On the Routing of Wide-Sense Circuit
Based on Algebraic Switching Fabric

Qian Zhan, Hui Li(✉), Li Ma, and Shijie Lv

Shenzhen Eng. Lab of Converged Networks Technology,
Shenzhen Key Lab of Cloud Computing Tech. & App, Shenzhen Graduate School,
Peking University, Shen Zhen, China
Zhanqian0218@gmail.com, lih64@pkusz.edu.cn,
mali5057@163.com, iamlvshijie@qq.com

Abstract. In order to ensure high quality of service for Next Generation Network, we focus our study on the Wide-Sense Circuit proposed in Flexible Architecture of Reconfigurable Infrastructure. First we construct the functional structure of Wide-Sense Circuit and then explore the switching mechanism and module for it, which is called the Multipath Self-routing Switching Mechanism. Its detailed working process includes traffic classification, establishment and adjustment of Wide-Sense Circuit and data forwarding three parts. For underlying data forwarding, we introduce an innovative Load-Balanced Multipath Self-routing Switching Architecture and start on the implementation on an Altera StratixIV FPGA. The inspiring test results prove that our theory and practiceguarantee the high communication transmission quality for Wide-Sense Circuit.

Keywords: Wide-Sense Circuit · Multipath Self-routing Switching Mechanism · Load-Balanced Multipath Self-routing Switching Architecture

1 Introduction

The present network's architecture is designed for data switching, and TCP/IP, as its fundamental mechanisms, suffers from single function and services as a barely satisfactory schema.That reveals the bottleneck of the general function innetworks, and also contributes the weak adaptability of the capability and structure of networks in different need, which could be seen as the incapability in current network to satisfy more advanced need, like ubiquitous, interconnection, quality, integration, isomerism, credibility, manageability, expandability.

In recent years, efforts have been made by many countries on new types of network architecture, reconfigurable technologies and, routing and switching architectures, such as FIND program in United States, AUTOI program from Challenge One Project in European Unionand AKIRA program in Japan.In China, Information Engineering University proposed the complete system of theories and application test platform of flexible architecture of reconfigurable infrastructure (FARI) [1], including network Atomic Capability (AC), Polymorphic Addressing, Routing and reconfigurable network. Among all, AC theory takes the enhancement of network

V.C.M. Leung et al. (Eds.): TridentCom 2014, LNICST 137, pp. 434–443, 2014.
DOI: 10.1007/978-3-319-13326-3_42

fundamental capacities directly as an entry point, rather than amending or expanding the original networks, which has drawn our academic attention.

AC theory holds that features and requirements of network businessesjobs are diverse and versatile variant with inspect to the finity and certainty. A feasible approach to mitigate this difference, as the core of network AC theory,is to abstract the features and requirements of network jobs as well as network service into a specific, top-down schema of business, Atomic Service(AS) and Atomic Capability.

According to the definition of AS, the network switching mechanism capable of this service is part of AC and satisfies new AS requirements through dynamical adjustment. Wide-Sense Circuit (WSC) [2], as a new basic data transfer mode, is introduced. WSC is an adaptive circuitry built dynamically for network business flow with identical transfer path, aiming to enhance basic data transfer mode in terms of performance, security, multicasting, mobility and extensionality.

Most of the large-scale switching architecture at present relies on principles of IO Queue and Slot scheduling, whose bottleneck is central scheduling and waiting delay induced from queuing. Additionally, the support of multicasting is realized by dividing into several multiple unicasts, so hardware logic fanout cannot be achieved. As a result, WSC cannot be implemented on current switching architecture. In this article, we propose a solution–a novel two-stage, load-balanced, multipath, self-routing switching architecture. The first stage has a load-balanced function and the second one is designed for self-routing. The self-routing module is based on concentrator which can absorb traffic bursts of network, and Algebraic Switching Fabric (ASF), is characterized by fully distributed self-routing, no scheduling of port matching, no delay and jitter of buffer, group building positionally and recursive extension. This solution is implemented on Stratix IV FPGA platform from Altera Inc. and results of tests show that this architecture can achieve 100% payload, low delay, and no jitter. So this proposal supports WSC in theory.

The content followed includes the structure and function of WSC(part II), setting up switching mechanisms corresponsive to WSC(part III), building switching models supporting WSC based on switching mechanisms(part IV), and the summary.

2 The Functional Structure of Wide-Sense Circuit

WSC is builtdynamically for network business stream with identical transfer path to support Quality of Service (Qos) of data transfer. To achieve this, the structure is designed as a ‘management-control-data’ model. The management layer is responsible for the deployment of WSC, the control layer is responsible for execution of management part, and the data layer is responsible for WSC data transmitting. These three layers exchange information to realize the function of WSC as shown in Fig. 1.

2.1 The Management Plane

The management plane’s functionality, realized by domain’s server, consists of acquitting current network status, computing WSC’s location, and transmitting orders, such as setting up, adjusting and removing WSC. This layer serves as the core layer and management center of WSC.

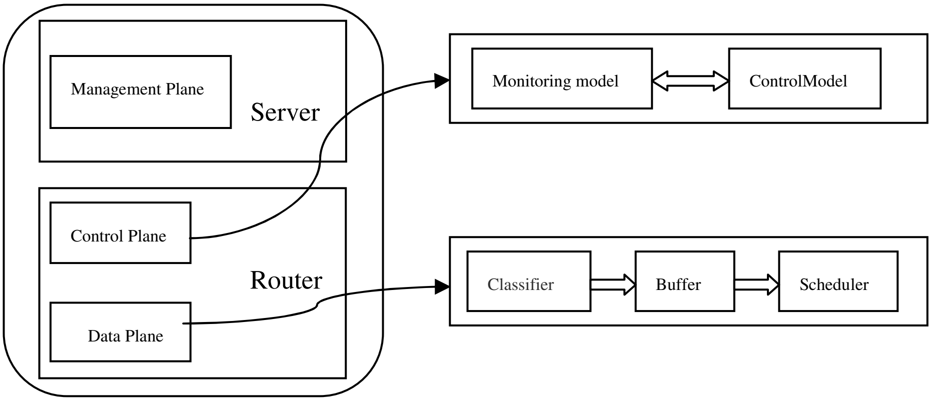


Fig. 1. The relationship of three planes of WSC

2.2 The Control Plane

This plane, realized in router, takes orders from the management layer and takes charge of execution, such as setting up, adjusting, and removing WSC according to WSC protocol. In this layer, monitoring module and control module are designed: the monitoring module collects information of flow and resource and reports it to the status acquisition module in the management layer; the control module executes orders from the management layer and communicates with other network WSC's control layer.

2.3 The Data Plane

Definition 1. Along the WSC, network set up a virtual circuit, which is called Wide-Sense Circuit Passway (WSCP).

This plane, realized in router, sets up WSCP and routes packets. When data is transmitted in WSC, WSCP is set up and corresponding labels are inserted into label switching table at entry point of WSCP. Afterwards, this layer will route packets and certify the QoS of different data flow simultaneously. This layer is the concrete key of realization of the functionality of WSC—for network nodes with built WSC, packets classifier, buffer and scheduler will react to the specific data flow accordingly to meet the requirements of data flow transmitting.

3 Switching Mechanism and Module for Wide-Sense Circuit

3.1 The Overall Structure of Reconfigurable Router

WSCis a new kind of data transmission mechanism, which guarantees the QoSaccording to the category of the traffic. Its working process includes traffic classification, establishment and adjustment of WSC and data forwarding three parts, which all happen in Reconfigurable Router. The overall architecture of Reconfigurable

Router is shown in Fig. 2. The flows are classified based on three key QoS indicators: delay, jitter and loss. The classification is mainly conducted by the classifier within Data Plane. The establishment and adjustment of WSC is under the decision of the server within Management Plane. And after making a decision, the work is completed by Control Plane in which the monitoring module is responsible for monitoring the bandwidth occupancy of various flows and the control module carries out the commands in detail. In Data Plane, if a WSC has been set up, there will be a private channel for it, calledWSCP, the place data switching proceeds. In addition, we need an advanced switching mechanism to ensure all types of QoS requirements. Next, we will carry on the detailed introduction of the three parts above-mentioned.

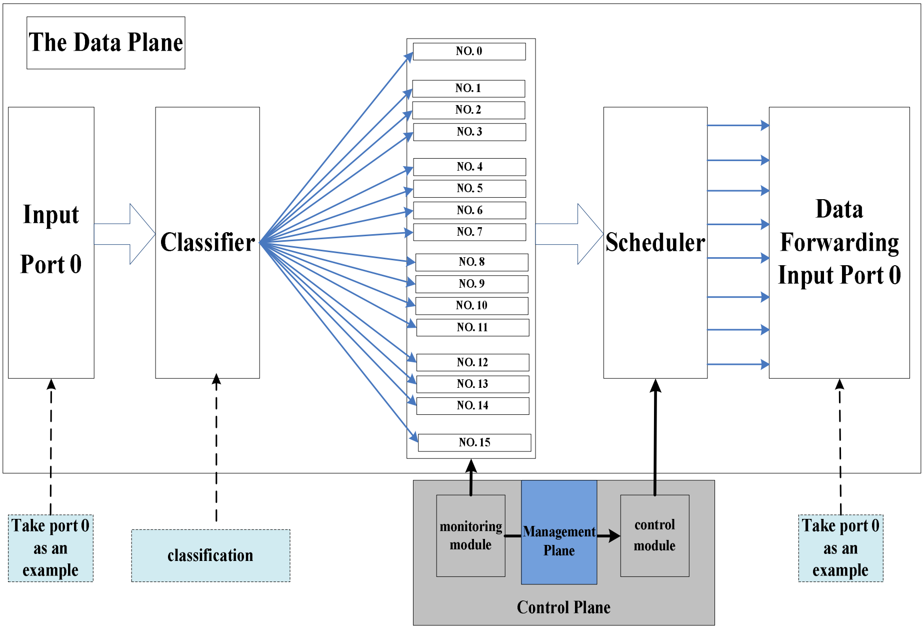


Fig. 2. The overall architecture of Reconfigurable Router

3.2 Traffic Classification

WSC sorts kinds of network flows in accordance with their different data transmission requirements, which consist of a series of QoS parameters. Here, to simplify our modeling, we just focus on delay, jitter and loss three indicators as the classification basis, and assume that boundary nodes have got all the relevant information.

Definition 2. The QoS requirements of data flow F is <delay, jitter, loss >.

Among them, delay indicates the time delay requirement, namely the time should be no more than delay (ms) when a packet is transferred from the source to its destination. Jitter defines the time jitter requirement, that is to say, the time jitter for a full packet must not exceed the maximum value: jitter (ms). Loss means the packet loss requirement, which limits the largest proportion of the discarded packets.

Next, we divide delay into [0, 100], (100, 400], (400, 1000] and (1000, ∞) four intervals. Jitter is divided into [0, 50] and (50, ∞) two intervals, and for loss: [0, 0.1%] and (0.1%, 1). In this way, we can use X, Y, and Z to represent the three QoS categories, and determine its specific number for each interval under a certain kind of category. The definite means are as follows:

1) If 0ms < delay ≤100ms, X = 1. If 100ms < delay ≤400ms, X=2. For 400ms < delay ≤1000ms, X=3. And if delay > 1000ms, X=4.

2) If 0ms < jitter ≤50ms, Y=1. If jitter > 50ms, Y=2.

3) For 0 ≤ loss ≤ 0.1%, Z = 1. And for loss > 0.1%, Z = 2.

Thus, we can get 16 types of flows according to the classification method: F_{XYZ}. And they are F_{111} F_{112}, F_{121}, F_{211}, F_{122}, F_{212}, F_{221}, F_{311}, F_{222}, F_{312}, F_{321}, F_{411}, F_{322}, F_{412}, F_{421} and F_{422}.

3.3 Establishment and Adjustment of Wide-Sense Circuit

According to the statistical analysis for different flows, the whole network establishes and adjusts WSC. Therefore the server, which is responsible for management, needs to collect each kind of traffic information on each link every once in a while to making the final decision. The basic principle is that when the link bandwidth occupied by one kind of network flow or other parameters have been above a certain threshold, the WSC is established on the link to guarantee the QoS requirements. Details are as the following three steps.

1) Calculating the link bandwidth utilization ratio UXYZ of flow FXYZ according to the traffic statistical information. This work is under the control by the monitoring module for real-time monitoring and the monitoring data will be transmitted to the server within the Management Plane.

2) The server should make the decision whether to establish WSC according to the setup criteria and related parameters. The basic idea of decision-making mechanism is that if the link bandwidth utilization ratio U_{XYZ} of flow F_{XYZ} has been above the threshold $U_{XYZ, th}$, just do it, otherwise, do nothing. For different flows, the value of $U_{XYZ, th}$ is not necessarily the same. Usually, for the business which requires high quality of service or high service priority, its $U_{XYZ, th}$ should be low, otherwise set high threshold, to achieve the purpose of distinguishing different service. To take an extreme example, setting $U_{111,th}$=$U_{112,th}$=$U_{211,th}$=$U_{212,th}$=0 and $U_{421,th}$=$U_{422,th}$=1, that is to say, we must establish WSC for flows F_{111}, F_{112}, F_{211}, F_{212} but not for flows F_{421}, F_{422}.Detailed scheme is in Table 1.

Table 1. Different values of $U_{XYZ, th}$ for different flows

NO.	0	1, 2, 3	4, 5, 6, 7	8, 9, 10, 11	12, 13, 14	15
flow	F_{111}	F_{112},F_{121},F_{211}	F_{122},F_{212}, F_{221},F_{311}	F_{222},F_{312}, F_{321},F_{411}	F_{322},F_{412}, F_{421}	F_{422}
threshold	3.4%	4.6%	5.7%	6.8%	7.9%	9.1%

We can see that threshold priority is divided into 3.4%, 4.6%, 5.7%, 6.8%, 7.9% and 9.1% six levels and the same threshold value can be used by some different types of flows, in which the smaller the label NO., the higher the internal priority. Firstly, we decide to set up WSC or not for a certain kind of network flow by comparing its link bandwidth utilization ratio with its threshold. After completing the establishment of all the qualified WSC, if the bandwidth resources are still remaining, we could use the rest flows to fill the bandwidth in the order of label numbers.

3) The control module within control plane performs the decisions such as establish or dismantle WSC, which is responsible for establishing WSCP, etc. When WSC deployment is completed, this WSC can provide sufficient bandwidth for the corresponding traffic. And to ensure kinds of QoS requirements, matched scheduling algorithm and discard algorithm are also be used for distinguishing or configuring.

3.4 Data Forwarding

WSC realizes virtual circuit connection with the method of label switching and packets are forwarded within WSCP. WSC supports Multiple Input Multiple Output (MIMO), therefore, packets can enter into WSC as long as there is a node within the coverage area of the WSC. Similarly, packets can leave in any node.

In WSC ingress node,when a network flow arrives, the node will retrieve the label switching table according to the category information and destination address of packets. In WSCintermediate node, the node will retrieve the label switching table again, but according to the message header of WSC this time, and then forwards the matched packets.In WSC exit node, the node will drop the message header and forward the data as common packets.

4 Switching Mechanism and Module for Data Forwarding

4.1 2×2 Basic Sorting Unit

The 2×2basic sorting unit is a sequential logiccircuit, with two inputs and two outputs (respectively called 0/1 port).According to the theory of algebraic distributive lattices [3], we define the two inputs as Ω_0 and Ω_1, each of which has three kinds of data: the one going to output0, the one going to output1 and the invalid data. As is shown in Fig. 3(a) and (b), the sorting unit has two essential states: Cross and Bar. That means the inputs go to the different outputs: input0/input1 to output1/output0 and input0/input1 to output0/output1, corresponding to Cross and Bar,respectively.If the inputs compete for the same one output, the state will be Conflict and the final choice of BAR or CROSS will depend on their priority, shown in Fig. 3(c).

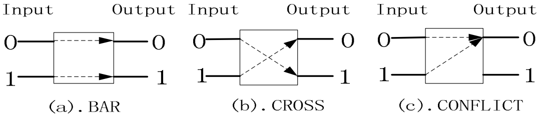


Fig. 3. 2×2 basic sorting unit and its states

4.2 Multipath Self-routing Switching Structure

An N×N ($N=2^n$) routing network is a Multistage Interconnection Network (MIN) built by 2×2 basic sorting units. By using first stage permutation σ_0, inter-stage permutation σ_1, $\sigma_2 \ldots \sigma_{(n-1)}$ and last stage permutation σ_n, the network can be represented as $[\sigma_0$: σ_1: σ_2:…: $\sigma_{(n-1)}$:$\sigma_n]$. Each colon symbolizes a stage of2x2 units.We can define a Trace sequence and a Guide sequence[4] as follows:

$$T_k=(\sigma_0\sigma_1\ldots\sigma_{K-1})^{(-1)}(n) \quad 1\leq k\leq n. \tag{1}$$

$$G_k=(\sigma_0\sigma_1\ldots\sigma_{K-1})(n) \quad 1\leq k\leq n. \tag{2}$$

Route is specified by Trace orGuide. AsFig. 2 shows,for the network [: (43): (42)(31): (43):], data from the origination address$I_1I_2I_3I_4$ finally gets to the destination$O_1O_2O_3O_4$ with the decision at each stage by the Trace (4, 3, 2, 1) or Guide (1, 2, 3, 4).

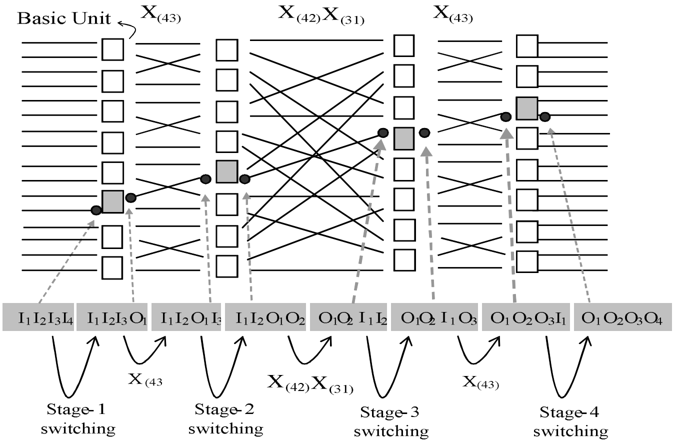


Fig. 4. An example of routing network

Multipath Self-routing Switching Structure (MSSS)[5]is an innovative structure, which combines MIN with concentrators.

To construct MSSS,we substitute each basic for 2G-to-G concentrator sorting unitand replace the single cable with a bundle of G cables. For the multipath structure (N=128 M=16 and G=8)which is based on a 16×16 routingnetwork, G shows the size of group, M is the number of group and N=M×G indicates the whole number of input/output ports ($G=2^{g,}$ $M=2^m$, $N=2^n$, n=m+g, n, m, g are positive integers).Acting as anindispensable part of MSSS, the 2G-to-Gconcentrator [6]separatesthelargerGsignals of the whole 2G inputs fromthe other G signals,finally forming two output groups. And the output orderwithin each group is arbitrary.

4.3 Load-Balanced Self-routing Switching System

As shown in Fig. 5,two MSSSs are used in series to compose the Load-BalancedMultipath Self-routing Switching Structure (Load-Balanced MSSS), with the VOGQs (Virtual Output Group Queues)[7]ahead of the firstfabric and the assemblages at the end of the second fabric. Actually, by using simple algorithms and small buffers,the first stage fabric serves as a balancer, which spreadsthe incoming traffic to all the ingress ports of the second stage fabric. Then the second stage fabric forwards the data in a self-routing manner to their final destinations. Every G inputs/outputs are bundled into an input/output group, thus N input lines form M groups on the input side(N=M×G), so is the output side. To ease presentation, IG/OG denotes input/output group, and MGrepresents a line group between the two stages.In the project, there are 4 IGs, 4 MGs and 4 OGs. Each group has 8 lines. .

VOGQs are responsible for storing packets and making data ready for IGs. We use VOGQ(i,j) to denote the VOGQ whose packets are destined forOGj from IGi.

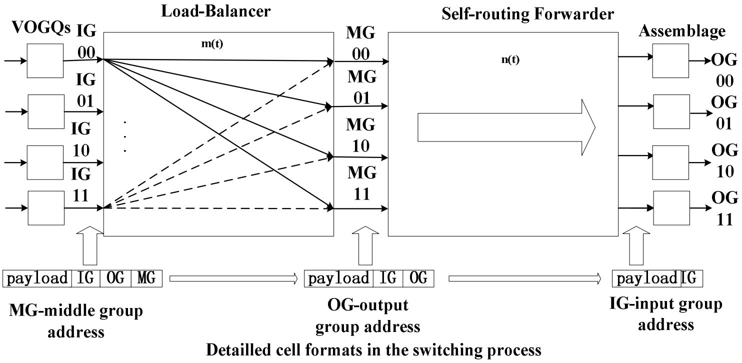


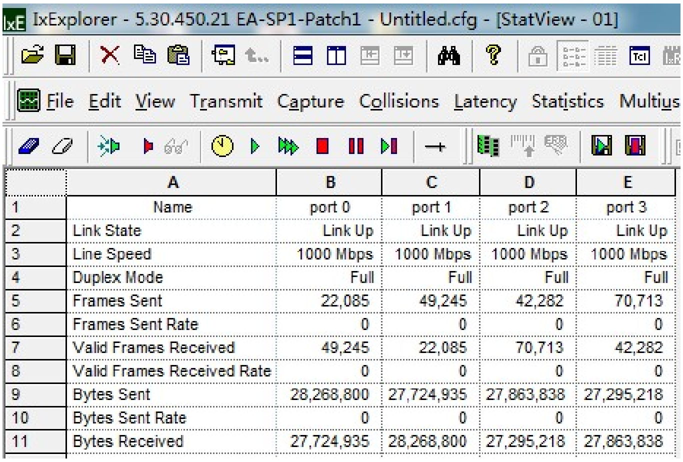
Fig. 5. Load-Balanced MSSS

Generally, for the structure we proposed, the processing of arriving packets in each time slot is composed by severalsequential phases which are shown as follows. In addition, to achieve maximum processing speed, we should use pipeline structure as far as possible.

1. *Preparatoryphase*: With checking and judging, the arriving packet which is destined forOGj from IGi is stored into VOGQ (i,j).
2. *Splitting phase*: Packets in VOGQs are split into cells. And each cell will be added with some certain packet headers.
3. *Balancing phase*: With the help of MG tags, cells will be routed to every middle group simultaneously and uniformly. When the cells reach the middle groups, MG tags will be dropped.
4. *Routing phase*: Cells are further to their final destinations directed by OG tags. When they get through the second stage fabric, OG tags will be discarded.
5. *Assembling phase*: Cells are to be assembled to originalpackets.When completed, packets will be output from the OGs.

4.4 System Test with Real Network Traffic

IXIA400T network tester is our leading network test instrument. We use four test modules of all the interfaces on the test board, which can generate or capture standard Ethernet frames transmitted at the rate of 10/100/1000 Mb/s. It is so powerful that we can set, if we want to, every Byte of a frame to be sent and get detailed and comprehensive information about the frames captured. The tester also provides remote management capabilities. And coupled with the automated platformset up by Tclscripting language, we can implement remote automated testing.


IxExplorer - 5.30.450.21 EA-SP1-Patch1 - Untitled.cfg - [StatView - 01]
File
Edit
View
Transmit
Capture
Collisions
Latency
Statistics
Multius

	A	B	C	D	E
1	Name	port 0	port 1	port 2	port 3
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	22,085	49,245	42,282	70,713
6	Frames Sent Rate	0	0	0	0
7	Valid Frames Received	49,245	22,085	70,713	42,282
8	Valid Frames Received Rate	0	0	0	0
9	Bytes Sent	28,268,800	27,724,935	27,863,838	27,295,218
10	Bytes Sent Rate	0	0	0	0
11	Bytes Received	27,724,935	28,268,800	27,295,218	27,863,838

Fig. 6. Statistic views of IXIA

Fig. 6 shows us the finalstatistical result of a test for four ports. According to our configuration, port 0 prepares to receive the output data sent form all the four ports (including itself). And port2, port3 and port4 follow the same way. We can see that there is no one Byte data dropped at each port in the case of a large number of input data.

5 Conclusions

Multipath Self-routing Switching Mechanism (MSSM) belongs to the physical vision of packet transmission in AC, which is also the foundation of WSC. It is a high quality basic network structure inherited from the existing network system, directly providing the network survivability and robustness. Above all it provides the material foundation for the construction of WSC and the reconfiguration. Based on the powerful transmittability of MSSM, WSC can also possess the new information foundation of the communication network and the new connotation of network interconnection transmission capacity. These enhanced foundational capabilities, especially for data transfer mode, ensure the lower delay of packet transmission, wire-speed forwarding, efficient multicast and security. MSSM gives effective technical support to build a new type of WSC, so as to make WSC be the supporting factor of reconfigurable ability, together with the packet transmission.

The research in this direction goes well: we have accomplished the single-stage switching system which supports unicast and multicast and the two-stage load balancing switching system which supports unicast only.Good results have been achieved in a series of tests. So far, our system is based on the MSSS (M=4, G=8). And next step, we plan to increase it to M=8, G=16. Meanwhile, the design of large-scale wire-speed multicast base on Load-Balanced MSSS we constructed will be the focus, which needs more excellent design and more thorough support system.

Acknowledgment. Our work is supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB315904, the National Natural Science Foundation of Chinaunder Grant61179028, the Natural Science Foundation of Guangdong Provinceunder Grant201101000923 and 2013020012822, the Basic Research of Shenzhenunder Grant201104210120A and 20130331144502026.We also acknowledge the valuable feedback from Le Yang and Qing Ma (Depaul University, Chicago, USA) for helping me improve my Englishduring the preparation of this paper.

References

1. Lan, J., Xinget, C., et al.: Reconfiguration Technology and Future Network Architecture. Telecommunications Science, 10.3969/j.issn.1000-0801.2013.08.003
2. The 2013 Annual Report of NationalBasic Research Program of China (973 Program, No. 2012CB315904), Zhenzhou
3. Nojima, S., et al.: Integrated services packet network using bus matrix switch. IEEE J. of Select Areas Commun. 5, 1284–1292 (1987)
4. Li, S.Y.R.: Algebraic switching theory and broadband applications. Academic Press (2001)
5. Li, H., He, W., Chen, X., Yi, P., Wang, B.: Multi-path Self-routing Switching Structure by Interconnection of Multistage Sorting Concentrators. In: IEEE CHINACOM 2007, Shanghai (August 2007)
6. Li, S.Y.R.: Algebraic Switching Theory and Broadband Applications. Academic Press (2001)
7. He, W., Li, H., Wang, B., et al.: A Load-Balanced Multipath Self-routing Switching Structure by Concentrators. In: IEEE ICC (2008)