The Axon Network Device: Prototyping with NetFPGA

Jeffrey Shafer, Mike Foss, Scott Rixner, Alan L. Cox

Rice University
Houston, TX
{shafer, mikefoss, rixner, alc}@rice.edu

Abstract

This paper introduces the *Axon*, a network device that employs *source-routed Ethernet* to improve the scalability of local-area networks. Axons provide transparent compatibility with unmodified hosts, which see the network as a single switched Ethernet segment. Prototypes of the Axon have been implemented on the NetFPGA platform and used to demonstrate the performance and backwards compatibility of source-routed Ethernet. Future work will combine prototype Axons with simulated Axons to evaluate network scalability for up to one million hosts.

1 Introduction

The design of datacenter and enterprise networks poses many challenges today, due to the shortcomings of existing devices such as Ethernet switches and IP routers. Such devices are limited in terms of scalability [3], support for virtualization [4], ease of administration, and security [1]. We developed a new network device – the *Axon* – and a new *source-routed Ethernet* protocol to address these challenges.

Axons replace Ethernet switches in a network and relegate IP routers to the network's edge for connectivity to other networks. Unmodified host systems communicate with Axons via traditional Ethernet frames and see the network as a single switched Ethernet segment. Internally, however, Axons communicate with other Axons via a novel source-routing datalink protocol over a standard Ethernet physical layer. Compared to distributed routing, employed by switches and IP routers, source routing provides a more scalable networking solution because routes are stored only at the edge of the network. The flexibility provided by source routing also enables network virtualization and security at a finer granularity than VLANs. Simply removing a route can deny access to a host, and modifying a route will cause traffic to take a different path over the physical network.

We have prototyped the Axon design using an FPGA to perform packet switching and an Atom processor to control switch operation. The prototype allows us to explore the interactions between Axon network devices and complex real-world hosts running a variety of operating systems. Although the Axon architecture is designed to be fundamentally scalable, we intend to fully test this by combining the Axon prototype with a simulator infrastructure to evaluate networks with up to one million hosts.

2 Axon Prototype

The Axon prototype was constructed from a NetFPGA PCI card and an Intel Atom processor in a mini-ITX motherboard, as shown in Figure 1. The Stanford NetFPGA project provides an integrated hardware/software environment for research and education in network systems architecture [2]. By combining NetFPGA with a low-power Atom processor, we were able to construct a highly capable, low cost prototyping platform.

The Axon prototype is divided into a *data plane* that does high speed packet forwarding in the FPGA, and a *control plane* that runs infrequent and/or complex management tasks in software. The data plane is built on the NetFPGA PCI card with 4 Gigabit Ethernet ports, a Virtex-II Pro 50 FPGA, several memories, and a 32-bit/33 MHz PCI interface to the control plane. Many pieces were reused from the NetFPGA project, including the PCI card, pieces of the Verilog design library (such as PCI DMA transfer engine), control utilities to program the FPGA, and ModelSim test suite. New components comprising 14,000 lines of Verilog were added, including a cut-through packet switching engine and modules to transparently convert between traditional and source-routed Ethernet frames.

The Axon control plane is executed on the Intel Atom x86 processor running Linux. A new 6,000 line control program uses the device driver provided by the NetFPGA project to directly manage the hardware data plane. In the

prototype, only management tasks – such as determining network topology and establishing source routes – run in software; packet forwarding is done purely in hardware.

Using software and hardware components from the NetFPGA project provided advantages in terms of design time and platform compatibility. The effective reuse of the Verilog design library accelerated implementation, and the ModelSim test framework improved the robustness of the prototype. The proven hardware allowed the prototype to be quickly developed in 3 months by 2 full-time graduate students, versus many months to design and fabricate a board from scratch. In addition, the NetFPGA project utilizes a host computer running Linux on an x86 processor, which simplifies development of the control plane software and ensures its portability. We will exploit this in future simulation work.

While reusing NetFPGA components provided many advantages, the board itself limited the prototype in several ways. First, the hardware limited the Axon prototype to only 4 Gigabit Ethernet ports – compared to 24-48 for a typical network switch – which restricts the network topologies that can be constructed. Second, the Virtex-II Pro 50 FPGA has limited resources. The current Axon design consumes 76% of the FPGA slices, which restricts future expansion. Finally, while conventional Ethernet switches use a Content-Addressable Memory (CAM), the Axon prototype does not, because a large CAM would not fit into any FPGA. Using a CAM small enough to fit on the FPGA would severely limit the number of hosts reachable on the experimental network and restrict the use of the prototype for research on scalability. Instead, a MAC address rewriting scheme was devised to store source routes in SRAM and access them via a direct lookup. While this improves scalability, it also increased the complexity of the prototype.

Constructing experimental prototypes enabled us to demonstrate that the Axon architecture provides full compatibility with unmodified hosts, and that it is possible to translate from traditional Ethernet to source-routed Ethernet at full network speed. In addition, the prototypes allowed us to do head-to-head comparisons between Axons, Ethernet switches, and IP routers, all while using real network hosts, using the test network shown in Figure 3.

3 Network Scalability

The current Axon prototype demonstrates the feasibility of transparent source-routed Ethernet. However, it would be unrealistic to build enough prototypes to prove the scalability of an Axon network, which we envision to handle as many as one million hosts. Instead, we plan to build a hybrid prototype/simulator infrastructure in which physical hosts are connected to physical Axon prototypes, which in turn are interconnected by multiple simulation servers. (See Figure 4.) Each simulation server will be responsible for emulating thousands of virtual Axons and network links.

The nature of the Axon architecture leads naturally to this hybrid approach. State is stored at the edge of the Axon network, not at the center. Thus, the metrics important to proving scalability – such as the number of bytes sent and received by the control software to converge to a routing path – can be measured directly on the prototype devices without requiring timing information from the simulator. Instead, only a simple functional simulator is needed to model the Axon packet forwarding, monitor link utilization and congestion, and execute control software for virtual Axons. The same control plane software that runs on the physical Axon will be run in multiple threads on the simulation server, ensuring simulator fidelity because the prototype Axons will interact with real control software. This benefit is made possible because the prototype control plane software was implemented in Linux running on an x86 processor.

4 Conclusions and Future Work

We developed source-routed Ethernet and a prototype Axon network device to implement it. This prototype demonstrates that high performance can be provided along with compatibility for existing network hosts. The use of the NetFPGA project greatly accelerated prototype implementation and testing. In the future, prototype Axons will be used to evaluate improvements in congestion control, virtual machine migration, and virtual networks made possible by the source routing architecture. Although experimental prototypes do not immediately allow us to explore how well the Axon architecture scales to large networks, we believe that this challenge can be overcome by coupling prototype Axons with simulated Axons.

References

[1] Casado et al. Ethane: taking control of the enterprise. In *SIGCOMM*, 2007.

[2] Lockwood et al. NetFPGA - an open platform for gigabit-rate network switching and routing. In *MSE*, 2007.

[3] Myers et al. Rethinking the service model: Scaling Ethernet to a million nodes. In *HotNets*, 2004.

[4] Wang et al. Virtual routers on the move: live router migration as a network-management primitive. In *SIGCOMM*, 2008.

5 Appendix

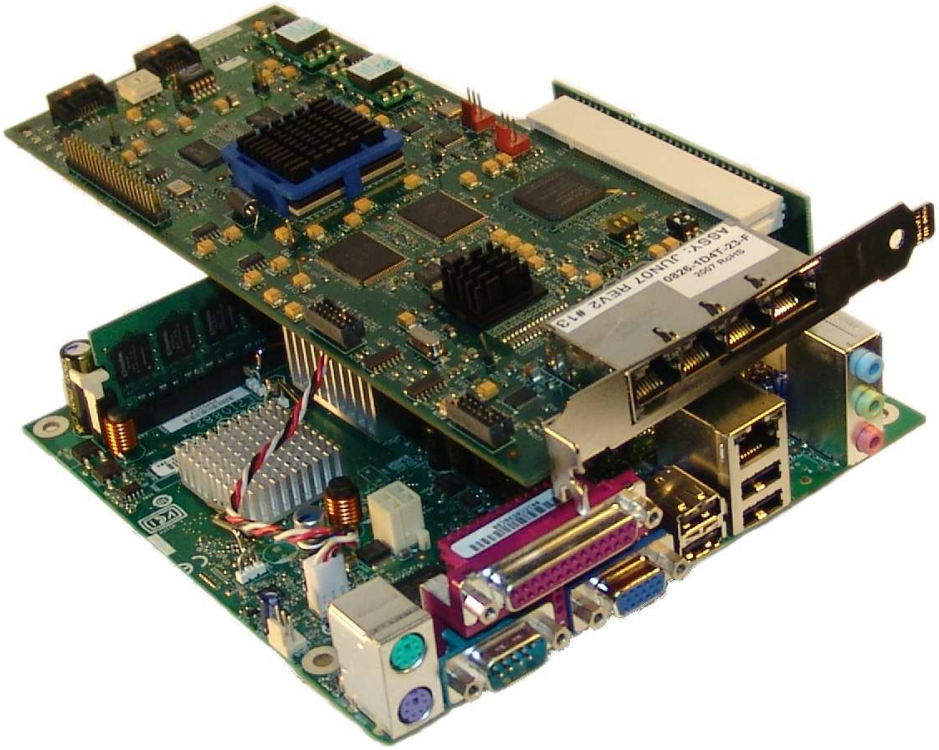

Figure 1: Prototype of the Axon Ethernet Device

Figure 1 shows the Axon prototype, which is comprised of an Intel Atom D945GCLF motherboard with a NetFPGA PCI card. Prominent features of the NetFPGA (on top) are its four gigabit ports, PCI connector, and Virtex-II Pro 50 FPGA. Below the NetFPGA card is an Intel Atom motherboard that powers the control plane of the Axon and performs tasks such as source route creation and management. Because the control plane tasks are lightweight and infrequent, a low-powered processor such as the Atom is more than powerful enough to drive them.

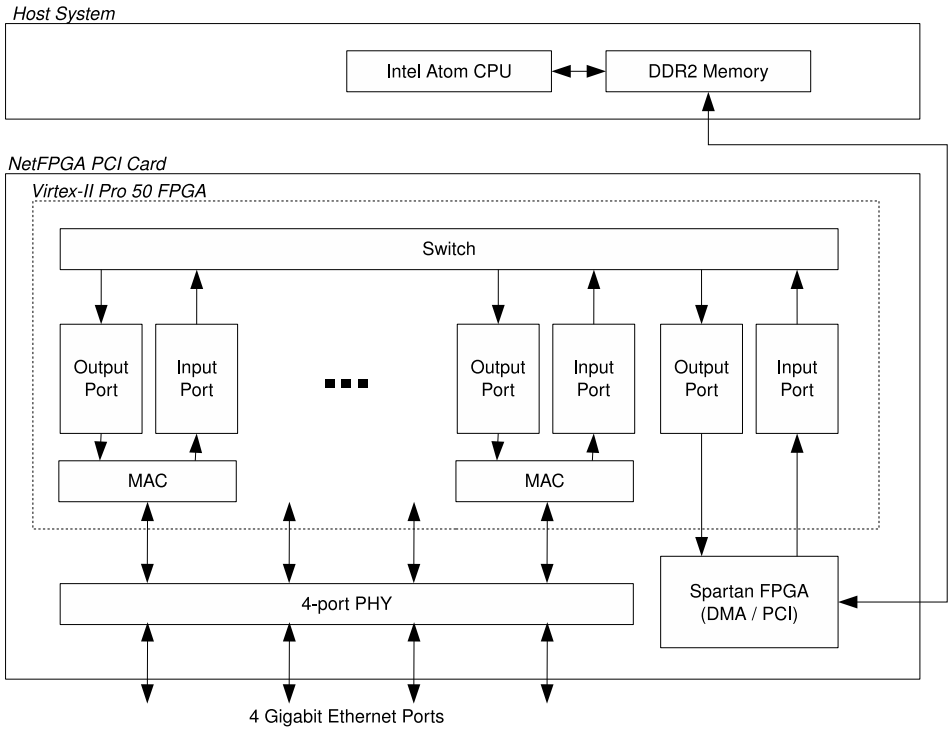

Figure 2: Axon Prototype Architecture

Figure 2 shows the architecture of the Axon prototype. The Virtex-II Pro 50 FPGA on the NetFPGA card implements the primary functions of the device, including the 4-port Ethernet MAC and switching logic. Also implemented in hardware are the input and output port units responsible for decoding the source-routed Ethernet datalink layer and providing transparent compatibility with traditional Ethernet. A Spartan FPGA provides the PCI interface and DMA engine for data transfer with the host system, in this case a motherboard with an Intel Atom processor.

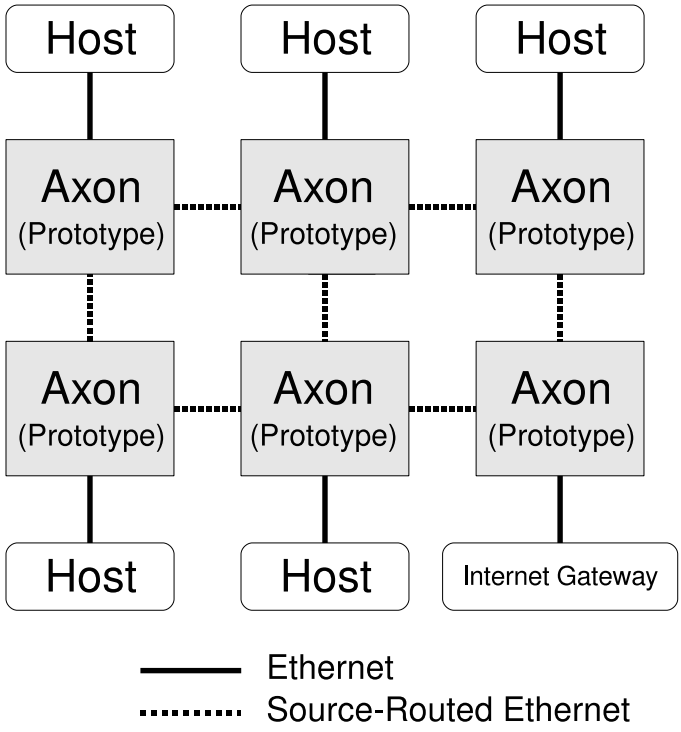

Figure 3: Experimental Testbed using Prototype Axons

Figure 3 shows an experimental testbed. This small-scale network is composed of 6 Axons in a mesh, along with 5 end hosts and a gateway router that links to the Rice University campus network and, by extension, the public Internet. Hosts communicate with Axons using standard Ethernet, and Axons communicate with other Axons using source-routed Ethernet. Axons can function as DHCP servers to their directly-attached hosts and can exchange traffic with the campus network (via the gateway router) if desired. Using this network, tests were performed to validate various aspects of the Axon design and to demonstrate its performance characteristics with regards to latency, bandwidth, and fairness.

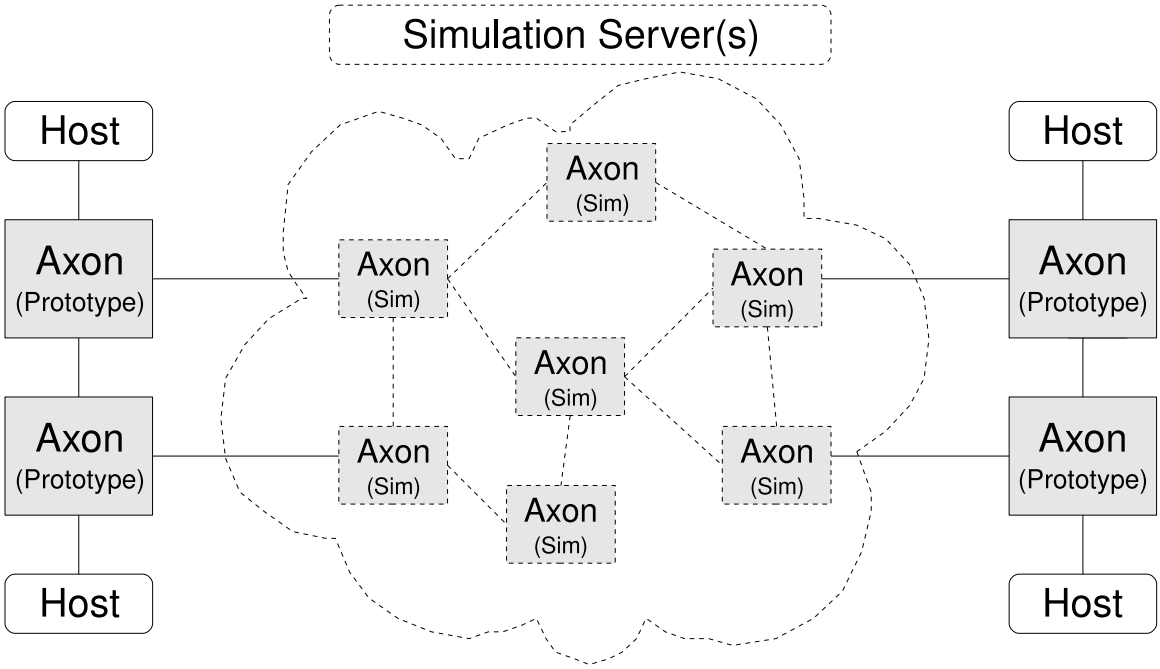

Figure 4: Experimental Testbed using Prototype and Simulated Axons

Figure 4 shows a proposed hybrid testbed composed both of physical Axon prototypes and simulated Axons that run in software on a simulation server.