**LITERATURE REVIEW**

**PAPER 1 --Elastic Tree: Saving Energy in Data Center Networks**

Data centers aim to provide reliable and scalable computing infrastructure for massive Internet services. To achieve these properties, they consume huge amounts of energy, and the resulting operational costs have been dominant over the capital cost (hardware cost). Most efforts have been focused on the reducing the energy consumed by the cooling system and the servers. Improvement basically includes better components as well as better software products. Since several data centers have deployed so the amount of energy consumption has by these data centers have been increasing rapidly. Networks are a shared resource connecting critical IT infrastructure, and the general practice is to always leave them on. Hence the energy can be saved by improving a network’s ability to scale up and down as the requirements of the traffic demand. Any dynamic energy management system that attempts to achieve energy proportionality by powering off a subset of idle components must demonstrate that the active components can still meet the current offered load, as well as changing load in the immediate future. The main aim is that the power savings must be worthwhile, performance effects must be minimal, and fault tolerance must not be sacrificed. In this paper, they present an Elastic Tree- a network wide power manager which dynamically adjusts the set of active network elements such as links and switches to satisfy changing data center traffic loads. They first compare multiple strategies for finding minimum-power network subsets across a range of traffic pattern. They implement and analyse Elastic Tree on a prototype testbed built with production OpenFlow switches from three network vendors. Further, we examine the trade-offs between energy efficiency, performance and robustness, with real traces from a production e-commerce website.

Their results demonstrate that for data center workloads, Elastic Tree can save up to 50% of network energy, while maintaining the ability to handle traffic surges. They fast heuristic for computing network subsets enables Elastic Tree to scale to data centers containing thousands of nodes. We finish by showing how a network admin might configure Elastic Tree to satisfy their needs for performance and fault tolerance, while minimizing their network power bill. Elastic Tree can maintain the robustness and performance, while lowering the energy bill.

**PAPER 2: EQVMP: Energy-efficient and QoS-aware Virtual Machine**

**Placement for Software Defined Datacenter Networks**

Cloud computing research becomes a hot topic in recent years. To provide various kinds of applications and services, datacenters need sufficient bandwidth to maintain QoS for communication among millions of network components, resulting in consuming tremendous energy. Hence, how to save energy and to provision sufficient bandwidth are important issues. Finally, we propose our solution to resolve these issues. Datacenters are designed to provide reliable and scalable computing services for massive users. One of the most important things in datacenters is to provide efficient and fault tolerant routing. Most of the efforts are made to reduce the energy consumption by the Data Centers. But both of energy and QoS are critical issue in datacenter networks. Many previous works proposed on VM placement policy guarantee VMs to have sufficient resources and utilize the network resources more effectively. However, unbalance and aggressive placement still induces severe congestion in datacenter networks.

To provide effective and reliable services, cloud datacenters need parallel computing and virtualization techniques. This work presents an improved virtual machine (VM) placement mechanism, called Energy efficiency and Quality of Service (QoS) aware VM Placement (EQVMP) to overcome the problem of unbalanced traffic load in switching on and off VMs for the purpose of energy saving. Therefore, they proposed an energy-efficient and QoS aware VM placement (EQVMP) mechanism. Experiments prove that EQVMP can provide better system throughput than other VM placement strategies. It determines a good VM placement considering energy consumption, hot delay and network throughput. EQVMP combines three key techniques: (1) hop reduction, (2) energy saving and (3) load balancing. Hop reduction can regroup VMs to lower the traffic load among them. Energy saving techniques aim at choosing the appropriate servers. The proposed load balancing updates VM placement periodically. Their experimental results show that the proposed scheme can lower energy consumption and maintain QoS. They propose an evaluation score to assess VM placement in terms of energy, delay and throughput. Comparing to other existing placement policies, their proposed mechanism can enhance system throughput by 25% and can have better evaluation score. Based on our scenario, they provide a long update period to relocate VM placement rather than frequently adjusting them.

**PAPER 3: Energy-aware routing in data center network**

Today’s data centers, containing tens of thousands of switches and servers, run data-intensive applications from cloud services such as search, web email, to infrastructural computations such as GFS, Cloud Store, and MapReduce. The goal of data center network (DCN) is to interconnect the massive number of data center servers, and provide efficient and fault-tolerant routing service to upper-layer applications. It is well known that the current practice of tree architecture in data centers suffers from the problems of low scalability, high cost as well as single point of failure. To overcome the problem of tree architecture in current practice, many new network architectures are proposed, represented by Fat-Tree, BCube, and etc. A consistent theme in these new architectures is that a large number of network devices are used to achieve 1:1 oversubscription ratio. However, at most time, data center traffic is far below the peak value. Specifically, traffic in data center network varies greatly between daytime and night. A clear diurnal pattern emerges: traffic peaks during the day and falls at night. Therefore, a great number of network devices work in idle state in these richly-connected data center networks. At the same time, the energy consumed by power-hungry devices now becomes a headache for many data center owners.

In this paper, they discuss how to save energy consumption in high-density data center networks in a routing perspective. They call this kind of routing as energy-aware routing. The key idea is to use as few network devices to provide the routing service as possible, with no/little sacrifice on the network performance. Meanwhile, the idle network devices can be shut down or put into sleep mode for energy saving. They then formally establish the model of energy-aware routing problem, and prove that it is NP-Hard. Then they proposed a heuristic routing algorithm to achieve their design goal. The algorithm works in the following way. First, the network throughput is computed, which is the most important performance metric for data-intensive computations, according to the routing on all data center switches. The corresponding routing is called basic routing. Second, they gradually remove switches from the basic routing, until when the network throughput decreases to a predefined performance threshold. Third, switches not involved in the final routing are powered off or put into sleep mode. Extensive simulations in typical data center networks is conducted to validate the effectiveness of our energy-aware routing algorithm. The results show that the energy-aware routing algorithm is a feasible and efficient method for saving energy consumed by network devices in data center network, especially under low network loads.

**Y. Shang, D. Li, and M. Xu, “Energy-aware routing in data center network,” in Proceedings of the ﬁrst ACM SIGCOMM workshop on Green networking, pp. 1–8, ACM, 2010.**

**PAPER 4: Energy-aware Virtual Machine Placement in Datacenters**

Modern Virtualized Data Center is increasingly a hosting platform for a wide spectrum of composite applications–such as search engines, social networks, video streaming, medical services, electronic commerce, grid computing and network-based cloud computing. As the trend towards more communication intensive applications in data centers and increasing size of data center will only continue, the energy consumption of network and server resources that underpin will grow. As the trend towards more communication intensive applications in data centers and increasing size of data center will only continue, the energy consumption of network and server resources that underpin will grow. On the one hand, with more communication intensive applications deployed in data centers, the bandwidth usage between virtual machines (VMs) is rapidly growing. This has drawn extensive attention from academia with respect to network energy consumption. On the other hand, with cooling energy demand and energy costs increasing, the server power consumption management has to be considered. Therefore, the engineering challenges and cost of managing the power consumption of large data centers and associated cooling drive the need to reduce data center energy use. While data center energy use has received much attention recently, there has been less attention paid to consider both server energy consumption and the energy consumption of the data center transmission and switching network. With many practical applications, minimizing the total energy consumption in a data center requires the formulation of a joint optimization problem**.** In contract to existing solutions that primarily focuses on only one specific aspect of management to reduce energy consumption, this paper explores the balance between server energy consumption and network energy consumption to present an energy-aware joint virtual machine (VM) placement. Given the definition of VM placement fairness, the basic algorithm of VM placement which fulfils server energy consumption constraints is conducted. Then, they further formulate the VM placement as an optimization problem which considers application dependencies to reduce network energy consumption. They design a joint algorithm that efficiently solves the VM placement problem for very large problem sizes. Using simulations, they conduct a comparative analysis on the impact of the data center architectures, server constraints and application dependencies on the potential performance gain of energy-aware VM placement. Compared to existing generic methods, we show a significant performance improvement such as efficiently reducing the number of physical machines used to save server energy consumption, decreasing the communication distance between VMs to obtain data center.

**D. Huang, D. Yang, and H. Zhang, “Energy-aware virtual machine placement in data centers,” GLOBECOM 2012, 2012**.