Protected Elastic-tree Topology for Survivable and Energy-efficient Data Center

Dieu-Linh Truong, Elena Ouro and Thanh-Chung Nguyen
School of Information and Communication Technology
Hanoi University of Science and Technology, Vietnam
E-mail: linhtd@soict.hust.edu.vn, elenaouro@gmail.com, ntchung92@gmail.com

Keywords: data center, energy efficiency, survivability, elastic tree

Received: January 1, 2016

Recently, using third party data centers has become a popular choice for storing enterprise data and deploying services. The fast growth of data centers in size and in number makes them become huge energy consumption points. Up to 70% of energy consumption is due to server running and cooling. In order to reduce energy consumption, recent researches have proposed turning off certain switches in data centers with little traffic flow. However, when those switches are turned off, the data center becomes vulnerable to failures due to low connectivity between servers. In order to overcome this weakness, this paper proposes to use path protection to ensure that all connections in the data center retain survivability upon any single failure. The paper also proposes an algorithm to calculate a tailored topology for the data center so that unnecessary switches can still be turned off. The simulation results show that the proposed solution makes data centers survivable while still saving energy significantly, mostly in big size data centers.

Povzetek: Nov algoritem poišče primerne varne rešitve v podatkovnih centrih pri varčevanju z energijo na osnovi elastične topologije.

1 Introduction

Cloud computing is currently the common choice for end users and enterprises to store their data and process their services by using third party data centers. End users and enterprises can now focus on their business issues without concerning themselves with the building and maintaining of their own storage servers or network devices since this infrastructure is hosted in data centers. This advantage leads to a quick growth of data centers in size and in number. However, these data centers consume a huge amount of energy for running servers, network devices in order to process user requests and transmit data between servers within data centers. According to US Federal Energy Management Program report in [1], data centers in 2013 accounted for 2.7% of the 3831 billion kWh used in the US. Federal data centers used about 5 billion kWh in 2013, or nearly 10% of federal electricity use. These numbers show the importance of utilising energy efficiently in data centers. It is stated in [2] that cooling infrastructure of data center facilities can require one to two times the energy used to power the IT equipment itself. Therefore, many researches look for efficient energy utilisation solutions for data centers either by locating data centres close to green energy resources, using energy efficient devices [3], or by limiting the number of running devices in data center.

Servers in a data center interconnect through a system of switches. Commonly, these switches are organised in hierarchical topology with 3 layers: a core layer, an aggregation layer and an edge layer. The switches in the edge layer link to servers (hosts). In upstream direction, traffic from source servers arrives at edge switches then is regrouped and routed to the aggregation layer switches. In turn, the aggregation layer switches regroup traffic to forward to a core switch. Traffic follows the downstream direction similarly to get to the destination servers.

Fat-tree topology [4] was proposed for data centers for the first time in 2008 [5]. A k-ary Fat-tree contains k modular data centers called PODs (Performance Optimized Data center) that links together through core layer switches. Each POD contains two layers of aggregation and edge switches where each switch of one layer links to all switches of the other layer. Figure 1 shows an example of 4-ary Fat-tree data center. Fat-tree structure has been developed to reduce the over-subscription ratio and to remove the single point of failures of the hierarchical architecture in the data center network through multiple-linking of a switch to other layers.

Fat-tree data centers tend to consume a lot of energy, however. Data centers in general, and Fat-tree data centers in particular, are usually designed with sufficiently high capacity to tolerate access traffic in peak hours, meanwhile the access traffic varies significantly within a day, a week, or month. As a result, data centers often work under capacity during normal hours. It is wasteful if all devices in a data center are kept working during this time.

With the objective to save energy a solution, Elastic tree, has been proposed in which the main idea is to turn off some devices in Fat-tree [6] when they are not needed. In

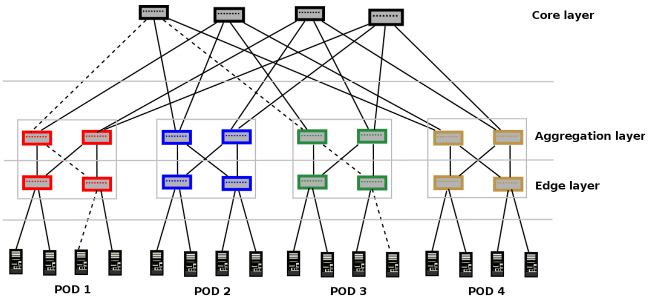


Figure 1: A Fat-tree data center with $k = 4$.

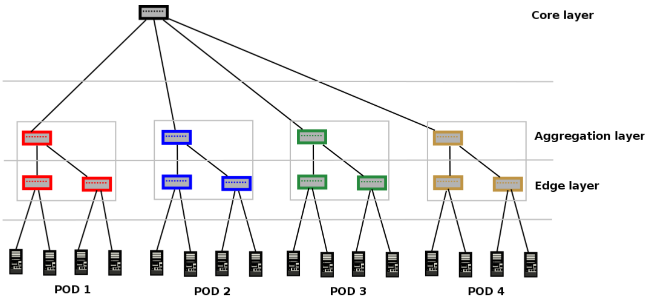


Figure 2: An Elastic-tree data center with $k = 4$ when 3 core switches and some aggregation switches are turned off.

Elastic tree, the traffic is compressed to some switches and links while the free switches and ports are turned off. The topology undergoes dynamic changes in concert with the change of traffic. That is why the topology is called Elastic tree. Elastic tree certainly saves more energy than Fat-tree. The idea of Elastic tree has also been developed further in RA-TAH [7] where besides turning off some devices, some others can also be put into sleep mode or into a reduced working rate for more efficient use of electricity. However, both Elastic tree and RA-TAH make data center vulnerable to failures due to low connectivities. For example, in Figure 2, after turning off 3 core switches and some aggregation switches, the data center becomes a tree with a single root, then there is a single path between any pair of servers. If a failure occurs on any link between the core and aggregation layers, a whole POD will be disconnected. If a failure occurs between the aggregation layer and the edge layer, an edge switch with all its associated servers will be disconnected from the data center. Although it is possible to wake up or to turn on the devices for taking over the traffic from the failed switches/links, the recovery delay remains important due to the slow starting/waking up process and long traffic re-direction process.

In this paper, we propose to add the protection capability to the Elastic tree so that even if some devices are turned off, the data center remains survivable upon any single failure. The proposed architecture is called Protected Elastic tree. We focus on protecting the communication between servers within the data center only. The main contributions of the paper are: i) proposition for the use of a path protection model in Elastic tree, ii) calculation of the minimum energy consumption topology for the Protected Elastic tree, and iii) analysis of the impact of different factors on the energy saving capability of the Protected Elastic tree topology.

A preliminary result of this research was presented in [8]. In that paper, due to limitations in the implementation, we were only able to perform the experiments on a single data center size with low traffic loads. In this paper, with the new implementation, we perform more extensive experiments and study further the impacts of network load and data center sizes on the energy saving level of the proposed Protected Elastic tree topology.

The remainder of this paper is organised as follows. The next section presents briefly the idea of conventional Elastic tree. Section 3 presents the protection model that we propose to use to reinforce Elastic tree data centers. Section 4 explains how to find the topology of the Protected Elastic tree. Section 5 shows how much energy can be saved with Protected Elastic tree and analyse the impact of traffic load and data center size on energy saving. Finally, Section 6 concludes the paper.

2 Elastic tree

Fat-tree term usually refers to the hierarchical topology where links closer to the root have greater capacities, although the Fat-tree concept here should be understood differently. It is a type of multistage circuit switching network. A k-ary Fat-tree contains k PODs. Each POD contains two layers of $k/2$ aggregation switches and $k/2$ edge switches. Each edge switch links to $k/2$ servers (or hosts) and $k/2$ aggregation switches of the upper layer. Each core switch links to k PODs by k aggregation switches. All switches are identical with k ports. Figure 1 shows the connection path between two servers through 3 layers of switches in dash line in a 4-ary Fat-tree data center. Fat-tree has a robust structure but it still consumes a significant amount of energy.

Elastic tree as proposed in [6] is a system for dynamically adapting the energy consumption of a Fat-tree data center network. Elastic tree is controlled by three logical modules Optimiser, Routing, and Power control as shown in Figure 3. The role of the Optimiser module is to find the minimum power network subset, which satisfies current (or statistical) traffic conditions. Its input consists of the topology, a traffic matrix, and the power model of each switch. The Optimiser outputs a set of active components to both the Power control and Routing modules. The Power control then turns on or off ports, linecards, or entire switches according to this output, while the Routing module chooses routes for all flows, then pushes routes to the switches in the network.

Elastic tree refers to the switch power consumption model proposed in [9]. Power consumption of a switch consists of a fixed component (chassis power and power

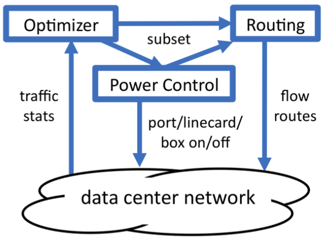


Figure 3: Control components of Elastic tree system (figure from reference [6]).

consumed by each linecard) and a variable component that depends on the number of active ports and the capacity and utilisation of each port.

$$\begin{aligned} P_{switch} = P_{chassis} &+ n_{linecard} \times P_{linecard} \\ &+ \sum_{i} np_{r_i} \times P_{r_i} \times F_u \end{aligned} \qquad (1)$$

Where P_{switch}, $P_{chassis}$, $P_{linecard}$ and P_{r_i} are respectively, the power consumed by a switch, a chassis, a linecard without active ports and an active port running at rate r_i. np_{r_i} is the number of active ports running at rate r_i. F_u is an utilisation scaling factor for each port. For simplicity, this factor is considered identical for all switch ports in Elastic tree.

It is clear that an active switch without traffic processing still consumes energy. Consequently, a minimum power network subset is sought for carrying the required network load for a data center. This network contains only switches and links that must be active in each layer in order to satisfy the network load. With the help of Software Defined Network technologies such as OpenFlow [10], unnecessary switches, line cards or ports are turned off to conserve energy. The network traffic is then routed so that it flows only over the active switches. Network traffic may change and the minimum power subset of active switches must be recomputed.

Some algorithms have been proposed to identify the minimum power network subset for a data center given a required network load. The algorithms include a multi-commodity flow formulation which is a mixed integer linear program, a greedy bin-packing algorithm and a Topology-aware Heuristic (TAH) [6]. Amongst these algorithms, TAH is the fastest with the least computational effort. TAH does not identify exactly which switches are a part of the minimum power network subset but instead identifies the number of active switches in each layer. From these numbers the total power consumed by the Elastic tree can be estimated.

The idea of TAH is as follows; based on the statistic of traffic between layers, TAH identifies the number of links required between an edge switch and the aggregation layer to support the up and down traffic from and to that edge switch. Assuming that the traffic is perfectly divisible, this number of links is equal to the traffic bandwidth divided by the link rate. The number of active switches in the aggregation layer is then equal to the number of links required to support traffic of the most active source from the upper layer or lower layer. Similar observation holds between the the aggregation layer and the core. Detailed computation can be found in [6]. In this paper, TAH is used to estimate the energy consumption by an Elastic tree.

In [6], the authors have also discussed adding a redundancy level to the Elastic tree by adding k parallel Minimum Spanning Trees (MSTs) that overlap at the edge switches. However, the study does not investigate further on how many MSTs would be sufficient.

3 Protection model for Elastic tree Data center

As seen in Figure 2, an Elastic tree is sensitive to failures. Its connections need to be protected. In this study we focus solely on single failure scenario. The single failure scenario assumes that there is, at most, one failure in the data center and this failure is repaired before another one may occur. The single failure scenario is a typical assumption in the research and also in practice since the frequency of failure is low [11] making the possibility of having multiple failures negligible.

There are several well-known topological protection schemes: path protection, link protection, and ring protection [12]. In the path protection model, the connection to be protected is called working path. Another path, called backup path, that shares the same end nodes with the working path is used to replace the working path when the latter fails due to link or node failures. When no failure is present the backup path is idle. Since the backup path should not be affected by any failure on the working path, the two paths should be disjoint. In link protection, each link in the working path is protected separately by a backup segment going from one end of the link to the other end. When a link fails, only one backup segment is used to replace the link. Intuitively, the link protection tends to require more backup resource than path protection since many backup segments are involved for protecting a single path. Ring protection is used in networks with ring topology. In that network, a connection follows a part of the ring and is protected by a backup connection following the remaining part of the ring in inverse direction. Due to the tree form topology of Fat-tree data centers, we focus on the path protection scheme.

A server-to-server connection in data center may travel through a suite of devices in the following order: source server - edge switch - aggregation switch - core switch - aggregation switch - edge switch - destination server. In Protected Elastic tree, we propose to use the path protection model to protect the part between the edge switches. The part between the source server - edge switch and the part between the edge switch - destination server will be left unprotected.

According to path protection model, the part between two edge switches will be protected by a disjoint path between the same two edge switches. From this principle, we apply three different protection configurations for Near, Middle and Far traffic. The notions of Near, Middle and Far traffic are defined as in [6] and [7].

- Near traffic refers to a flow between source and destination servers linked to the same edge switch. For this kind of flow, no protection is offered.

- Middle traffic refers to a flow between source and destination servers linked to different edge switches of the same POD. This kind of flow passes through an intermediate aggregate switch then returns to the edge layer within the POD. The backup flow simply uses another aggregate switch in order to guarantee that the working flow is disjointed. See Figure 4 for an example.

- Far traffic refers to a flow between source and destination servers linked to different PODs. A flow in this traffic passes through an aggregation switch of the source POD, to a core switch, then to an aggregation switch of the destination POD, and finally to the destination edge switch. The backup flow needs to use different aggregation and core switches than the working flow in order to guarantee the disjointedness. Readers are referred to Figure 5 for an example of a working flow and its backup flow of the far traffic.

Looking at the topological aspect only (without link capacity consideration), the protection configuration for Middle traffic is always possible since all aggregation switches and edge switches of the same POD link to each other. Protection configuration for Far traffic is also always possible. Since each core switch links with all PODs then with $k \geq 4$ there are always at least four core switches linking to both the source and destination PODs. Amongst these four core switches there exists at least two core switches linking to two different pairs of aggregation switches; one pair in the source POD, and the other in the destination POD. Therefore, there always exists two disjoint paths between two edge switches through these two different pairs of aggregation switches. If one path is serving as the working path then the other can be utilised as the backup path.

Now consider the bandwidth capacity aspect. In non-protection mode, a Fat-tree data center with unified link capacity is fully loaded when all servers saturate their links to edge layer. In protection mode links between edge, aggregation and core layers have to carry not only the working flows but also the backup flows, therefore the total consumed bandwidth is doubled in comparison with non-protection case. Consequently, under high traffic load, it is still possible that the data center will not have enough capacity to allocate backup flows for all working flows.

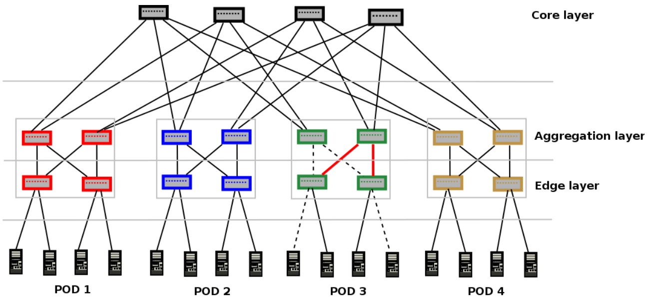


Figure 4: Example of a backup configuration for Middle traffic. The working flow is in dash line and its backup flow is in red line.

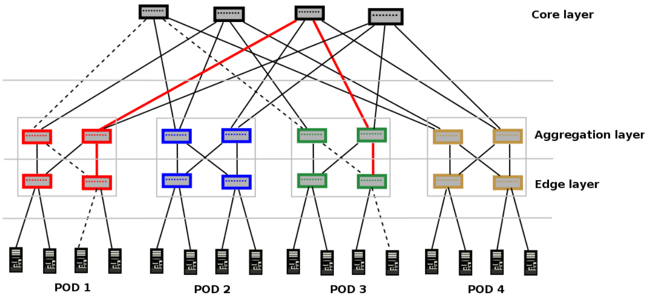


Figure 5: Example of a backup configuration for Far traffic. The working flow is in dash line and its backup flow is in red line.

4 Minimum active topology with path protection

Similar to the conventional Elastic tree, the Optimiser in Protected Elastic tree must find the minimum active topology, however in this instance the path protection mechanism is integrated. The minimum active topology with path protection is a minimal network subset that can accommodate not only all working flows for a given traffic matrix but also a backup flow for each working flow. The traffic matrix considered here contains the demanded flows between pairs of edge switches. Since we offer the protection between edge switches, we are interested in flows between edge switches only. A flow of traffic between a pair of edge switches is in fact an aggregation of multiple flows between servers. Here we assume that a rough traffic matrix between edge switches is known in advance. This assumption is quite practical since this rough traffic matrix can be estimated easily by observing the traffic statistics over time.

In order to find such a minimal network subset, we will try to accommodate working and backup paths for all flows in the traffic matrix. During this accommodation process, switches and ports are activated gradually. We assume that all switches and links in the data center are identical. Switches have k ports and link rate is r. The switches in data center are linked in k-ary Fat-tree topology. We also assume that flows are perfectly divisible. Working flows between edge switches are allocated one by one, followed by the backup flows. The flows are aggregated maximally to active links, switches and ports until those active devices are fully used before activating new devices. The following steps describe how to build the minimum active topology with path protection.

- We start with an initial active topology as the set of switches forming a Minimum Spanning Tree (MST) with a core switch as root and all servers are leaves.
- For allocating the working flow from a source edge switch e_s to a destination edge switch e_d, we browse all active paths between e_s and e_d to find the one with largest residual bandwidth. In the case that all paths are fully used, we browse from all remaining possible paths between e_s and e_d to find the one that requires least additional power consumption for activating. The power consumption of an newly activated path includes the power for activating new switches, ports and line cards as estimated by (1).
- Once all working flows are routed their backup paths will be accommodated. In order to find a backup flow for a working flow between e_s and e_d, we first exclude the aggregate switches of its working flow from the list of usable switches for the backup path. This exclusion ensures that the backup path is disjoint from the working path. Then, the backup flow is sought in similar manner as when we allocate a working flow.

If the algorithm fails to find backup paths for some of the working flows, the Optimiser ends with a false status. Once the algorithm terminates successfully each connection request in the traffic matrix will have a backup path which is disjoint with the working path. Therefore, when a single failure occurs in the network the working path or the backup path will not be affected, thus one of them is available for carrying the traffic. Consequently, the data center is 100% available to carry the traffic in the traffic matrix under any single failure.

Although in the above procedure we have to browse all the possible paths between two edge switches, the special structure of Fat-tree limits the number of paths available to browse. From an edge switch there are $k/2$ choices of aggregation switches to go up. For each aggregation switch, there are again $k/2$ choices of core switches. From a core switch there is only a single choice of route to an edge switch. Therefore, there are at most $k^2/4$ possible paths between two edges switches in Far traffic. Similarly, there are at most $k/2$ possible paths between two edge switches in Middle traffic.

Once the Optimiser has calculated the minimum active topology and the working and backup paths for flows in the traffic matrix, it gives the Routing module this information, e.g., the list of active switches and the list of working and backup paths for flows between edge switches. Later on, when the Routing module receives a connection request between two servers, it can easily identify the two edge switches associated with the two servers thanks to child-parent relationship of these devices. The Routing module will route the connection request over the working path registered for the flow between the two switches while the corresponding backup path is used for protection.

5 Evaluation of energy consumption of Protected Elastic tree data centers

The advantage of path protection scheme integrated in Protected Elastic tree data center is obvious as the data center is 100% survivable upon any single failure. However, with the presence of a backup path in parallel with a working path for each flow between edge switches, the data center consumes more energy than the conventional Elastic tree data center. In order to evaluate how much of energy the protection scheme imposes into the data center, the energy consumption of the Protected Elastic tree is compared against the fully active Fat-tree data center and against the conventional Elastic tree data center without protection.

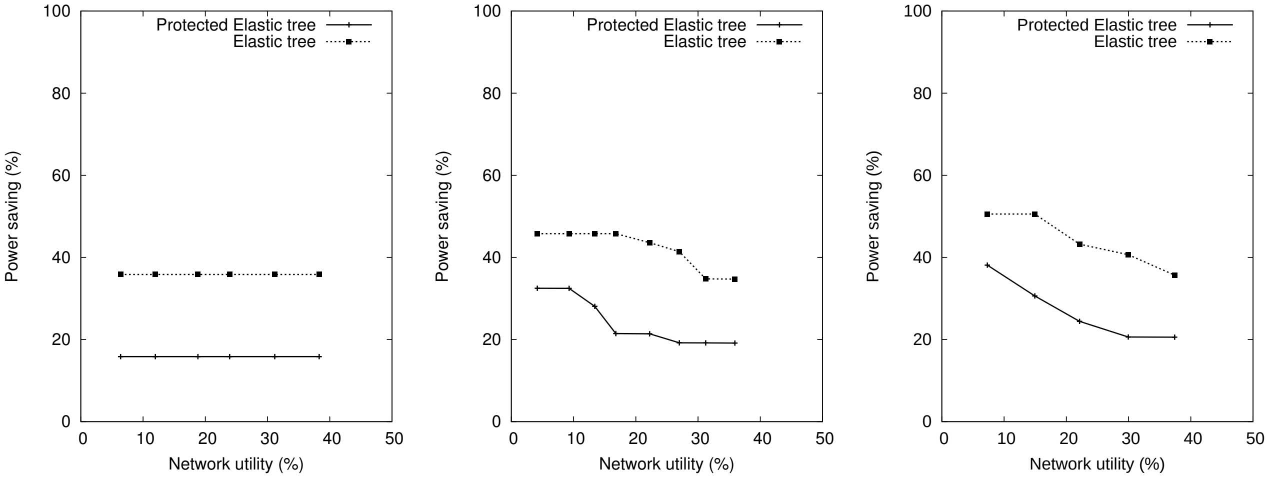
The best topology for the Protected Elastic tree is identified by “Minimal active topology with path protection” algorithm proposed in Section 4. In the previous work in [8], the algorithm was implemented in a data center emulation platform: Ecodane [13] which did not allow tests with a high traffic load due to the large computation effort that they involve. In this paper, the algorithm has been implemented in Scilab [14], an open source numerical computational package. With the new implementation, we can perform more extensive experiments. The minimal topology for the conventional Elastic tree is calculated using the TAH algorithm which has also been implemented in Scilab.

Let λ_{ij} be the total requested bandwidth from server i to server j. Let $\#server$ be the number of servers in the data center. The total bandwidth capacity provided by links between servers and edge switches is $\#servers \times r$. Since a flow between server i and server j uses two links between server layer and edge layer, the maximum total acceptable load between servers in the data center is $0.5\#servers \times r$. From this observation, we define the network utilisation index as the ratio between the total requested bandwidth between servers in data center and the maximum total acceptable load of the data center. It is calculated by:

$$u = \frac{\sum_{ij} \lambda_{ij}}{0.5\#servers \times r} \times 100\% \tag{2}$$

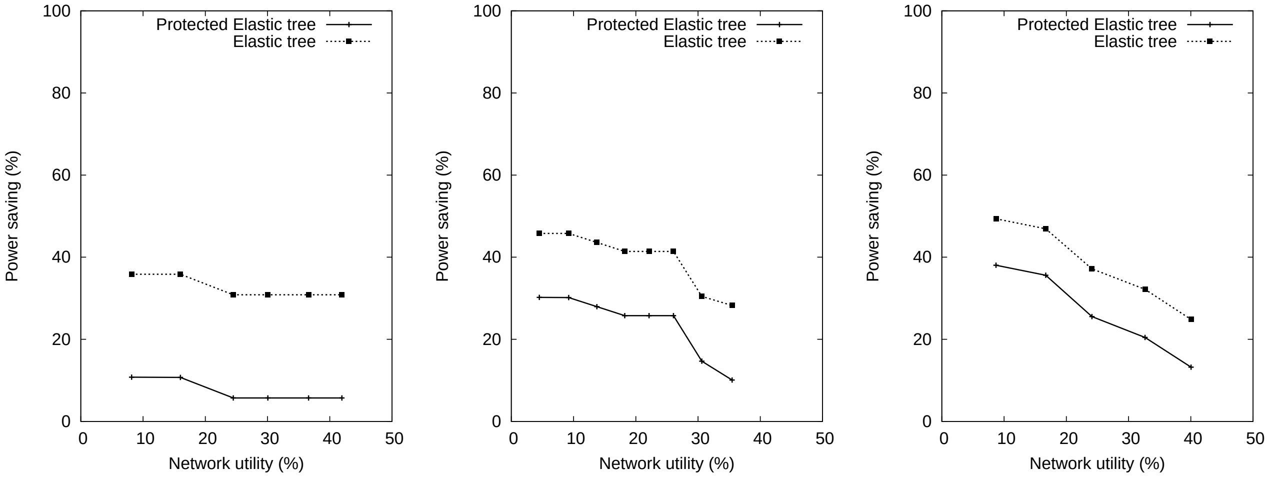
In case on non-protection, when $u = 100\%$ the data center is maximally loaded. In case of protection, for each pair of servers, a flow for working path needs another flow for backup path. Therefore, the maximum network utilisation for protected data center is $u = 50\%$.

In all simulations, switches are wired in Fat-tree topology. Data center sizes vary with $k = 4, 6, 8$. All links



(a) k=4. (b) k=6. (c) k=8.

Figure 6: Power saving level with Middle traffic



(a) k=4. (b) k=6. (c) k=8.

Figure 7: Power saving level with Far traffic

are bi-directional with bandwidth capacity of $r = 10$ Gbps. Traffic requests are generated for 3 traffic models: Far traffic, Middle traffic and Mixed traffic. Mixed traffic does not include Near traffic since we are not interested in protecting Near traffic in this research. In order to reserve enough spare capacity for protection, traffic is generated with increasing network utilisation until it approaches $u = 50\%$ but without saturating the data center. Each server-to-server connection requests a bandwidth uniformly distributed in range [0-1] Gbps.

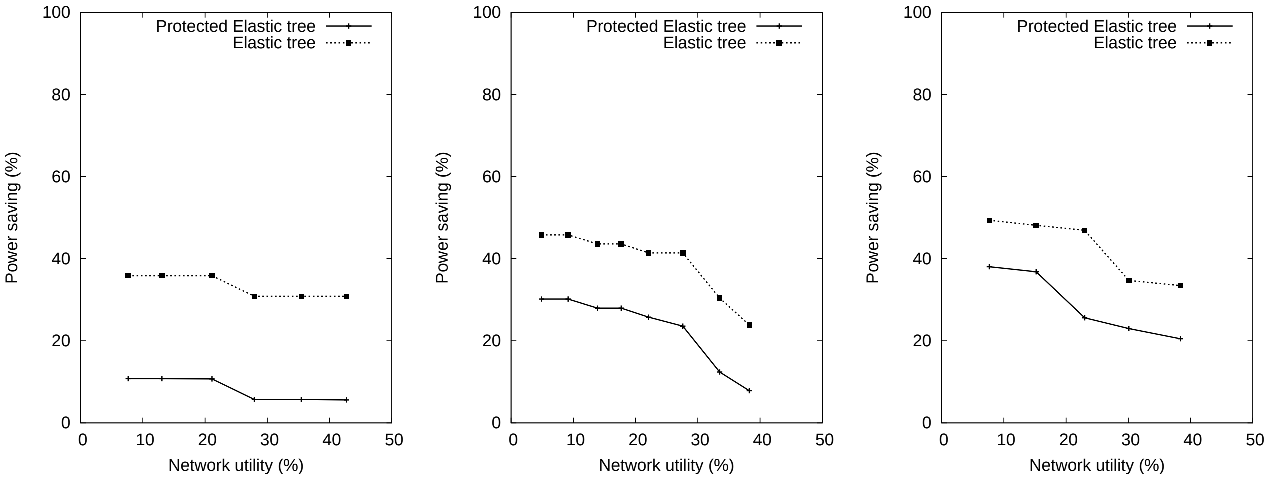
The total energy usage by Fat-tree and conventional Elastic tree are evaluated based to their number of active devices. According to (1), the power consumption of a switch consists in a fixed part including power consumption of chassis and line cards, and a dynamic part including power consumption of ports. Let us denote the fixed part of the power consumption by P. Let us also consider that the power consumption of a port is constant regardless of its working rate and is denoted it by p.

It is easy to prove that a k-ary Fat-tree data center has a total of $5k^2/4$ switches, $5k^3/4$ ports and $3k^3/4$ links. Since they are all active, the power consuming by a Fat-tree, regardless of traffic matrix, is

$$P_{Fat} = \frac{5}{4}k^2P + \frac{5}{4}k^3p \qquad (3)$$

In Elastic tree, all edge switches must be active in order to be ready to receive data from the servers, therefore, the number of active edge switches is always $k^2/2$. Assume that a conventional Elastic tree topology, identified by TAH for a given traffic matrix, uses x aggregation switches ($x \leq k^2/2$), y core switches ($y \leq k^2/4$) and has n_p active ports, then the total energy consumed by the Elastic tree is

$$P_{Elastic} = \frac{k^2}{2}P + xP + yP + n_pp \qquad (4)$$



(a) k=4. (b) k=6 (c) k=8

Figure 8: Power saving level with Mixed traffic

Th Energy saving level of the Elastic tree over the Fat-tree is defined as:

$$\left(1 - \frac{P_{Elastic}}{P_{Fat}}\right) \times 100\% \qquad (5)$$

We denote the energy consumed by Protected Elastic tree by $P_{Protected.Elastic}$. The energy consumed by Protected Elastic tree is also evaluated by using (4) where x, y and n_p are the number of active aggregation switches, the number of active core switches and the number of active ports in Protected Elastic tree. Those numbers are obtained from the calculations in the algorithm “Minimal active topology with path protection”. The energy saving level of the Protected Elastic tree over Fat-tree is defined as:

$$\left(1 - \frac{P_{Protected.Elastic}}{P_{Fat}}\right) \times 100\% \qquad (6)$$

In all tests, the power consumed by a switch chassis including line cards is set $P = 146$ watts, the power consumed by a port is set $p = 0.9$ watts. These numbers have been chosen after analysing the study of switch power consumption in [9].

Figure 6, 7 and 8 show the power saving level of the Protected Elastic tree and Elastic tree with $k = 4, 6, 8$ for Middle, Far and Mixed traffic. The detailed results are shown in Table 1, 2 and 3 where the columns show the following information:

- Network utility (in percentage)
- Power consumed by Fat-tree topology P_{Fat}
- Number of active switches, number of active ports, total power consumption and energy saving level (in percentage) of the Protected Elastic tree
- Number of active switches, number of active ports, total power consumption and energy saving level of the conventional Elastic tree

It is clear that the Protected Elastic tree consumes more energy than the conventional Elastic tree. The observed gap between energy saving level of Protected Elastic tree and Elastic tree varies roughly between 10% and 25%.

In the following sessions the impact of network utility, traffic model and data center size on the energy saving performance of the proposed Protected Elastic tree topology are analysed.

5.1 Impact of network utility on energy saving

We can observe in Figures 6, 7, 8 that, except in Figure 6a, the power saving levels of both conventional Elastic tree and Protected Elastic tree decrease when the network utility increases. The obvious reason is that when the network utility increases, both the conventional Elastic tree and Protected Elastic tree have to use more switches in order to accommodate the increase in traffic load.

In the case of the data center with size $k = 4$ for Middle traffic in Figure 6a, the power saving levels of both Elastic tree and Protected Elastic tree are constant regardless of the network utility. This special case can be explained as follows; with the Middle traffic, all flows (including working and backup) travel only inside a POD and do not involve core switches. In a non-protected Elastic tree data center, all edge switches are active in order to be ready to receive traffic from the servers. At the aggregation layer, one aggregation switch per POD must be used to carry all traffic from the POD’s edge switches. With the network utility at 50% or under, one aggregation switch with two aggregation-edge links is sufficient to carry the traffic from/to the four servers of the POD. Therefore the total number of active switches is always: 8 (edge) + 4 (aggregation) =12 switches. Now let us consider the Protected Elastic tree data center, again 8 edge switches must be active. Both aggregation switches inside a POD must be active in

		Protected Elastic tree				Elastic tree			
Utility	P_{Fat}	Nb. swt.	Nb. ports	Power	Saving	Nb. swt.	Nb. ports	Power	Saving
Far traffic									
8.16	2992	18	46	2669.4	10.78	13	24	1919.6	35.84
16	2992	18	48	2671.2	10.72	13	24	1919.6	35.84
24.49	2992	19	52	2820.8	5.72	14	28	2069.2	30.84
30.05	2992	19	52	2820.8	5.72	14	28	2069.2	30.84
36.59	2992	19	52	2820.8	5.72	14	28	2069.2	30.84
41.95	2992	19	52	2820.8	5.72	14	28	2069.2	30.84
Middle traffic									
6.4	2992	17	40	2518	15.84	12	16	1766.4	40.96
11.98	2992	17	40	2518	15.84	12	16	1766.4	40.96
18.81	2992	17	40	2518	15.84	12	16	1766.4	40.96
23.91	2992	17	40	2518	15.84	12	16	1766.4	40.96
31.16	2992	17	40	2518	15.84	12	16	1766.4	40.96
38.29	2992	17	40	2518	15.84	12	16	1766.4	40.96
Mixed traffic									
7.65	2992	18	46	2669.4	10.78	13	24	1919.6	35.84
13.08	2992	18	46	2669.4	10.78	13	24	1919.6	35.84
21.11	2992	18	48	2671.2	10.72	13	24	1919.6	35.84
27.91	2992	19	52	2820.8	5.72	14	28	2069.2	30.84
35.45	2992	19	52	2820.8	5.72	14	28	2069.2	30.84
42.73	2992	19	56	2824.4	5.6	14	28	2069.2	30.84

Table 1: Test results with k=4

		Protected Elastic tree				Elastic tree			
Utility	P_{Fat}	Nb. swt.	Nb. ports.	Power	Saving	Nb. swt.	Nb. ports	Power	Saving
Far traffic									
9.22	6813	32	96	4758.4	30.16	25	48	3693.2	45.79
13.75	6813	33	100	4908	27.96	26	52	3842.8	43.6
18.21	6813	34	104	5057.6	25.77	27	56	3992.4	41.4
22.13	6813	34	104	5057.6	25.77	27	56	3992.4	41.4
26.04	6813	34	104	5057.6	25.77	27	56	3992.4	41.4
30.6	6813	39	132	5812.8	14.68	32	74	4738.6	30.45
35.47	6813	41	156	6126.4	10.08	33	78	4888.2	28.25
Middle traffic									
9.28	6813	31	84	4601.6	32.46	24	36	3536.4	48.09
13.41	6813	33	92	4900.8	28.07	24	36	3536.4	48.09
16.79	6813	36	106	5351.4	21.45	24	36	3536.4	48.09
22.22	6813	36	110	5355	21.4	25	40	3686	45.9
26.99	6813	37	114	5504.6	19.2	26	44	3835.6	43.7
31.23	6813	37	116	5506.4	19.18	29	58	4286.2	37.09
35.92	6813	37	118	5508.2	19.15	29	62	4289.8	37.04
Mixed traffic									
9.16	6813	32	96	4758.4	30.16	25	48	3693.2	45.79
13.87	6813	33	100	4908	27.96	26	52	3842.8	43.6
17.7	6813	33	100	4908	27.96	26	52	3842.8	43.6
22.09	6813	34	104	5057.6	25.77	27	56	3992.4	41.4
27.59	6813	35	108	5207.2	23.57	27	56	3992.4	41.4
33.51	6813	40	142	5967.8	12.41	32	76	4740.4	30.42
38.26	6813	42	162	6277.8	7.86	35	86	5187.4	23.86

Table 2: Test results with k=6

		Protected Elastic tree				Elastic tree			
Utility	P_{Fat}	Nb. swt.	Nb. ports	Power	Saving	Nb. swt.	Nb. ports	Power	Saving
Far traffic									
8.7	12256	51	164	7593.6	38.04	42	84	6207.6	49.35
16.7	12256	53	172	7892.8	35.6	44	92	6506.8	46.91
24.12	12256	61	246	9127.4	25.53	52	122	7701.8	37.16
32.64	12256	65	288	9749.2	20.45	56	152	8312.8	32.17
40.04	12256	71	300	10636	13.22	62	176	9210.4	24.85
Middle traffic									
7.32	12256	51	154	7584.6	38.12	40	64	5897.6	51.88
14.94	12256	57	202	8503.8	30.62	40	64	5897.6	51.88
22.13	12256	62	232	9260.8	24.44	46	92	6798.8	44.53
29.95	12256	65	266	9729.4	20.62	48	116	7112.4	41.97
37.41	12256	65	272	9734.8	20.57	52	144	7721.6	37
Mixed traffic									
7.67	12256	51	164	7593.6	38.04	42	84	6207.6	49.35
15.19	12256	52	168	7743.2	36.82	43	88	6357.2	48.13
22.99	12256	61	238	9120.2	25.59	44	92	6506.8	46.91
30.09	12256	63	268	9439.2	22.98	54	134	8004.6	34.69
38.36	12256	65	282	9743.8	20.5	55	142	8157.8	33.44

Table 3: Test results with k=8

order to provide disjoint working and backup paths for each connection in the POD. Therefore, 8 aggregation switches will be involved. Since Protected Elastic tree topology is built from a MST with a core switch as root then one core switch will also be involved. Consequently, the total number of active switches is always 8 (edge) + 8 (aggregation) + 1 (core)=17 switches regardless of network load. This phenomena of constant power saving does not happen with Middle traffic for bigger data center sizes.

5.2 Impact of traffic model on energy saving

Since Middle traffic does not go through core switches as it happens in Mixed and Far traffic, those core switches can be turned off. This characteristic leads to several advantages in Middle traffic:

- The Protected Elastic tree with Middle traffic saves, in general, slightly more energy than Far and Mixed traffic for the same data center and with the same network utility (see Tables 1, 2, 3).
- Protected Elastic tree data centers under Middle traffic always save a certain amount of energy regardless of the network utility. This amount is equivalent to the energy consumed by core switches. That is the reason why the energy saving level lines of Middle traffic become stable as the network utility increases, meanwhile, those for Far and Mixed traffic continue to decrease. We can observe that even for networks with high utility (close to 40%) the saving ratios are still considerable, i.e., 15.84%, 19.15% and 20.57% when $k = 4, 6$ and 8 respectively.

We can also notice that the power saving level lines for Far and Mixed traffic have similar shapes. This similarity demonstrates that the Far traffic has a stronger impact on power saving than Middle traffic. The reason is that Far traffic involves more links and switches due to their long connection paths and Middle traffic can profit from those links and switches for its connections without activating additional elements.

5.3 Impact of data center size on energy saving

The performance of Protected Elastic tree over different data center sizes can also be observed in Figures 6, 7, 8. We have two remarks:

- Power saving lines of both Elastic and Protected Elastic tree shift up slightly from $k = 4$ to $k = 6$, and the same observation can be made from $k = 6$ to $k = 8$, for all traffic patterns. This means that both the Elastic tree and the Protected Elastic tree become more energy efficient as the size of the data center increases.
- The power saving lines of the Protected Elastic tree for Far and Mixed traffic are closer to those of the Elastic tree when size of the data center increases. The conclusion to be drawn from this is that the extra energy consumption for protection becomes less important in large data centers. The main reason of this phenomena is that larger data centers have a higher connectivity and, therefore, more possibilities to route backup flows over switches and ports that are already active to carry working flows.

These two observations show that the Protected Elastic tree topology is an option worthy of consideration for large size data centers.

6 Conclusions

The Fat-tree topology has been proposed as a topology for data centers characterised by a low over-subscription and a high availability. Elastic tree has been proposed in order to reduce the energy consumption of Fat-tree data centers by deactivating unnecessary switches and links. Although Elastic tree is a highly energy efficient topology, it severely damages the availability of the Fat-tree since the network connectivity is reduced remarkably. In this paper, we proposed the possibility of adding protection capabilities to the conventional Elastic tree by allocating backup paths for each aggregated flow between edge switches. Many backup paths can profit from existing active switches and ports while some others require the activation of additional elements in the network. This results in a Protected Elastic tree topology where 100% of connections are survivable for any single failure. The simulation results show that the path protection scheme generates a decrease in the energy saving ratio of 10%-25% for the proposed Protected Elastic tree in comparison with the conventional Elastic tree. However, the Protected Elastic tree data center saves a notable amount of energy if it is compared with the Fat-tree, mostly for the Middle traffic model. Moreover, the proposed solution becomes more energy efficient as the data center size increases.

Acknowledgement

The authors would like to thank Assoc. Professor Nguyen Huu Thanh for his encouragement to perform this research.

References

[1] FEPM, "FEMP Data Center Program Overview," Tech. Rep. DOE/EE-1191, Federal Energy Management Program, March 2015.

[2] Anthesis, "Data Center Efficiency Assessment," Tech. Rep. IP:14-08-A, US Natural Resources Defense Council, August 2014.

[3] K. K. Nguyen, M. Cheriet, M. Lemay, V. Reijs, A. Mackarel, and A. Pastrama, "Environmental-aware virtual data center network," *Computer Networks*, vol. 56, no. 10, pp. 2538 – 2550, 2012. Green communication networks.

[4] C. Leiserson, "Fat-trees: Universal networks for hardware-efficient supercomputing," *Computers, IEEE Transactions on*, vol. C-34, pp. 892–901, Oct 1985.

[5] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," in *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication*, SIGCOMM '08, (New York, NY, USA), pp. 63–74, ACM, 2008.

[6] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, "Elastictree: Saving energy in data center networks," in *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*, NSDI'10, (Berkeley, CA, USA), pp. 17–17, USENIX Association, 2010.

[7] T. N. Huu, N. P. Ngoc, H. T. Thu, T. T. Ngoc, D. N. Minh, V. G. Nguyen, H. N. Tai, T. N. Quynh, D. Hock, and C. Schwartz, "Modeling and experimenting combined smart sleep and power scaling algorithms in energy-aware data center networks," *Simulation Modelling Practice and Theory*, vol. 39, pp. 20 – 40, 2013. S.I.Energy efficiency in grids and clouds.

[8] D. L. Truong and T. C. Nguyen, "Protected elastic-tree topology for data center," in *Proceedings of the Sixth International Symposium on Information and Communication Technology, Hue City, Vietnam, December 3-4, 2015*, pp. 129–134, 2015.

[9] P. Mahadevan, P. Sharma, S. Banerjee, and P. Ranganathan, "A power benchmarking framework for network devices," in *Proceedings of the 8th International IFIP-TC 6 Networking Conference*, NETWORKING '09, (Berlin, Heidelberg), pp. 795–808, Springer-Verlag, 2009.

[10] "Openflow." `https://www.opennetworking.org/sdn-resources/openflow/`.

[11] M. To and P. Neusy, "Unavailability analysis of long-haul networks," *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 100–109, Jan. 1994.

[12] J. Zhang and B. Mukheriee, "A review of fault management in WDM mesh networks: basic concepts and research challenges," *IEEE Network: Magazine of Global Internetworking*, vol. 18, pp. 41–48, Mar. 2004.

[13] H. T. Nguyen, D. C. Bui, D. T. To, N. N. Pham, Q. T. Ngo, T. H. Truong, and M. N. Tran, "Ecodane: A customizable hybrid testbed for green data center networks," in *Advanced Technologies for Communications (ATC), 2013 International Conference on*, pp. 312–317, Oct 2013.

[14] "Scilab." `http://www.scilab.org`.