Design of Large-scale Wire-speed Multicast Switching Fabric Based on Distributive Lattice

[1]CUI Kai, [2]LI Ke-dan, [1]CHEN Fu-xing, [1]ZHU Zhi-pu, [1]ZHU Yue-sheng

1. Shenzhen Eng. Lab of Converged Networks Technology, Shenzhen Key Lab of Cloud Computing Tech. & App., Shenzhen Graduate School, Peking University, Shenzhen 518055, China.
2. Class 1, Grade 2, International Division, Shenzhen Senior High School, Shenzhen 518040, China.

Abstract — Ensuring high quality of service (QoS) of multicast video stream over a next generation network is a challenging issue, and how to implement the wire-speed multicast with hardware logical support in the network nodes of every hierarchy is a key solution to achieve high QoS multicast. Currently, the multicast packets are processed in this way, in which they are copied and then scheduled by ports, lastly, sent respectively. But this approach cannot ensure the high QoS in real-time applications. Moreover, the traditional hardware solutions can`t achieve large-scale scalability well owning to their own bottlenecks. In this project, using distributive lattice theory we have constructed a wire-speed multicast switching fabric based on a multi-path self-routing fabric structure which we developed previously, and implemented it on an Altera StratixIV FPGA chip. Also, we have investigated how the structure is used in large scale multicast switching fabric and designed the signaling system and control mechanism to support the process of self-routing and wire-speed fan-out copy of multicast packets.

Keywords — multicast switching fabric, distributive lattice, Multi-path Self-routing Fabric Structure, FPGA

I. INTRODUCTION

According to a research report [1] published by Arbor company and University of Michigan, the video service becomes the major internet applications. The video stream is characterized by multicast in two ways. One is multiple unicast software scheduling, and the other is wire-speed fanout hardware copy. The former performs poorly in real-time and QoS feature, but, the latter can approve the latency and also achieves excellent performance for multicasting. Therefore, looking for a high QoS multicast solution which can provide hardware logical support in the network nodes of every hierarchy is a key R&D point. Currently, the multicast packets approaches cannot ensure the high QoS in real-time applications because the packets are copied and then scheduled by ports, lastly, sent respectively. Also, the traditional hardware solutions can`t achieve good large-scale scalability. In this paper, using distributive lattice theory we have constructed a wire-speed multicast switching fabric based on a multi-path self-routing fabric structure developed previously by us, and implemented it on an Altera StratixIV FPGA chip.

The rest of the paper is organized as follows. The large-scale wire-speed multicast switching fabric system is described in Section II. The hardware implementation in FPGA is given in

Section III. In Section IV the real multicast stream test of the system is shown, and Section V is our conclusions.

II. THEORETICAL BASIS OF THE SYSTEM

The large-scale wire-speed multicast switching fabric system that we presented mainly consists of two components. One is Multi-path Self-routing Switching Structure based on group theory. The other one is sorting unit, which supports wire-speed multicast, based on distributive lattice.

A. Multi-path Self-routing Switching Structure

Multi-path self-routing switching structure [2] is completely self-routing [3] and modular. In such structure, parameter G indicates the group size, and M indicates the quantity of groups [4]. The packet loss rate caused by traffic fluctuation and sudden flow will exponentially decrease when we increase the parameter G [5]. As shown in figure 1 is a multi-path self-routing switching structure in which M is 16 and G is 8.

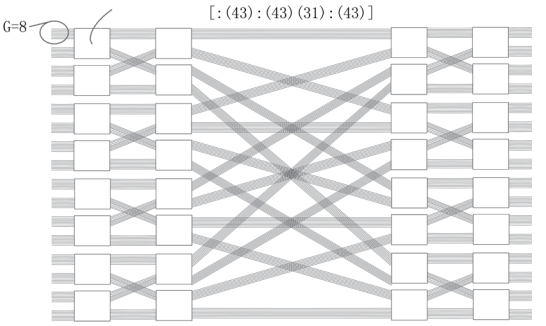


Figure 1. M=16, G=8 Multipath self-routing structure.

B. Sorting unit

Figure 2 describes the normal 2×2 sorting unit. It is the smallest element in switching fabric.

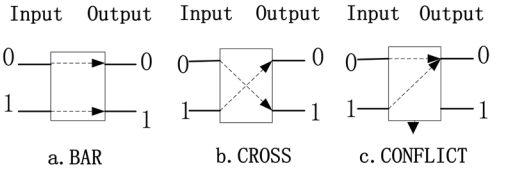


Figure 2. 2×2 sorting unit and its conditions

Such 2×2 sorting unit can act upon this in-band signaling: 10<00<11. The details are shown in Table 1.

TABLE I. Two bits` in-band signaling control mechanism

Conditions		INPUT-1 : A1D1		
		10	00	11
INPUT-0: A1D1	10	CONF	BAR	BAR
	00	CROSS	EITHER	BAR
	11	CROSS	CROSS	CONF

When CONF(CONFLICT), priority decides condition.

Based on distributive lattice, we further defined Ω_{route}*={0-bound,1-bound,idle}*. As a result, the previous ordering relation *10<00<11* equals *0-bound<idle<1-bound*.

When the sorting unit is used for multicast, we should define the multicast condition for the unit, as is shown in figure 3.

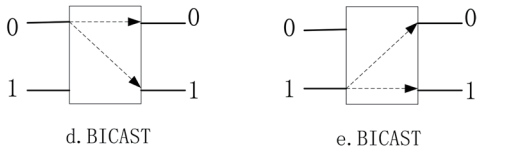


Figure 3. 2×2 sorting unit and its multicast condition

On the basis of the multicast supported sorting unit, we give the new in-band signaling and corresponding control mechanism in Table 2.

TABLE II. 2×2 Control mechanism for multicast sorting unit

Conditions		Input-1			
		0	1	B	I
Input-0	0	CONF	BAR	BAR	BAR
	1	CROSS	CONF	CROSS	CROSS
	B	CROSS	BAR	EITHER	BICAST
	I	CROSS	BAR	BICAST	EITHER

B: BICAST; I: IDLE;

III. SYSTEM DESIGN AND ANALYSIS

The system has been implemented on StratixIV FPGA of Altera. The parameters G and M are 8 and 4 respectively. On the FPGA, the system mainly comprises of user define path and register system.

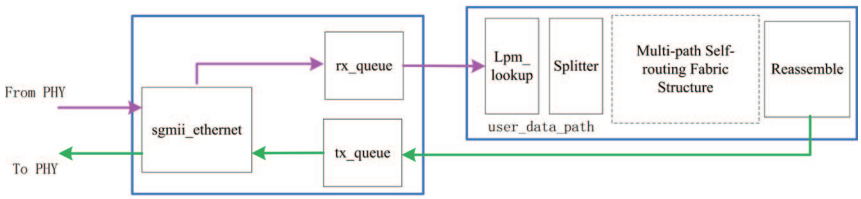


Figure 4. User define path.

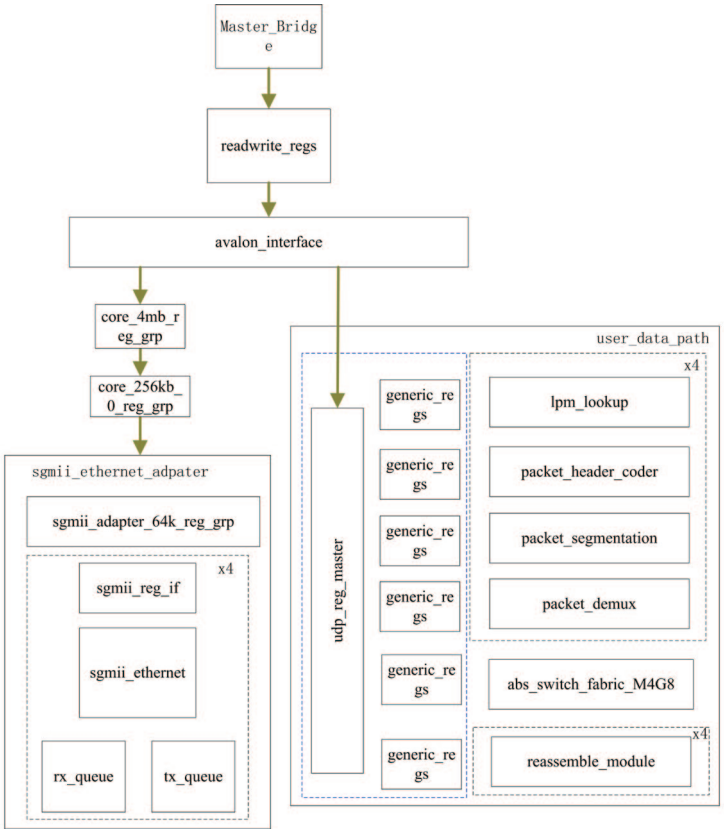


Figure 5. Register system

A. User define path

Figure 4 describes the structure of user define path. It mainly consists of seven sub modules.

Sgmii_ethernet: the interface of system and external PHY chip;

Rx_queue: extract the information of the data packets, such as length, and generate the splitter header;

Lpm_lookup: routing table lookup and generate the lpm header;

Splitter: Split the data packets to cells in certain length and generate the routing control header and cells reassemble header;

Multi-path Self-routing Fabric: Switching fabric;

Reassemble: Reassemble the cells which split in splitter and generate the starting index header;

Tx_queue: Send the data packets to sgmii_ ethernet module and complete switching.

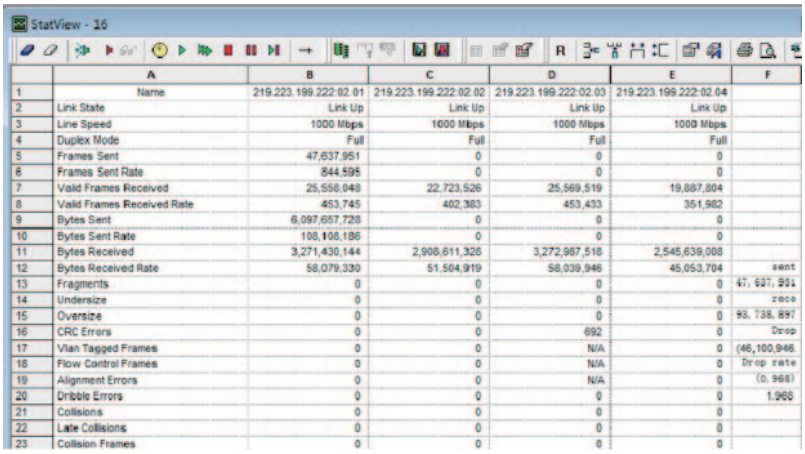
While a packet passing through the system, every sub module will extract the relevant information to generate various headers. The headers will be attached in front of the former packet. Additionally, the system adds two bits` control signal to assist us in data identification and processing. The control signal and packets will transmit in parallel.

B. Register system

The register system not only configures the sub modules in user define path, but also extracts the internal signals of the user define path to help us debug the system. The structure of the register system is revealed in figure 5.

We adopted pipeline architecture to design our register system. Registers are serially connected by specific interface. Every register only responses to the requests which belonging to it. Compared to the star architecture, it is much more convenient when we add another module into the system.

The register system is constructed in Qsys development environment which is attached in Quartus II. We utilized Jtag_Avalon_Master_ Bridge and made the register interface for every sub module based on Avalon Bus Specification. Through the register system, we can use a computer to exchange signals with the system on FPGA.



	A	B	C	D	E	F
1	Name	219.223.199.222:02.01	219.223.199.222:02.02	219.223.199.222:02.03	219.223.199.222:02.04	
2	Link State	Link Up	Link Up	Link Up	Link Up	
3	Line Speed	1000 Mbps	1000 Mbps	1000 Mbps	1000 Mbps	
4	Duplex Mode	Full	Full	Full	Full	
5	Frames Sent	47,637,951	0	0	0	
6	Frames Sent Rate	844,595	0	0	0	
7	Valid Frames Received	25,558,048	22,723,526	25,569,519	19,887,804	
8	Valid Frames Received Rate	453,745	402,383	453,433	351,982	
9	Bytes Sent	6,097,657,728	0	0	0	
10	Bytes Sent Rate	108,106,186	0	0	0	
11	Bytes Received	3,271,430,144	2,908,611,328	3,272,987,518	2,545,639,008	
12	Bytes Received Rate	58,079,330	51,504,919	58,039,946	45,053,704	sent
13	Fragments	0	0	0	0	47, 637, 951
14	Undersize	0	0	0	0	reco
15	Oversize	0	0	0	0	93, 738, 897
16	CRC Errors	0	0	692	0	Drop
17	Vlan Tagged Frames	0	0	N/A	0	(46,100,946
18	Flow Control Frames	0	0	N/A	0	Drop rate
19	Alignment Errors	0	0	N/A	0	(0. 968)
20	Dribble Errors	0	0	0	0	1.968
21	Collisions	0	0	0	0	
22	Late Collisions	0	0	0	0	
23	Collision Frames	0	0	0	0	

Figure 6. Test window

IV. SYSTEM TEST WITH REAL MULTICAST STREAM

We utilized IXIA 400T network tester to test our system. That tester supports 10/100/1000 Ethernet standard and also can generate and count and capture all kinds of streams.

Figure 6 is the test window of the tester. In the test shown in the window, the tester generated a stream from port1 to all the four ports. Now, the system has passed all the normal functional tests including unicast and multicast. Next step, we will implement more complicated tests to figure out the performance of our system in all kinds of network environment.

CONCLUSIONS

This paper constructs multicast sorting unit, on the basis of distributive lattice. Combining with the Multi-path Self-routing Fabric Structure, we constructed the wire-speed multicast switching fabric, implemented the system on StratixIV FPGA, and tested it with real network stream.

The system we implemented on FPGA is not large enough for network size nowadays. However, through it, we get much more familiar with the theoretical basis and development environment. In the near future, we will further research the large-scale utilization of our system.

ACKNOWLEDGMENT

The paper "Design of large-scale wire-speed multicast switching fabric system based on distributive lattice" does not include any other published articles and research achievement excluding the ones listed in the Reference above.

REFERENCES

[1]. Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, Farnam Jahanian; — Internet Inter-Domain Traffic, ACM SIGCOMM 2010；

[2]. Hui Li, Wei He, Xi CHEN, Peng Yi, Binqiang Wang, "Multi-path Self-routing Switching Structure by Interconnection of Multistage Sorting Concentrators", IEEE CHINACOM2007, Aug.2007, Shanghai；

[3]. He W, Li H, Wang B, et al. A Load-Balanced Multipath Self-routing Switching Structure by Concentrators[C]. IEEE ICC 2008；

[4]. 李挥，王秉睿，黄佳庆，安辉耀，雷凯，伊鹏，汪斌强 " 负载均衡自路由交换结构 "，《通信学报》 Vol.30 No..5, 2009 年 5 月 pp.9-15; (EI 20092712164275)；

[5]. 李挥，林良敏，黄佳庆，王蔚，安辉耀，伊鹏，汪斌强，" 具有最小缓存复杂度的负载均衡交换方法 "《电子学报》Vol.37,No.11,2009 年 11 月，pp2367-2372；