A10-169

作品名稱

盲路由負載平衡交換網路: 一個無網路處理器的Fat-tree拓撲 Route-blind Load-balanced Switching Network: An NP-free Fat-tree Topology

隊伍名稱

平衡 Balance

隊長

闕宏時 清華大學通訊工程研究所

隊員

謝承宏・吳政家 清華大學資訊工程研究所

作品摘要

根據網路製造商的建議，如Cisco或是Juniper，目前網際網路的建構為連接路由器的樹枝狀拓撲(tree topology)。資料流量透過樹枝狀節點的路由器查表被逐層匯流並轉送。然而，隨著科技的進步，頻寬的增加已快速超越網路處理器可以負荷的速度。目前以網路處理器查表為基礎的樹枝狀拓撲，終將無法跟上網路服務的要求及頻寬的增加。

在這份文件中我們提出了一個前所未有的想法，我們把整個網路視為一個虛擬的交換機。在虛擬的交換機中，網路的拓撲成為交換核心。我們提出利用負載平衡布可夫馮紐曼交換機當作虛擬交換機的核心，因為負載平衡布可夫馮紐曼交換機有高度的擴展性且可以保證有百分之百的效能。緩衝器和交換節點在虛擬交換機中被連結成一個fat-tree。Fat-tree表示一個樹狀結構的連結頻寬從末端節點至樹根節點逐層增加。我們利用負載平衡布可夫馮紐曼交換機的對稱式分時多工(S-TDM)配對的連結方式免除了網際網路位址的查詢。我們稱這樣的特性為盲路由特性。

然而, 連結頻寬在fat-tree會成指數成長。為了解決這個問題，我們提出了位元逆轉連結方式的設計大幅減少fat-tree的連結頻寬。這樣的設計，交換核心將完全不再需要虛擬輸出佇列。所有的虛擬輸出佇列將被放置在fat-tree的末端。如此一來，將適用於超高速的網路，因為這樣的網路最大的連結頻寬往往會超越記憶體的存取速度。此外，這樣的設計也適用於沒有虛擬輸出佇列在核心網路的電路交換網路。

我們採用先進電信運算架構的機櫃為基礎以建造fat-tree拓撲的盲路由負載平衡的交換雛型。我們設計了一個可被程式化為線卡或是交換核心的18層印刷電路板硬體平台以實現八個節點規模的盲路由fat-tree交換雛型。除此之外, 我們設計的印刷電路板也支援當交換機運轉時可以將線卡插入先進電信運算架構機櫃並自動載入儲存在電路板上快閃記憶體中的程式之熱插拔功能。

在未來的工作，我們將提出結合位元逆轉設計以及速率匯流設計。虛擬交換機不但具備盲路由的特性還將能支援不同的網路存取速度。

指導教授

李端興 清華大學資訊工程學系、清華大學通訊工程研究所

- 李端興教授1983年在台灣拿到清華大學電機學士，並且在1987年和1990年在美國紐約哥倫比亞大學拿到電機碩士及博士。1990年到1998年在美國紐澤西NEC的C&C實驗室擔任研究員。於1998年回國任教於清華大學資訊工程學系，並在2003年8月升為教授。2006年獲得徐有庠基金會最佳論文獎。李教授目前為IEEE的Senior Member。
- 研究領域：包含網路交換機和路由器的設計、無線網路、網路效能分析及排隊理論。

Abstract

Complying with the suggestion from main network equipment manufacturers such as Cisco and Juniper, routers are connected in a tree topology in the present Internet. That is, data flows are aggregated and routed by the routers located in the tree nodes based on Internet address lookup. However, as technology advances, link speeds increase more quickly than processor speeds. The current connection of routers in tree topology with address lookup by network processors will eventually become infeasible as demands for network services and more bandwidth continue.

In this document we propose an unprecedented idea, in which we view the entire network as a large virtual switch. In this virtual switch, the network topology becomes the switching core. We propose load-balanced Birkhoff-von Neumann switches as the core of the virtual switch as load-balanced Birkhoff-von Neumann switches are highly scalable and can guarantee 100% throughput. Buffers and switching nodes are connected by a "fat" tree in this virtual switch. By a fat tree, we mean a tree in which the link speeds increase as one traverses from any leaf node to the root of the tree. The symmetric time division multiplexing connection patterns of the load-balanced Birkhoff-von Neumann switch eliminates the need for Internet address lookups. We call this property the route blind property. The main drawback is that the link speeds in the fat tree increase exponentially. To solve this problem we propose bit reversal design which dramatically decreases the link speeds in the fat tree. In this design, the switch core is completely void of virtual output queues. All virtual output queues are located in the leaves of the fat tree. Thus, this design is particularly suitable for an ultra speed network in which the maximum link speed exceeds the memory access speed. In addition, this design is suitable for a circuit switched network, in which there is no virtual output queue in the core of the network.

We have built a prototyped route-blind load-balanced switch with fat-tree topology in an AdvancedTCA chassis. We have also designed a single eighteen-layered printed circuit board (PCB) hardware platform, which could be programmed to function as a line card or as a switching fabric to realize a route-blind fat tree with 8 nodes. In addition, our PCB design supports hot-swap capability which allows line cards to be plugged into the AdvancedTCA chassis and program code be automatically loaded from compact flash memory on the board while the switch is operational.

For future work we propose to combine the bit reversal design with the rate aggregation design. The result is a virtual switch that is route blind and can support multiple access speeds.