Green computing concept and history; (II) Cloud computing challenge and requirement; (III) Recent research done in Green Cloud Computing, including: microprocessors, task scheduling algorithms, virtualization technology, cooling systems, networks and disk storage. Furthermore, we introduce recent work done in GCC based on networks, including microprocessors, task scheduling algorithms, virtualization technology, cooling systems, networks and disk storage. After that, we present the works on GCC from our research group. Finally, we give the conclusion and some future works.

Our future work will focus on the engineering implementation of the above control scheme in a multi-cloud computing environment that is being developed by several universities as part of the VCL open source community. We observe, finally, that our proposed model can be expanded to address additional parameters related to other aspects of multi-cloud computing, including, for instance, the level of computing security that user applications may require, or even storage availability or dependability. We also try to carry out our above FD schemes in IBM Cloud Computing platform, where there are some actual problems that should be solved before they are used in our living.

XI. REFERENCE


(Visited on Feb 26, 2010).


