RiceNIC: Prototyping Network Interfaces

Jeffrey Shafer and Scott Rixner
Rice University
Houston, TX
{shafer,rixner}@rice.edu

ABSTRACT

RiceNIC is a reconfigurable and programmable Gigabit Ethernet network interface card (NIC). It was created as an open platform for public use and is freely available for research and education. It can be used to prototype new network server architectures, and has proved invaluable in recent research efforts. Using a commercial development board saved significant time and expense compared with custom fabrication. In addition, using a modern FPGA optimized for system-on-a-chip applications allowed the NIC to meet its performance goals with minimal effort.

1. RICENIC

We have created RiceNIC, a reconfigurable and programmable Gigabit Ethernet NIC. RiceNIC is freely available and provides researchers with a flexible baseline network interface that can be easily modified to support a wide range of research and educational activities, providing a significant head-start compared with implementing an experimental prototype from scratch [2].

RiceNIC has been used in several network research projects both at Rice University and at other institutions. For example, RiceNIC hardware and software was modified to support a new network virtualization technique called Concurrent Direct Network Access (CDNA). This allows multiple virtual machines running on the same system to concurrently control a single NIC, improving the efficiency of network virtualization [3]. RiceNIC's flexible design was easily modified, and the final implementation ran at full network speed, allowing highly accurate performance measurements to be taken that strongly demonstrated the benefits of the new network virtualization architecture.

The CDNA prototype using RiceNIC proved very useful in experimentally determining key architectural features, such as the minimum size of on-NIC packet buffers per virtual machine to fully saturate the Ethernet link. This was an important contribution of the published research, because it was found that the per-OS buffer requirements of 384 KB were well within the capability of a modern NIC and could be implemented at a low cost. Estimating the buffer size in a virtualized system is difficult, because buffer requirements are directly linked to non-obvious scheduling decisions by the virtual machine monitor that determine how frequently each guest operating system can execute. The less frequently an OS runs, the more buffering it needs to maintain network throughput. With the RiceNIC prototype, a series of focused experiments on a heavily-loaded virtualized system were conducted at a wide range of buffer sizes in order to determine the best size for optimal performance. Performing these experiments in simulation would have been difficult because saturating the buffers required tens of seconds of execution time. Most simulations only model a fraction of a second of real time in order to reduce the length of the simulation. This research effort shows both the value of having an experimental prototype and the value of using RiceNIC as a design platform.

RiceNIC is built upon the commercial Avnet Virtex-II Pro development card and provides a custom FPGA design to make all of its raw components function as a system. The NIC includes multiple FPGAs, large on-NIC memories in excess of 256 MB, two 300 MHz embedded PowerPC processors, and a copper Gigabit Ethernet interface. Custom NIC firmware is provided along with a Linux device driver. Using only a single PowerPC processor, RiceNIC can saturate the Gigabit Ethernet network link with maximum-sized packets. This leaves significant resources—including 50% of the reconfigurable logic elements on the Virtex-II Pro FPGA, a spare PowerPC processor, and hundreds of megabytes of memory—available on RiceNIC to use for experimental networking research [1].

2. LESSONS LEARNED

Several lessons can be drawn from the construction and use of RiceNIC in research programs that are valuable to researchers contemplating similar efforts.

2.1 Commercial Development Board

Using the commercial Avnet development card directly contributed to the success of this project. The Avnet card is a solid foundation for a high performance NIC because it contains a large FPGA, PCI bus, Gigabit Ethernet PHY, and on-chip memory resources. Purchasing a board accelerated the project since it arrived fully-tested and documented and thus enabled RiceNIC implementation to begin immediately. In addition, it reduced the up-front project expenses in comparison with designing and producing a custom platform. For a low-volume research product such as RiceNIC, it might never be cheaper to fabricate a custom board.

2.2 FPGA Capabilities

The Virtex-II Pro 30 FPGA provided on the Avnet development card is targeted for system-on-a-chip applications such as RiceNIC. This FPGA includes two embedded PowerPC

processors, on-chip memory, and digital clock managers for multiple clock domains, and these features were extensively used. The amount of reconfigurable logic on the chip was sufficient for RiceNIC and newer FPGAs should have even greater resources. Using an FPGA obviates the need for a custom-designed ASIC, which would have been cost and time-prohibitive for the RiceNIC project.

The key reason the Xilinx FPGA performed so well in the RiceNIC, however, was because it was able to meet the performance demands of the system with only minimal effort. A 100MHz clock for the memory and data buses and a 300MHz clock for the embedded processors was sufficient to provide a full Gigabit per second of network throughput for the NIC, as well as provide headroom for the implementation of new network architectures such as CDNA. These clock frequencies are well within the capabilities of modern FPGAs.

In terms of modern network hardware, a 1-Gigabit NIC is no longer cutting-edge. This certainly made it easier to develop RiceNIC in a timely fashion using commercial tools. The use of a commercial development board or FPGA, however, would not preclude the development of a 10-Gigabit version of RiceNIC, as newer development cards from Avnet feature larger FPGAs, 10-Gigabit copper network ports, and utilize PCI Express. Thus, much of the RiceNIC development methodology would transfer easily. One possible exception is that while the 300MHz PowerPC processors on RiceNIC could, with optimization, support higher than 1-Gigabit data rates, they would be unable to achieve 10-Gigabit data rates. Since newer FPGAs do not include significantly faster processors, a 10-Gigabit FPGA NIC would need to be implemented largely in the reconfigurable logic arrays.

2.3 Design Time

It took a single graduate student slightly more than 1 year to design and implement the RiceNIC hardware. In addition, it took about 1 month for a second student to write the software, including the device driver and NIC firmware.

This development time can be divided into three stages. First, at least one month was spent learning how to use the Xilinx ISE and EDK tools effectively. However, any development tool would require time to properly learn, and the Xilinx software is no better or worse in that regard. Second, the actual design and initial implementation took about 9 months of productive design time. Finally, testing and bug fixes took at least 3 months at the end of the project.

The PCI core in particular required extensive iterative testing because it must work correctly on a wide variety of host computer systems with non-deterministic bus transfer and error timing characteristics. This type of problem would not exist in a software simulator that ignores these low-level issues. A working prototype, however, must function perfectly at even the most detailed level in order to produce useful research results. The redeeming grace of dedicating this effort to fixing low-level bugs is the confidence that it lends to any experimental research. RiceNIC is a fully-functional NIC and is accurate at even the lowest-levels, so no simulation approximations are present that would dilute the experimental results. The RiceNIC implementation time was not prohibitive, and the benefits are well worth the cost.

3. CONCLUSIONS

Perhaps the most important question related to RiceNIC is: would we build it again, knowing what we now know? The answer is an unqualified yes. The effort to build a reconfigurable and programmable NIC was successful and produced a working tool suitable for research and education uses. RiceNIC has already been used in several research projects, especially involving virtualized systems, and has enabled the collection of invaluable experimental results on actual hardware that would have been impossible to achieve with any other network interface. Given the challenges associated with network server architecture simulation, experimental research involving the network system is not only possible, but is essential. The complex, asynchronous interactions among system components demand high performance prototyping for accurate experimental research. RiceNIC, and other tools like it, are critical for the development and understanding of future computer systems.

All of the custom design elements of the RiceNIC platform can be freely downloaded for research and education uses. This includes the FPGA configuration, VHDL source code, embedded PowerPC firmware, and Linux device driver. However, the Avnet Virtex-II Pro development board with a Virtex II Pro 30 FPGA must be obtained from Avnet Electronics Marketing, along with other inexpensive accessories such as a JTAG programming cable, compact flash card reader, and larger SDRAM SODIMM. In addition to the hardware, Xilinx development software such as ISE, EDK, and Chipscope are useful for modifying the FPGA configuration. Finally, hardware customization requires new licenses for the Xilinx MAC and PCI cores. The terms of the Xilinx licenses restrict distribution of the IP source code or netlist, and thus these files cannot be provided with the rest of the public RiceNIC source code.

To learn more about RiceNIC, please visit:
http://www.cs.rice.edu/CS/Architecture/ricenic/

4. ACKNOWLEDGMENTS

This project is supported by gifts from Advanced Micro Devices and Xilinx, and grants from the Los Alamos Computer Science Institute and the National Science Foundation under grant No. CCF-0546140.

5. REFERENCES

[1] J. Shafer and S. Rixner. A reconfigurable and programmable gigabit ethernet network interface card. Technical Report TREE0611, Rice University, Dec. 2006.

[2] J. Shafer and S. Rixner. RiceNIC: A reconfigurable network interface for experimental research and education. In *Proceedings of the Workshop on Experimental Computer Science (ExpCS'07)*, June 2007.

[3] P. Willmann, J. Shafer, D. Carr, A. Menon, S. Rixner, A. L. Cox, and W. Zwaenepoel. Concurrent direct network access for virtual machine monitors. In *Proceedings of the International Symposium on High-Performance Computer Architecture (HPCA'07)*, Feb. 2007.