AFRL-RI-RS-TR-2013-169

ENGAGING STUDENTS VIA INNOVATIVE MILITARILY USEFUL TECHNOLOGIES

CORNELL UNIVERSITY

JULY 2013

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2013-169 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/ S /

RICHARD FEDORS
Work Unit Manager

/ S /

JULIE BRICHACEK, Chief
Information Systems Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
JULY 2013	FINAL TECHNICAL REPORT	NOV 2011 – MAR 2013

4. TITLE AND SUBTITLE
ENGAGING STUDENTS VIA INNOVATIVE MILITARILY USEFUL TECHNOLOGIES

5a. CONTRACT NUMBER FA8750-11-1-0247
5b. GRANT NUMBER N/A
5c. PROGRAM ELEMENT NUMBER 62788F

6. AUTHOR(S)
Kenneth Birman, Robbert van Renesse and Hakim Weatherspoon

5d. PROJECT NUMBER S2UN
5e. TASK NUMBER EX
5f. WORK UNIT NUMBER PL

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cornell University
4230 Upson Hall
Ithaca, NY 14853

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/RISC
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2013-169

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. PA# 88ABW-2013-3174
Date Cleared: 10 JUL 2013

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Faculty and researchers in the Department of Computer Science at Cornell University have been studying the challenges of information management and high-assurance computing for a variety of demanding settings, including cloud computing and data transmission in complex network environments. The main objective of the project titled “Engaging Students via Innovative Militarily Useful Technologies” was to expose students to the research and use of these technologies which have potential application in military scenarios and in nationally important critical infrastructure areas. By exposing students to this type of research, we can come closer to realizing the potential application of these tools, and the students themselves benefit from the experience in many ways. Historically, a significant percentage of Cornell students have pursued careers in the military, with military contractors, and in the civilian critical infrastructure areas. During the course of this program students were involved in four main areas: Scalable Landmark Recognition (ePaparazzi), the Software-defined Network Interface Card (SoNIC), Smart Grid Security (GridControl), and additional projects relating to specific aspects of the Isis2 software platform.

15. SUBJECT TERMS
Information management, high-assurance computing, cloud computing, landmark recognition, network interface card, smart power grid, Isis2 platform, student research

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					RICHARD FEDORS
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPONE NUMBER (*Include area code*)
U	U	U	UU	29	**N/A**

Table of Contents

List of Figures ii
List of Tables ii
1.0 SUMMARY 1
2.0 INTRODUCTION 2
3.0 METHODS, ASSUMPTIONS, AND PROCEDURES 9
3.1 Scalable Landmark Recognition 10
3.2 Software-defined Network Interface Card 12
3.3 Smart Grid Security 14
3.4 Isis2 Projects 15
4.0 RESULTS AND DISCUSSION 16
5.0 CONCLUSIONS AND RECOMMENDATIONS 18
6.0 REFERENCES 19
APPENDIX 20
BOOM 2012 Projects 20
BOOM 2013 Projects 23
LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS 24

List of Figures

Figure 1. ePaparazzi Recognized Photos 4
Figure 2. SoNIC: Software-defined Network Inferface Card 5
Figure 3. Representation of the GridCloud Platform 7
Figure 4. Isis2 and its Interactive Data Analysis Layer, the IDA 8
Figure 5. Organizational Structure of Projects 10

List of Tables

Table 1. Student Research Briefings 3
Table 2. Proposed Tasks, Schedule, and Milestones 9

1.0 SUMMARY

The main objective of the project titled *Engaging Students via Innovative Militarily Useful Technologies* was to expose students to computing technologies that include a range of solutions (some purely commercial, some developed at Cornell University, and some including a mixture of technologies) with potential for military or critical-infrastructure use. Faculty and researchers in the Department of Computer Science at Cornell have been studying the challenges of information management and high-assurance computing in a variety of demanding settings, including cloud computing, data transmission in complex network environments, and the exploration of the underlying scientific basis for high-assurance computing. These research efforts have resulted in the development of many useful technologies, such as sophisticated application development tools and platforms, advanced security frameworks and tools, and purely conceptual tools such as new theories of high-assurance that respond to the most stringent requirements. In particular, there has been recent focus on cloud computing and the creation of highly resilient, secure, cloud-hosted services.

Often the research is guided by thinking about important military scenarios, and there is potential for considerable positive impact in the application of these technologies to real problems in real military settings. However, making this a reality depends on training a new generation of students in the use of these Cornell-developed technologies and sparking interest in relevant career paths. Therefore, the focus of the project was to involve students by exposing them to these tools and showing them non-classified examples of problems that might be applicable to military scenarios.

The process for achieving this objective started by recruiting students interested in mission-oriented and information-based computing, and then continued with defining and completing projects based on the students' specific interests and abilities. Having the students present their work in an appropriate setting with high quality was also considered part of the process. With the help of funding provided by the Air Force Research Laboratory (AFRL), Professors Ken Birman and Hakim Weatherspoon and Principal Research Scientist Robbert van Renesse were able to mentor many undergraduate and master's students. Several Ph.D. students and postdoctoral associates additionally took on supervisory roles within the project teams. The work performed exposed these students to some of the technologies described above and led to the successful completion of four main projects leveraging these tools: Scalable Landmark Recognition (ePaparazzi), the Software-defined Network Interface Card (SoNIC), Smart Grid Security (GridControl), and a group of projects relating to Birman's Isis2 platform.

The ePaparazzi project was a collaboration between the distributed systems and computer vision research groups at Cornell and focused on creating a scalable, publically-available system, able to directly match user-submitted photographs with corresponding real-life geographic landmarks on the Earth. SoNIC is a unique measurement and monitoring apparatus that allows researchers and network operators to take a closer look at networks that interconnect our military networks. The GridControl project has focused on hardening the power grid and building a framework that will look into smart grid security issues and help protect the grid from cyber attacks. Finally, the remaining projects have focused on specific topics relating to Birman's Isis2 platform, which is a programming library designed to help developers incorporate high-assurance computing properties into cloud-based applications. [1]

As mentioned above, carefully prepared presentations of the results were also considered part of the process. For the ePaparazzi project, students presented their work at the Cornell Bits On Our Minds (BOOM) event in April of 2012. This annual event showcases student projects and considers their application of core computer science ideas in accessible technology. The ePaparazzi work performed won the team a Cornell BOOM 2012 Innovation Award and the Department of Computer Science's Master of Engineering Group Project of the Year award for 2012. Also at BOOM 2012, the research project "Detection of DDoS Attacks Using Gossip" received the 2012 AFRL Achievement Award, presented by Dr. Mark Linderman. The Appendix section contains a list of the relevant information management student projects presented at BOOM in 2012 and 2013.

Students' efforts in completing projects also contributed to publications for the International Conference on Principles of Distributed Systems [2]; the Euromicro International Conference on Parallel, Distributed and Network-Based Processing [3]; the USENIX Symposium on Networked Systems Design and Implementation [4]; the IEEE/IFIP International Conference on Dependable Systems and Networks [5]; as well as several student undergraduate and master’s degree research projects. Overall, the work completed in each of the projects under this award was a success, and the students involved will benefit from this exposure as they advance to the next steps in their career paths. Historically, a significant percentage of Cornell students have pursued careers in the military, with military contractors, and in the civilian critical infrastructure areas.

2.0 INTRODUCTION

Faculty and researchers in the Department of Computer Science at Cornell University have been studying the challenges of information management and high-assurance computing in a variety of demanding settings, including cloud computing, data transmission in complex network environments, and the exploration of the underlying scientific basis for high-assurance computing. These research efforts have resulted in the development of many useful technologies, such as sophisticated application development tools and platforms, advanced security frameworks and tools, and purely conceptual tools such as new theories of high-assurance that respond to the most stringent requirements. In particular, there has been recent focus on cloud computing and the creation of highly resilient, secure, cloud-hosted services.

Often the research is guided by thinking about important military scenarios as well as other nationally critical infrastructure challenges, and there is potential for considerable positive impact in the application of these technologies to real problems in real military settings. However, making this a reality depends on training a new generation of students in the use of these Cornell-developed technologies and sparking interest in relevant career paths. Therefore, the focus of the project titled *Engaging Students via Innovative Militarily Useful Technologies* was to involve students by exposing them to these tools and showing them non-classified examples of problems that might be applicable to military and important civilian scenarios.

With this AFRL funding, we were able to mentor many undergraduate and master’s students interested in mission-oriented and information-based computing. Several Ph.D. students and postdoctoral associates additionally took on supervisory roles within the project teams. At the start of the grant period, we held a Symposium on Student Advancement via AFRL-sponsored computer science (CS) Research at Cornell, where student research briefings were conducted. Table 1 shows a list of these presentations. The work performed by the students exposed them to some of the technologies described above and led to the successful completion of four main projects leveraging these tools: Scalable Landmark Recognition (ePaparazzi), the Software-defined Network Interface Card (SoNIC), Smart Grid Security (GridControl), and a group of projects relating to Birman's Isis2 platform.

Table 1. Student Research Briefings

Project Title	**Team Members**
Compositional Gossip Protocols	Lonnie Princehouse, Nate Foster, Ken Birman
Fault-tolerant TCP for a Stable BGP	Robert Surton, Ken Birman, Robbert van Renesse
Elastic Replication for Scalable Consistent Services	Hussam Abu-Libdeh, Haoyan Geng, Robbert van Renesse
SoNIC: Software-defined Network Interface Card	Han Wang, Ki Suh Lee, Hakim Weatherspoon
Small-World Datacenters	Ji Yong Shin, Hakim Weatherspoon
Monitoring and Visualization Framework for Isis2	Patrick Dowell, Qi Huang, Daniel Freedman, Ken Birman
Scalable Landmark Recognition	Hee Jung Ryu, Scott Phung, Kaushik Nataraj, Ansu Abraham, Qi Huang, Daniel Freedman, Noah Snavely, Ken Birman

The ePaparazzi project—a collaboration between the distributed systems and computer vision research groups at Cornell University—aimed to create a scalable, publically-available system, able to directly match user-submitted photographs with corresponding real-life geographic landmarks on the Earth. ePaparazzi does not rely on any use of geotagging, metadata, or other non-image information, but instead deconstructs the visual information contained in a photograph and matches it to a corpus of three-dimensional reconstructed locations. A user can upload an image containing a landmark and the system uses computer vision content-based image retrieval (CBIR) techniques to return the latitude/longitude of where the photo was taken in relation to the landmark.

Our work can take a photograph and -- using visual information alone -- compute the exact position and orientation of the camera by matching it to a large database of three-dimensional models (or determine that the photo is not recognized). Our method is extremely accurate and scalable, recognizing locations for photos in a few seconds by quickly matching to thousands of sites, even if no location information is known about the photo in advance. Our work leverages large-scale distributed computing using the Isis2 framework (a new distributed computing library that simplifies the creation of strongly consistent, secure, fault-tolerant cloud computing and HPC services, available open source from Cornell at <isis2.codeplex.com>), and has applications in forensics as well as consumer photographs. Figure 1 shows a few recognized photos displayed on a map at their estimated latitude and longitude. Cornell faculty leaders for the ePaparazzi project included Professor Ken Birman from the systems group and Professor Noah Snavely from the computer vision team.

Figure 1. ePaparazzi Recognized Photos

The next project engaged students in the development and use of SoNIC. The research group led by Professor Hakim Weatherspoon at Cornell has built SoNIC as a unique measurement and monitoring apparatus that allows researchers and network operators to take a closer look at networks that interconnect our military networks. Our military has moved from being enabled by the network to being dependent on it. The network must work, and it must work securely, for everything from connecting the Department of Defense cloud together (the global information grid), to coordinating military action, ordering supplies, and checking email. However, the network is physically complex where an end-to-end path may include fiber-optic links, satellite, and line-of-site communication.

To study, monitor, and understand these military networks, Software-defined Network Interface Card (SoNIC) was built as a specialized measurement apparatus designed for the sensitive timings required of the high data rates and diversity of these networks. SoNIC is a real-time network adapter that can control and change the physical layer encoding in software in a similar fashion that a software-defined radio allows a wireless medium access layer to be controlled and changed in software, thus allowing a 10 gigabit network stack to be studied at a heretofore inaccessible level. Figure 2 depicts the concept behind SoNIC. SoNIC provides users unprecedented access to the physical and data link layers of network protocol stack in software and in realtime. By implementing the creation of the physical layer bitstream in software and the transmission of this bitstream in hardware, SoNIC provides complete control over every bit in software. SoNIC consists of commodity off-the-shelf multi-core processors and a field-programmable gate array (FPGA) development board with high-bandwidth Peripheral Component Interconnect Express (PCIe) interface. With software access to the physical layer, SoNIC can perform precise network measurements, characterize network components, and enable novel network research applications that were not previously feasible.

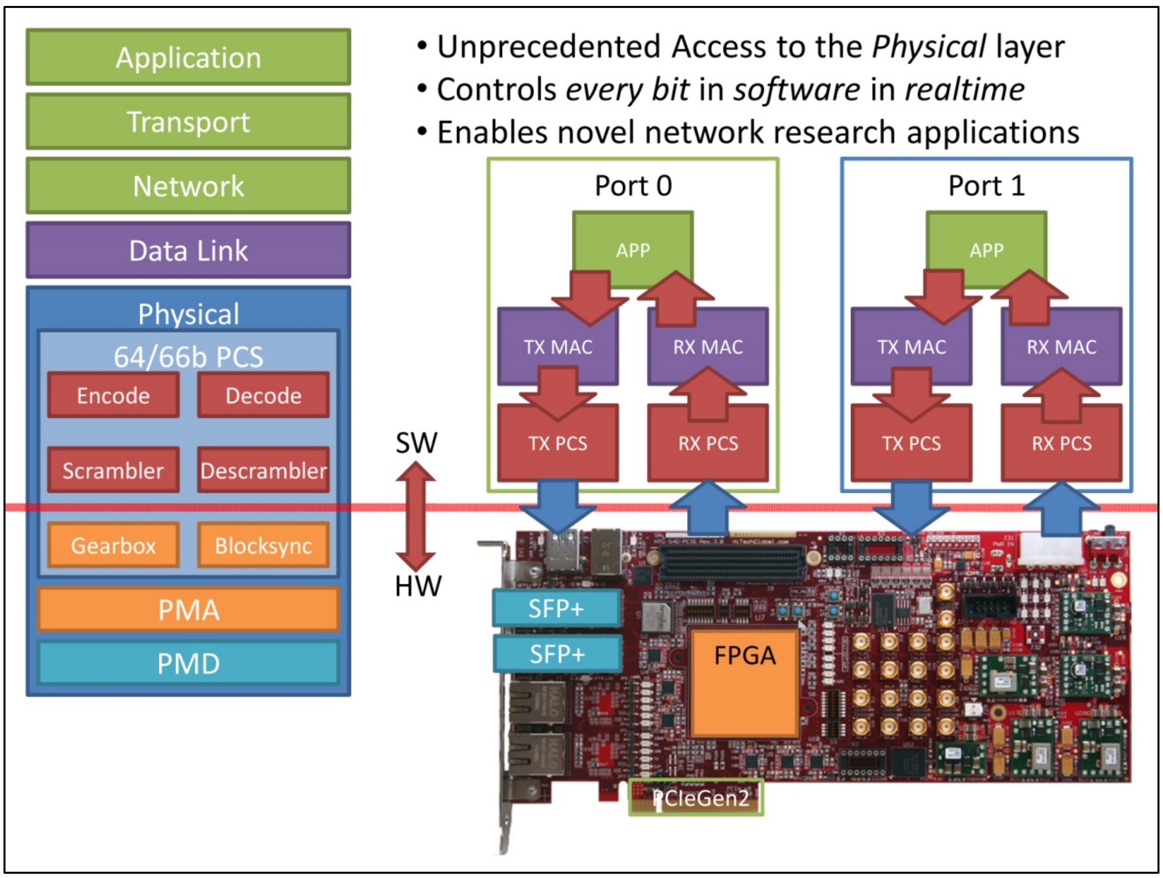

Figure 2. SoNIC: Software-defined Network Inferface Card

The research group has been designing, implementing, and using SoNIC for state-of-the-art research. One of the applications developed using SoNIC demonstrates how one can create a covert channel over a 10 gigabit Ethernet (GbE) using a technique called SoNIC Steganography, which is made possible because SoNIC can control and modify the line encoding in real-time in an undetectable manner. State-of-the-art military networks require state-of-the-art methodologies to understand and use them efficiently. SoNIC-enabled networks are a crucial enabling step. Informed by the improved understanding these devices facilitate, we expect to develop better networks and protocols for moving large quantities of data securely and reliably over our military networks.

During the summer of 2012, we brought students onboard to the GridCloud project led by Birman and Principal Research Scientist Robbert van Renesse for research into hardening the power grid. GridCloud is an effort within a larger project we call GridControl, being undertaken in collaboration with Washington State University and under primary sponsorship by the Advanced Research Projects Agency-Energy (ARPA-E) Green Electricity Network Integration (GENI) program, operated out of the Department of Energy. Today's smart grid is increasingly getting integrated with the cyber infrastructure for various kinds of information processing. This puts the grid into a vulnerable position with respect to cyber attacks from external malicious parties.

We have embarked upon the building of a framework that will look into smart grid security issues and help protect the grid from cyber attacks. GridCloud is a cloud-hosted, high-assurance platform we are creating to monitor and manage complex sensor networks with real-time properties. Figure 3 is a visual representation of how GridCloud works. Many thousand phasor measurement units (PMUs) collect data which is replicated and used in a state estimation algorithm. Creating a high-assurance platform would help prevent blackouts such as the one shown in Figure 3 (note the extensive blackout area from a 2003 satellite image).

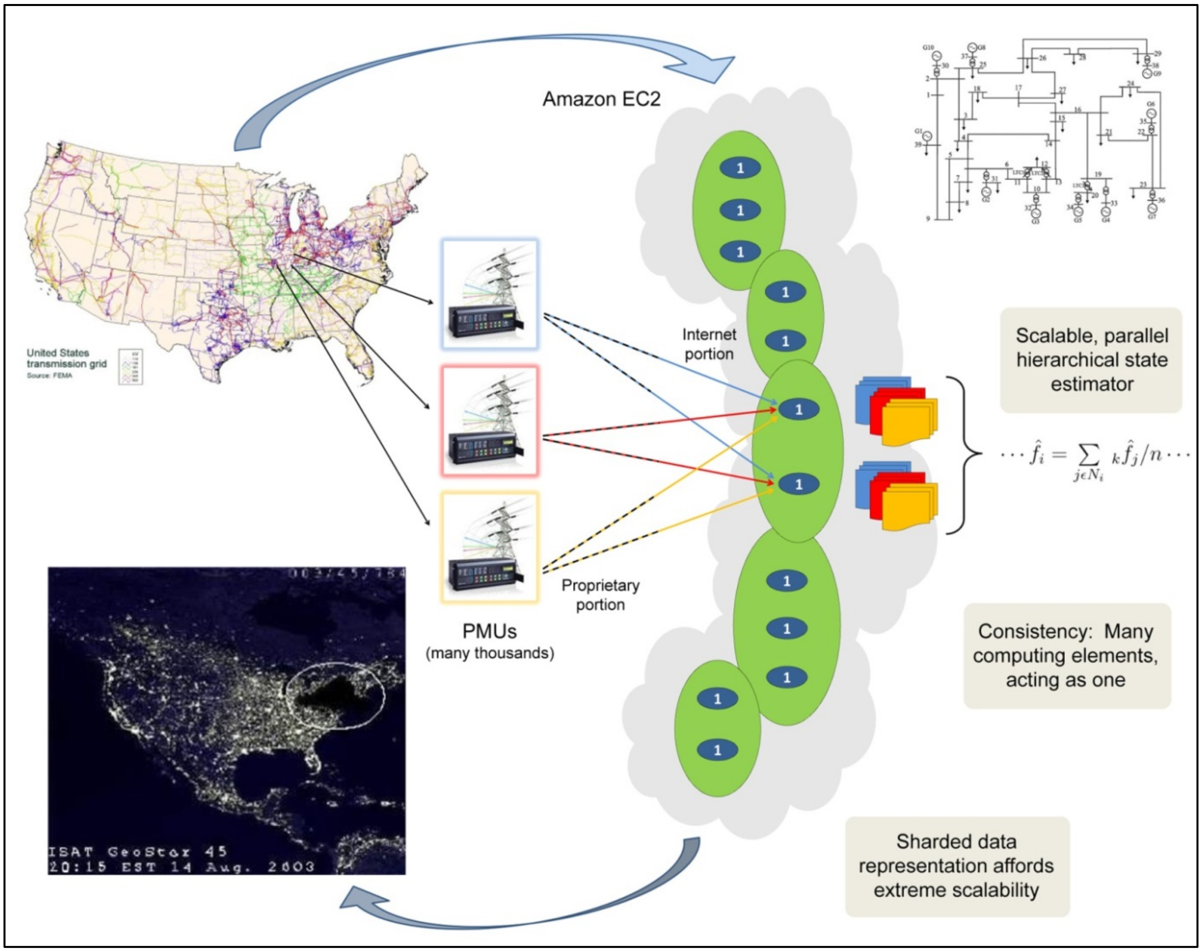

Figure 3. Representation of the GridCloud Platform

Finally, during the 2012-2013 academic year, we also involved students in projects relating to Birman's Isis2 platform. [1] Isis2 is a new platform developed by Birman at Cornell under Defense Advanced Research Projects Agency (DARPA) funding. It assists the developer who uses it in creating high-assurance services and big-data repositories for use in cloud computing settings. By automating such tasks as fault-tolerance, security, coordination, consistency preservation and health monitoring, and triggering reconfigurations after crashes, Isis2 makes it easy to build a self-healing and strongly assured cloud computing solution even for demanding, very large-scale use cases.

The Isis2 projects covered specific topics relating to the system, such as experimenting with the new distributed hash table (DHT) architecture, integrating Isis2 with Cornell's older Live Distributed Objects platform, and making Isis2 useable from programs written in other languages. Live Objects enables a powerful drag-and-drop style of collaborative application development. Using it, a non-programmer can create and share a specialized "digital dashboard" customized for a unique situation that has arisen suddenly and in which nimble responsiveness is key to mission success. For example, when a cloud computing system is used to monitor the electric power grid, prevention of an outage may require a reaction faster than the propagation speed of the disruption through the grid, which occurs roughly at the speed of sound. Obviously, networks (which move packets at the speed of light) are faster, but for a computing system to have time to read a message in, process it, and then react (for example by adjusting trip-points on circuit breakers) demands a form of very nimble responsiveness.

The Isis2 DHT is used to spread the data widely at very low cost and with strong reliability guarantees. Figure 4 depicts the system architecture of Isis2 and Ida, its Interactive Data Analysis layer. As seen in the illustration, systems such as Isis2 are immensely complex. Today's cloud infrastructures lack the needed technologies to support high assurance, and leave it to the developer to create all the required tools. High assurance services offer multiple resiliency properties, notably data security, fault-tolerance, consistency, self-repair, etc. Different use cases require different mixtures of such properties.

With Isis2, our Cornell students can build applications of a kind that would normally require major teams and lengthy development cycles even at cutting-edge companies like Microsoft, Google or Facebook.

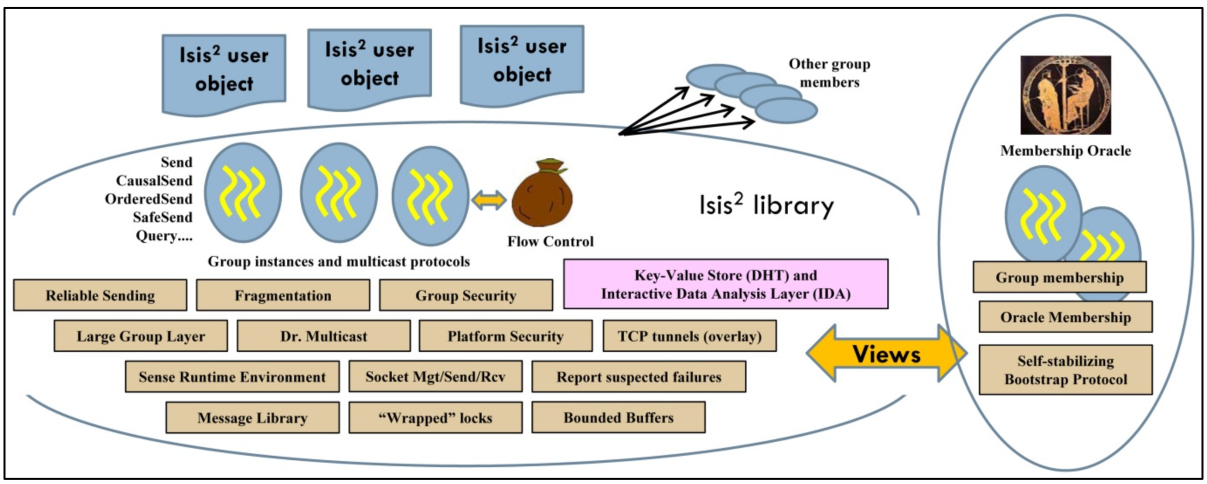

Figure 4. Isis2 and its Interactive Data Analysis Layer, the IDA

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

We were able to recruit several undergraduate and master's students to participate in projects under this award. Cornell's Master of Engineering program in Computer Science requires students to complete a project as part of their degree, and therefore we were in an ideal position to expose students to a variety of project ideas.

Projects were chosen based on the students' abilities and interests. We aimed to find projects that would pose intellectual challenges but that could be solved and would lead to a solution demonstrating the desired capability. Carrying out this effort required close supervision of the students. Faculty leads worked directly with the students, but we also engaged postdoctoral associates and Ph.D. students in supervisory roles. The final step for a project was to present the research in a high-quality format, such as a poster displaying the work, participation in the Cornell BOOM event, or publication of the results in a paper. The Appendix section contains a list of the relevant information management student projects presented at BOOM in 2012 and 2013. Table 2 shows the list of tasks and milestones we proposed for carrying out the projects. Figure 5 shows the organizational structure for each of the four main projects.

Table 2. Proposed Tasks, Schedule, and Milestones

Task	Sub-task description	Time period	Milestones
Recruit student Participants	Identify candidates interested in working with our effort	Start of each semester	Pool of candidates selected
Develop project plan	Done by each student or student team	First two weeks of semester	Approval of plan by supervising faculty or researcher
Implement technology, evaluate	Carried out by the students or student teams	Subsequent 10-12 week period	Meetings once every week or two weeks with supervising faculty or researcher. Possible re-planning if the work encounters a setback or obstacle.
Document through report, poster	Carried out by the students or student teams	End of semester prior to final due date for grades	Must be approved by supervising faculty member or researcher
Present solutions at BOOM	Carried out by the students or student teams, with AFRL representatives present to review work done and assist in judging BOOM projects, awarding prizes.	March/April timeframe	Required with some exceptions for special situations.

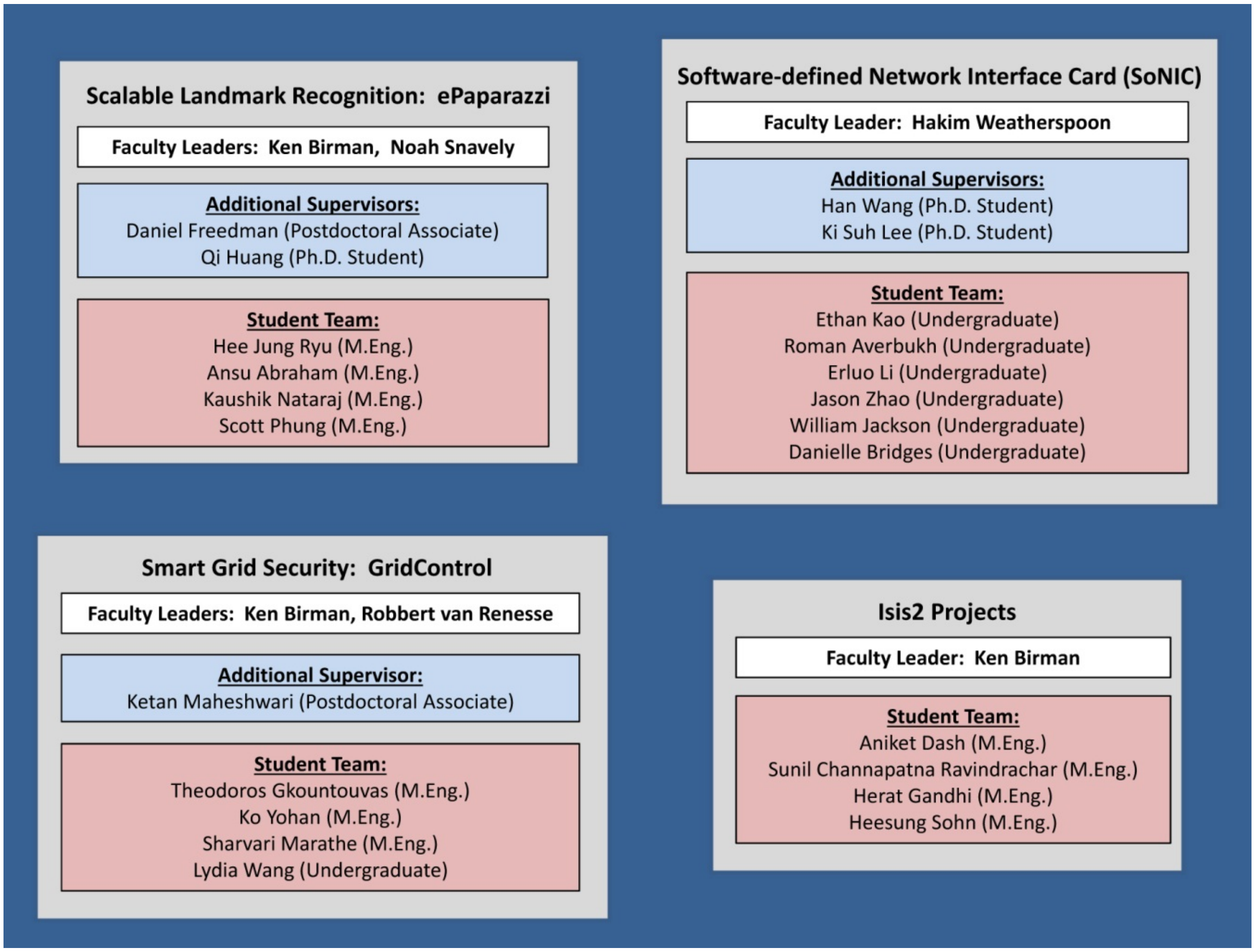

Figure 5. Organizational Structure of Projects

3.1 Scalable Landmark Recognition

For the ePaparazzi project, Cornell's computer vision community supplied the research project to perform the image recognition and landmark association, while the distributed systems team worked on leveraging an infrastructure platform (Isis2) to transform such a single-user image program into a unified system—one that scales to large numbers of simultaneous users searching a sizeable landmark corpus.

Faculty leaders for the project included Birman from the systems group and Snavely from the computer vision team. Postdoctoral associate Daniel Freedman and Ph.D. student Qi Huang directly supervised the student project team. This team was comprised of four students in the Computer Science M.Eng. program: Hee Jung Ryu, Ansu Abraham, Kaushik Nataraj, and Scott Phung. The students themselves brought interesting and diverse backgrounds to bear. Hee Jung has interned at Facebook, Ansu at IBM, and Kaushik at both IBM and Tech Mahindra, while Scott worked full-time for SAP for a number of years. Individually, they were a motivated, hardworking group, and collectively they were anxious to deliver a working system that provides unique user experiences while leveraging Cornell-developed computer vision and distributed systems research.

From a technical management perspective, the project was divided into four different divisions:

1) Performance enhancements of the front-end, user-facing module
2) Scale-out of the back-end (doing feature matching and analysis of landmark coincidence)
3) Creation of deployment and configuration tools
4) Design and implementation of monitoring and analysis toolset

Hee Jung served as both the student project manager as well as the engineer charged with wrapping and integrating the computer vision image search algorithm into the Isis2 infrastructure. Ansu worked on developing the front-end, user-facing web application, as well as the Isis2 groups that manage and load-balance the image workflow from front-end to processing nodes. Kaushik designed test suites to ensure that both the ePaparazzi application and the underlying Isis2 platform deliver the behavior that their interfaces claim. Scott focused on the design and implementation of the Isis2 back-end computational services that interface with the underlying image search algorithms.

A major theme underlying our progress was the successful transition to a cloud computing platform. In short, we abandoned our earlier small collection of local servers and integrated our architecture and codebase within the Amazon Web Services (AWS) model, with real-world scalability benefits accruing from elasticity and carefully metered cost.

During the first quarter of 2012, we made the following progress in each of the four divisions:

1) On the front-end side, we utilized a specialized implementation of the scale-invariant feature transform (SIFT) computer vision algorithm, which is tailored for parallelized workloads on multi-core nodes. We also integrated our system with Amazon's Elastic Compute Cloud (EC2) load balancer to enhance both per-request speed and multi-requests load distribution

2) On the back-end of our implementation, we switched from using physical Internet Protocol multicast (IPMC) to a mode that generates appropriate overlays atop the Transmission Control Protocol (TCP), in order to support the AWS model. We also began investigating issues with concurrency stability on our path to scale to larger system sizes, such as 50~100 nodes, for example.

3) With respect to our deployment and configuration tools, we released a beta version of such functionality. We adopted Amazon’s EC2 application programming interfaces (APIs) and our own ‘daemon’ (type of program that runs unobtrusively in the background) and demonstrated automatic deployments, upgrades, system launches, and system halts atop any EC2 instance.

4) Regarding our monitoring and analysis tools, we customized them to collect application-level logs for each node in our cloud deployment and performed off-line analysis to extract latencies of message transit throughout the application. We then began to extend this to visualize internal system events with a dedicated group of logging nodes.

Research continued throughout the spring of 2012, and our team worked on scaling the project from internal Cornell servers into Amazon's EC2. We were able to scale from 19 virtual machine instances on two Windows machines to 160 instances on 70 EC2 nodes. This was the big achievement as we came up with and tested an architecture to support this, based on load balancing and multiple feature matching groups. We also tested the accuracy and speed of the system in regard to distributed feature matching corpus shard size (degree to which the data is replicated; normally 2 or 3), and dynamically resizing image resolution.

We were able to increase throughput and reliability from 5 concurrent requests to 40 as well as reduce the latency of a request from 8 seconds to 3. We scaled the system horizontally by introducing multiple feature matching groups. Each group is one self-contained Isis2 group, and the groups are joined together by an Amazon Load Balancer and TCP socket calls. One of the challenges of Amazon was that it did not support multicast over nodes. To reduce latency during feature extraction, we used a multi-thread SIFT implementation called OpenMP over David Lowe's slower single threaded SIFT. In addition, the team worked on a live logging framework to capture statistics during individual stages as well as a new user interface and a deployment/fault-tolerance/elasticity piece.

The project team also made significant progress in terms of deployment and configuration tools, developing a customized server daemon, integrating with Amazon's EC2 APIs, and demonstrating automatic deployments, upgrades, system launches, and system halts atop any EC2 instance. Finally, the monitoring analysis is similarly sophisticated. The team has created applications to aggregate customized application-level logs for each node in our cloud deployment, perform off-line analysis to extract latencies of message transit throughout the application, and examine visually the internal system events with a dedicated group of logging nodes.

3.2 Software-defined Network Interface Card

At the start of the award period, the SoNIC project, led by Weatherspoon, included two Ph.D. students and one undergraduate student. Han Wang is an Electrical and Computer Engineering Ph.D. student with a strong background in designing for FPGAs, and Ki Suh Lee is a Computer Science Ph.D. student who has significant background in operating systems. Ethan Kao completed his undergraduate degree while working on this project.

Throughout the course of the award, several additional undergraduate students joined the project. Roman Averbukh, Erluo Li, and Jason Zhao joined the research group during the spring semester of 2012. Two additional undergraduates joined for a Research Experience for Undergraduates (REU) during the summer of 2012: William Jackson from Cornell and Danielle Bridges from Jackson State University.

There were three areas in particular that the SoNIC research group was studying during this project:

1) **Network steganography**. SoNIC enables the creation and detection of a covert channel over 10GbE. This is possible because SoNIC can control and modify the line encoding in real-time in an undetectable manner. SoNIC Steganography allows the military (e.g. AFRL) to detect, suppress, or even instigate covert channel attacks as low as the network physical layer.

2) **Network monitoring, filtering, and fingerprinting**. Ultimately, SoNIC plugs into a commodity server, a software router, like any network interface card (NIC) and can be used to monitor, filter, and fingerprint network traffic in real-time with extremely high precision. This can be combined with other software router research, which can operate from 10 to 100 gigabits per second (Gbps).

3) **Network measurements**. SoNIC allows researchers to measure and characterize military networks at a heretofore inaccessible level, and to do so in software. We believe that these measurements will help explain observed packet loss at low data rates and reveal new aspects of network behavior, and inform new, more secure network protocol design.

Within the project team, Han Wang has been responsible for ensuring that the hardware side of SoNIC is able to interface with the network on one side and the host processor on the other. This requires seeking a balance between implementing functionality in hardware such as clock recovery and functionality in software. Further, Han has worked on creating a path to map SoNIC onto a NetFPGA board. Ki Suh Lee has been responsible for implementing the software side of SoNIC and ensuring its performance. Ki Suh has also worked on creating a SoNIC-enabled software router that will perform the deep packet inspection for the data in motion and alert system.

During the first quarter of 2012, Han and Ki Suh were able to plug two SoNIC cards (that we developed) into one commodity server and receive on one card and send on the other. Han used an FPGA on the cards to receive or transmit the physical signal over the fiber and to also direct memory access (DMA) transfer the raw bits into the host processor. Ki Suh developed the software to do layer 1 and 2 processing of the raw bits. He designed one core to descramble the bits and another core to decode the bits; we then used a separate core to encode and scramble the bits before sending. All of this was successful and we were able to do higher level processing on the raw bits such as implementing covert channels. Examples off higher level processing would be HPC computations that are used to perform power grid state estimation, or to compute the adjustments needed to set-points of power grid circuit breakers.

Their research continued throughout the spring of 2012, and they have been able to demonstrate the networking functionality and precision of SoNIC. In particular, we are able to use SoNIC to profile the characteristics of routers with picosecond precision. We can also use SoNIC to identify how different routers affect the performance of network communication. Moreover, we are able to validate previous experiments of a different system called BiFocals since we obtain similar results. Created by Dan Freedman and Hakim Weatherspoon, BiFocals supports dual-resolution network monitoring. One very high resolution monitoring channel operates close to the fiber optic medium. The second lower resolution one monitors the data channels that reach user-level application processes, after data has travelled through the network interface card (NIC). The system uses a mixture of off-the-shelf analog and digital hardware components with a novel software infrastructure.

Undergraduate students Jason Zhao and Erluo Li have been working with our servers to increase our routing capabilities from 10 Gbps to 40 Gbps. They have had to contact the Myricom NIC vendors to ensure that the commodity NICs interact with the host processors properly (e.g., evenly distribute packets between cores). They were able to configure and program our commodity servers and commodity NICs from Myricom such that the servers could function as 40 Gbps routers. This enables the servers to be not only routers, but also general high performance packet processors (e.g. the server can possibly do deep packet inspection of every packet at 40 Gbps). We then used SoNIC to profile the timing characteristics of the servers as they function as routers.

Ethan Kao and Roman Averbukh worked on designing the storage and timing covert channels. Ethan completed the design of a storage covert channel, and Roman investigated the bounds of a timing channel; namely, the minimum interpacket gap spacings to identify and differentiate a 1 from a 0—the basic elements of communication. Danielle Bridges continued where Ethan and Roman left off, using SoNIC to design, create, and test a timing channel. During the summer of 2012, William Jackson used SoNIC to measure and uniquely identify network devices such as switches and routers to create “network signatures.”

3.3 Smart Grid Security

The GridControl project was led by Birman and van Renesse, with postdoctoral associate Ketan Maheshwari supervising students as well. The team included three students in the Computer Science M.Eng. program— Theodoros Gkountouvas, Ko Yohan, and Sharvari Marathe—along with Lydia Wang, an Electrical and Computer Engineering undergraduate student.

Within the project team, Theo has been working on the integration of Isis2 into the GridCloud platform. He is carefully studying the network connection's behavior and analyzing ways to implement Isis2 functionality to provide an efficient fault-tolerant framework which will help a continuous availability of PMU data in the wake of network failures and fault conditions. We are closely looking into the possibilities to integrate the Isis2 library to provide a reliable and fault-tolerant dataflow occurring simultaneously across the cloud and the web over thousands of TCP channels. He has been looking closely at the possibility of implementing proprietary transmission protocols such as those recommended by the North American SynchroPhasor Initiative Network (NASPInet) consortium.

The goal of Theo Gkountouvas’s work was to look at how we can script management actions such as launching programs, providing them with parameterization and configuration data, monitoring them as they run, and adaptively healing an application that is disrupted by a failure. He has been doing this by building a distributed framework that works, as much as possible, with standard Unix-style "make" files, redefining them to make sense in distributed settings and adding the needed distributed framework to have this actually work under production conditions. Theo designed the solution during the fall of 2012 and began implementing it during the winter period, and continued this work into the spring semester of 2013.

Ko Yohanhas been studying the PMU data storage and archival strategies. Since the input/output (I/O) speeds are tens of orders slower than the computations speed, efficient I/O management is essential and can potentially cause a bottleneck otherwise. To address this challenge, we are closely looking into cache management policies and concurrent I/O techniques. Currently, a simple data storage policy is in place. It implements a lightweight, cyclical, fixed-size log system to avoid clogging up the disc. A future version of data storage and archival will implement a concurrent, cached and managed version of PMU data storage and archival.

Sharvari Marathe has been researching the implementation of a simple but high precision state estimation algorithm. Her implementation takes the linear state estimation approach estimating the state from continuously obtained emulated PMU observations and an error vector. The method addresses the different elements of PMU data and iterates in quasi-real-time over this data to give a near-live state estimation matrix. The current implementation takes into account a hierarchical arrangement of state estimators in the wake of hundreds of source PMUs. The data is received from configurable multiple sources and the state is sent out over the network on predefined ports respectively. This allows another state estimator downstream to gather this data for further processing and refinement.

Lydia Wang has been researching how to test and maintain the framework, and she has also worked on writing and enhancing some basic Python programs of the GridCloud framework. Lydia has helped with the installation of supporting programming libraries on our experimental test-bed, along with the setup of an environment for experiments. She has additionally worked on refining the documentation and manuals for the GridCloud platform.

3.4 Isis2 Projects

Professor Ken Birman also supervised a few students on projects that relate to specific aspects of his Isis2 platform [1]. These projects involved four additional M.Eng. students who are all finishing their degrees during the spring of 2013: Aniket Dash, Sunil Channapatna Ravindrachar, Herat Gandhi, and Heesung Sohn.

Aniket Dash and Sunil Channapatna Ravindrachar have worked with us on creating a way for unmanaged C++, C or non-.NET Python and Java programs to communicate with our new Isis2 library. These types of programs cannot easily link to the Isis2 library because it currently runs only in .NET— Isis2 runs on Linux and Windows, and can be used from C#, C++/CLI, IronPython and Visual Basic, but not easily from pure Java, native C++ or C, or other non-.NET languages. Aniket spent the fall of 2012 learning the format of Isis2 messages and writing a C++ module to marshal into and out of the Isis2 format. During the winter he continued this work and was able to get some simple Isis2 system calls working. In the spring semester, Aniket and Sunil have made very strong progress and are wrapping up their "outboard" library now (where some components of the library reside in a daemon that runs outside the address space of the application using it). However, they were not able to support the full Isis2 API and some functionality still can only be used from C++/CLI (a version of C++ that runs in .NET).

Herat Gandhi is working to integrate the Isis2 system with Cornell's older Live Distributed Objects platform. The Live Objects system was a focus of our work until about five years ago, but since then has been idle, available in open-source form but not actively maintained or enhanced. Herat's goal is to get Live Objects to run over Isis2 multicast, which should give big speedups and new flexibility. He also is porting Live Objects to run on more modern versions of Windows.

Heesung Sohn has been working with Professor Birman to carry out experiments on a new distributed hash table (DHT) architecture aimed at bringing strong consistency to the space of dynamically updated "big data" systems. The DHT itself is a part of Birman's new Isis2 platform and was developed with DARPA funding, managed by AFRL under Patrick Hurley; the basic idea is to capture very large amounts of "key-value" data into a scaled-out structure in such a way that if we have more key-value data, we can just add more nodes to handle the extra capacity. But unlike standard DHT systems, the Isis2-based solution (which we are calling Ida, the Interactive Data Analysis framework) is also highly assured with strong consistency and security properties. This work is the basis of a paper we recently submitted to the Association for Computing Machinery (ACM) Symposium on Operating Systems Principles (SOSP).

4.0 RESULTS AND DISCUSSION

Cornell has developed many useful technologies in information management and high-assurance computing which have the potential for application in real military scenarios; however, in order for this potential to be realized, we need to expose a new generation of students to the use of these techniques. Therefore, the main focus during this award was to engage students in projects that utilized these Cornell-developed, militarily useful technologies. Our first task was to recruit students from our undergraduate and master's programs who are interested in mission-oriented and information-based computing. We were able to recruit several students to work on such projects, and we additionally involved Ph.D. students and postdoctoral associates in supervisory roles.

The next task was to define projects based on the students' abilities and interests. We successfully carried out four main types of projects that incorporated Cornell-developed technologies: Scalable Landmark Recognition (ePaparazzi), the Software-defined Network Interface Card (SoNIC), Smart Grid Security (GridControl), and projects relating to specific aspects of Birman's Isis2 platform. We also aimed to have the students present their research and showcase the results. This task is an important skill for students to learn, and we were able to provide mechanisms for achieving this step.

Our ePaparazzi team presented a poster along with a live demonstration of the application at the Cornell BOOM event in April of 2012. During the demonstration, all processing and landmark recognition was performed in real-time in the AWS cloud. This annual event allows the showcase of student projects and considers their application of core computer science ideas in accessible technology. The work we did won the team a Cornell BOOM 2012 Innovation Award and the Department of Computer Science's Master of Engineering Group project of the year award for 2012. The Appendix section contains a list of the relevant information management student projects presented at BOOM in 2012 and 2013.

In the SoNIC project, the research agenda and platform that we created allowed students to engage in network research with a very high degree of control, flexibility, and insight, and use SoNIC to make scientific contributions. In particular, Erluo Li used SoNIC to characterize and classify different routers and networks.

Both graduate students Han Wang and Ki Suh Lee have been instrumental in fully implementing, evaluating, writing, and presenting SoNIC. Our paper titled "SoNIC: Precise Realtime Software Access and Control of Wired Networks" [4] was accepted for publication at the 2013 USENIX Symposium on Networked Systems Design and Implementation (NSDI). Han and Ki Suh presented this paper at a Cornell Systems Lunch seminar in March and then again at the NSDI conference in April. In the paper, we discuss SoNIC's architecture and performance and demonstrate its utility, precision, and flexibility. We are able to create exact packet generators and captures from software and use this capability to profile different network elements such as a variety of routers and NICs. Moreover, we are able to create a timing channel that is orders of magnitude less detectable than previous techniques (i.e. a 0 or 1 can be sent with nanosecond interpacket gaps instead of millisecond interpacket gaps). We are currently testing if such precise timing channels can propagate across a network.

Undergraduates Jason Zhao and Erluo Li, who joined the project in the spring of 2012 and continued working throughout the summer, were able to complete independent research projects in the fall. Jason's research involved configuring and programming our commodity servers and commodity NICs from Myricom such that the servers could function at 40 Gbps routers. This enables the servers to be not only routers, but also general high performance packet processors (e.g. the server can possibly do deep packet inspection of every packet at 40 Gbps). We then used SoNIC to profile the timing characteristics of the servers as they function as routers. Since completing this project, both Jason and Erluo have accepted offers of admission to Ph.D. programs—Jason at the University of California, Berkeley, and Erluo at Cornell.

Ethan Kao and Roman Averbukh wrote a semester/term paper on their experience investigating and creating storage and timing covert channels using SoNIC. Ethan completed the design of a storage covert channel, and Roman investigated the bounds of a timing channel—namely, the minimum interpacket gap spacings to identify and differentiate a 1 from a 0, which are the basic elements of communication. Danielle Bridges continued this work, using SoNIC to design, create, and test a timing channel. Han and Ki Suh are now researching this topic and are planning to write a paper. During the summer of 2012, William Jackson used SoNIC to measure and uniquely identify network devices such as switches and routers to create "network signatures."

In the GridControl project, Theodoros Gkountouvas is in the unusual situation of finishing his M.Eng. degree but then starting in the computer science Ph.D. program in the fall, having been admitted from one program to the other. For his work on GridCloud, he has built a new system, DMake, which uses Isis2 internally and provides secure and consistent management services for an application that might be spread over large numbers of nodes in a cloud setting. DMake uses the familiar Linux “Makefile” syntax for its control files, but extends the “Make” model to work in a decentralized manner. Our collaborators at Washington State University are about to shift to using DMake, and a paper on the work is planned.

Engaging students in this type of research was our priority, and the students involved in these projects have benefited from the experience in many ways. By participating in research outside of the classroom, students are provided with an opportunity to see real applications of the theories they study; furthermore, they are trying to solve problems that might not have clean, expected solutions like the ones they encounter in their courses. The students worked in teams since the projects were large, and teamwork skills are invaluable no matter what direction their careers take. In the past, however, we have seen students involved in these types of projects continue on to careers in defense industries and hold positions at organizations such as AFRL.

5.0 CONCLUSIONS AND RECOMMENDATIONS

Our main objective for this AFRL project titled *Engaging Students via Innovative Militarily Useful Technologies* was to expose students to the research and use of information management tools and high-assurance computing technologies developed at Cornell which have the potential for application in military scenarios and in nationally important critical infrastructure areas. We recruited several undergraduate and master’s students to participate in projects under this award. Under the guidance of faculty leaders in addition to Ph.D. students and postdoctoral associates, the students successfully completed four main types of projects utilizing some of these technologies: Scalable Landmark Recognition (ePaparazzi), the Software-defined Network Interface Card (SoNIC), Smart Grid Security (GridControl), and additional projects relating to specific aspects of Birman’s Isis2 platform.

We conclude that exposing students to high-value military scenarios and technologies is an effective technique for training a new generation of employees who might well enter the US military supply chain or vendor community, but do not believe that this level of student is likely to be able to innovate or play leadership roles in solving the US military’s most challenging open problems. That is, we can train students to use best of breed technologies, and believe that this will ultimately contribute towards improved national security if carried out on a large scale. However, students at this level cannot be expected to invent new solutions where none currently exist. Our firm belief is that continued investment to support researchers in confronting today’s tough computing challenges is vital to maintaining a process by which the state of the art can slowly be evolved, and also enabling rare breakthroughs that can be game changers in areas such as security, reliability and scalable real-time computing.

REFERENCES

[1] K. Birman, "Isis2 Cloud Computing Library," [Online]. Available: http://isis2.codeplex.com/.

[2] R. van Renesse, C. Ho and N. Schiper, "Byzantine Chain Replication," in *The 16th International Conference On Principles Of Distributed Systems (OPODIS)*, Rome, Italy, 2012.

[3] R. van Renesse, H. Johansen, N. Naigaonkar and D. Johansen, "Secure Abstraction with Code Capabilities," in *The 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*, Belfast, Northern Ireland, 2013.

[4] K. S. Lee, H. Wang and H. Weatherspoon, "SoNIC: Precise Realtime Software Access and Control of Wired Networks," in *The 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, Lombard, IL, 2013.

[5] R. Surton, K. Birman and R. van Renesse, "Application-Driven TCP Recovery," in *The 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Budapest, Hungary, 2013.

APPENDIX

BOOM 2012 Projects

Detection of DDoS Attacks Using Gossip

Summary:

A gossip-based monitoring system to detect DDoS bandwidth depletion attacks by using statistical analysis of network bandwidth usage in egress and ingress links of each node in the network. The program can abstractly be seen as an implementation of an overlay network with a gossip-based monitor to oversee network flow between nodes. MiCA is used to implement the networking portion of the system. Each peer node runs its own monitoring system and actively gossips its health status to other nodes in the network, and handles requests for other nodes to either confirm or deny whether they believe that a particular peer node is under attack. A GUI interface displays current data on network usage in relation with past collected data, as well as information of suspected attacks underway (if any).

Faculty Advisor: Ken Birman

Presenters: Vera Kutsenko, Olson Jaimes Carrillo

Distributed Cache using Cornell's new MiCA gossip programming language

Summary:

Memcached is a distributed in-memory cache that stores key-value pairs for rapid lookup. Create a gossip system that helps memcached nodes coordinate by speculatively caching popular keys and evicting unpopular ones.

Faculty Advisor: Ken Birman

Presenters: Sharvari Marathe, Liyaqatali Nadaf

Achieving Scalable Landmark Recognition on the Cloud through Isis2

Summary:

Our project allows the user to upload a photo containing a landmark to our website and receive back the location of the landmark in the world. At BOOM 2012, we will provide a demo of the current system running on Amazon EC2.

Background:

Professor Snavely's Computer Vision research on Context-based Image Retrieval (CBIR) has resulted in his new Co-occurrence RANSAC Landmark Recognition algorithm which improves the accuracy and speed of existing Landmark Recognition approaches. The problem is the speed of CBIR algorithms – they are computationally expensive in nature, since a search must analyze the contents of an image and perform matching against a very large in-memory landmark corpus. The solution, which is our M.Eng project, is to transform the algorithm to run on a distributed system hosted on Amazon's EC2 cloud. To build the distributed system, we are using Isis2, a new and exciting cloud computing platform developed by Professor Birman. A distributed system is a solution to our problem since one can scale computation and memory resources as needed while taking advantage of parallel image processing.

In our team of 4 M.Eng students, our goals are to build a system that is reliable under load, handles a large number of concurrent requests, provides scalability of nodes and offers a low response time. Our targets are: serve up to 100 requests/second, use up to 100 EC2 nodes, and have a response time under 5 seconds. We also will build automated deployment & process control and testing frameworks. We plan to publish a paper on our experiences, architectural decisions and roadblocks encountered on building a large scale landmark recognition system using Isis2 on Amazon EC2.

Faculty Advisors: Ken Birman, Noah Snavely

Presenters: Scott Phung, Kaushik Nataraj, Shivendra Singh, Abdelrahman Kamel

Timber

Timber is a distributed, scalable cloud logging application, running as a service. It is written in Python to receive requests and communicate internally using a gossip-like protocol. It uses vector clocks to maintain partial ordering of events. Users have access to a public application programming interface (API) which asynchronously logs and persists critical data. Users can then poll the service to get back data about their application's performance and stability.

Faculty Advisor: Ken Birman

Presenter: Chet Mancini

GridControl: A Software Platform to Support the Smart Grid

The transformational control and power deployment concepts that will characterize the next generation of the power grid share a common weakness: the most exciting concepts are so far beyond the norm for "cloud computing" that they would force every power research team to become experts in high-assurance distributed computing. Cornell University and Washington State University have joined forces to create GridControl, a powerful and comprehensive software platform that will slash the time and difficulty required to prototype and demonstrate new smart-grid control paradigms.

Power systems developers will employ GridControl as a tool that simplifies their most challenging problems. It will include new architectural options for power systems monitoring, management and control, and overcome the diverse technical hurdles of cloud computing in real settings. Our effort focuses on putting well understood, working technologies into the hands of development teams seeking to explore innovative power grid control concepts. We argue that such efforts to date have been hobbled by inadequacies in the most common, widely available, production technology platforms. We see ourselves in a role of having fixes for those shortcomings and solutions that already demonstrably bridge this how-to gap.

Faculty Advisors: Ken Birman and Robbert van Renesse

Presenters: Ali Goheer, Marcus Lim, Sean Ogden, Ashik Ratnani

Peer to Peer MapReduce

The project would showcase how peer to peer architecture can be used to distribute the high workload for more efficiency and throughput.

Faculty Advisor: Ken Birman

Presenter: Rudhir Gupta

Cloud RAM

The Cloud RAM project is an implementation of an interface that allows an operating system to request random access memory (RAM) from a cluster located elsewhere in a network rather than from its available local RAM. It applies distributed computing strategies to achieve high assurance, and dynamically and seamlessly manages both internal and external use of RAM memory for the client. This type of interface is useful in data centers where network speeds combined with fast RAMs can be greater than that of hard drive accesses, and where processes may benefit from extensive RAM usage.

Faculty Advisor: Ken Birman

Presenters: Jose Rosello, Suryansh Agarwal

BOOM 2013 Projects

GE Software Cloud Execution

Developing a framework to optimize the cost and execution time of MAPS on a Linux cluster. The idea is to implement map reduce jobs on hadoop framework in Amazon EC2.

Faculty Advisor: Robbert van Renesse

Presenters: Vivek Sharma, Nishant Patel

Live Distributed Objects with Isis2

This project aims at creating robust and wide range of distributed applications using the power of Live Distributed Objects and Isis2. The Live Distributed Objects provides flexibility and easy interface to create applications while Isis2 provides robustness.

Faculty Advisor: Ken Birman

Presenters: Herat Gandhi, Maneet Bansal, Rakesh Chenchu

Porting Isis2 to C++

Faculty Advisor: Ken Birman

Presenters: Sunil Channapatna Ravindrachar, Aniket Dash

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

ACM	Association for Computing Machinery
AFRL	Air Force Research Laboratory
API	application programming interface
ARPA-E	Advanced Research Projects Agency-Energy
AWS	Amazon Web Services
BOOM	Bits On Our Minds
CBIR	content-based image retrieval
DARPA	Defense Advanced Research Projects Agency
DHT	distributed hash table
DMA	direct memory access
EC2	Amazon's Elastic Compute Cloud
FPGA	field-programmable gate array
GbE	gigabit Ethernet
Gbps	gigabit-per-second
GENI	Green Electricity Network Integration
I/O	input/output
IPMC	Internet Protocol multicast
NASPInet	North American SynchroPhasor Initiative Network
NIC	network interface card
NSDI	Networked Systems Design and Implementation
PCIe	Peripheral Component Interconnect Express
PMU	phasor measurement unit
REU	Research Experience for Undergraduates
SIFT	scale-invariant feature transform
SoNIC	Software-defined Network Interface Card
SOSP	Symposium on Operating Systems Principles
TCP	Transmission Control Protocol