

Report of Independent Study  
Matthew Kelly <mkelly@cs.odu.edu>  
July 29, 2011  
Old Dominion University Summer Term, 2011

I. Introduction

During the Summer 2011 term, I worked on and investigated various projects with the primary motivation being to investigate future project through exposure to many research ideas. A brief description of each of these projects follows with more detailed information as well as the worked performed for each of these projects in subsequent section of this document.

II. Project List

- a. **Archive Facebook** is an on-going software development project for the Web Science and Digital Libraries research group at ODU (<http://ws-dl.blogspot.com/>) to which I was first introduced in Spring 2011 semester. The group had originally developed a Firefox browser add-on to allow a user to archive his/her Facebook profile information with archival intentions, per the group's motivation.
- b. **Motes Investigation** is an exploratory precursor to further research in which TinyOS-based embedded system devices were used to fulfill certain requirements relating to sensors and intelligent ad-hoc networks.
- c. **Tmix** is an extension onto the network simulation package "ns" that allows for the replication of data read from real network traffic to be abstracted into software and readily replicated. The extension as originally created for the older ns-2 project but is in need of being adapted to the more modern and structure descendent of ns-2, ns-3.
- d. **VAST** was a problem solving competition (<http://hcil.cs.umd.edu/localphp/hcil/vast11/index.php>) affiliated with IEEE relating to

visual analytics used for problem solving complex scenarios as documented by supplied data.

- e. **USRP Investigation** is an exploratory project wherein devices that were purchased by ODU and have high potential for future research project require pre-research to be completed on them to allow them to be effectively used.

### III. Student's Contribution to Projects

*Archive Facebook* was the first project for which I did work, having volunteered to investigate and correct the project after it was known to no longer function in Spring 2011. In March 2011, having never been exposed the architecture of browser add-on/extension developing, I investigated what needed to be done to bring the project back to a functional state and learn and document why it broke in the case of the add-on needing to be repaired again in the future. ArchiveFacebook's original implementation was done by Carl Northern, another ODU graduate student, who extended a different Firefox add-on, Scrapbook, to account for the dynamics of Facebook where it failed to successfully archive. Per Carl, 99 percent of the code in Archive Facebook is Scrapbook code. This fact also meant that Carl or any of the previous developers of Archive Facebook were not familiar with 99 percent of the inner working of the add-on to which their names are attributed, so much of the investigation as to why it was broken needed to be original research. Having a large deal of experience in Javascript development, this task was not completely daunting but instead allowed me to use my experience to rewrite much of the original Archive Facebook code to make it be applicable to the task it needed to perform as well as strip away the parts of Scrapbook that weren't applicable to the Archive Facebook project.

Programming against a live website through the means of scraping is a frustrating endeavor, as a change in the website tends to break anything that was based on it. This was the reason for Archive Facebook's lack of functionality when I took the project on. The newer version of Scrapbook used a

more efficient local storage means, so first I acquired the Scrapbook source for a basis. While doing this, I also acquired the version from which Archive Facebook was based and diff'd it against the Archive Facebook code to observe the small amount of code that was originally added to Scrapbook to create the end-product, Archive Facebook.

As the basis of the original implementation had long since been improved, I scrapped the entirety of the original implementation and proceeded to code the new version of the plug-in with the more robust storage engine in the newer version of Scrapbook. While creating original code, I also stripped away parts of Scrapbook that would be unnecessary to the niche functionality of Archive Facebook so as to reduce the memory footprint of the end result. The first release of Archive Facebook was about 70% original code and 30% Scrapbook code. Since the original release, I have released various updates to keep up with the ever changing Facebook layout.

In late June 2011, I attended the National Digital Information Infrastructure and Preservation Program and National Digital Stewardship Alliance (NDIIPP/NDSA) joint Partner Meetup in Washington, DC. There, in addition to being exposed to much more than I realized was being done in the field of archiving, I also presented the Archive Facebook project to a group of participants whose focus was web archiving. After returning, I wrote a blog post on the WS-DL blog at <http://ws-dl.blogspot.com/2011/07/2011-07-25-ndsandiipp-partner-meetup.html> consisting of a comprehensive summary of the Meetup as well as highlighting the projects presented from Old Dominion University.

The second project that I worked on during the term was an independent investigation of motes. Motes are TinyOS-based programmable wireless devices that have been loosely explored in various other projects at ODU. As I hoped to work with them further in a future semester, I became familiar with how to program the devices, the structure of the environment in which to do so and the dynamics of the mote/TinyOS architecture. No concrete goal was establish beyond extensive exploration

with little guidance (a good thing: I was free to explore) and the project's wiki for reference. I gathered various materials from past projects including an implementation done for a Networking Sensors class that occurred in a previous semester that modified the base (Deluge) framework.

There was little end-product for the motes investigation, being mostly exploratory, but one tool I did create to ease development was a Graphical User Interface (GUI) via Python that allowed for simplification of the overhead usually needed to program a mote, as the current TinyOS interface with the particular motes is buggy. Beyond the overhead, I also implemented a means to fetch the various version of the Deluge framework that had been developed (the base Deluge and the one modified in the Sensors class) and be able to swap them out of the environment on-the-fly with a click of the button in the GUI. Before this, this was a tedious task that risked corrupting the development environment if a developer was not careful.

The main reason beyond curiosity to explore this platform is for a future research project with Dr. Weigle that I hope to be involved in Fall 2011. Prior familiarity with the motes will reduce the learning curve required to perform the tasks outline in the already-accepted proposal.

The third project for which I did work this Summer was the *Tmix* project. Tmix is an extension onto a network simulator, ns-2, that allows one to generate realistic traffic on a simulated network. As of this document's writing, I am still actively involved in this project. My tasks for this project were extensive, ranging from validation of results obtained from the latest build of software as compared to a real-traffic basis to integration with the latest version of the simulator, ns-3, which uses a completely disjoint framework than ns-2. Much of the time interacting with this project was in learning ns-2 and tmix's workings, which required significant overhead but because the package has been used in other projects at ODU, the overhead was necessary, as familiarity with the package will benefit future projects.

The fourth project in which I was involved over the Summer semester was the **IEEE VAST Challenge**.

Kalpesh Padia, another student at Old Dominion, worked on this project for his independent study for the Summer semester. The project was broken into three smaller parts and one grand challenge. Dr. Weigle advised that in the time frame in which I would work on the project that I only attempt to tackle one mini-challenge while Kalpesh another. I was added onto the project much later than Kalpesh, with him having about a month and me two weeks. During the two weeks my mission was, given a large corpus consisting of over four thousand news articles, determine the where, when, who, what, how and why of an imminent terror threat. As I have a very brief understanding of Machine Learning, quite a bit of research of ways to accomplish this task needed to be investigated. The problem descriptions can be seen at <http://hcil.cs.umd.edu/localphp/hcil/vast11/index.php/taskdesc/index> .

My approach was meta in that I created a tool that allowed one to investigate the corpus through work selection and TF-IDF processing, grouped articles together with a custom similarity algorithm and allowed a user to create a sub-collection of article to be used as evidence to support his/her claim. I then “ate my own dog food”, established evidence for my hypothesis, wrote a report about my hypothesis (<http://matkelly.com/projects/VAST2011/> ), created a video describing my tool’s usage (<http://matkelly.com/projects/VAST2011/video/index.html> ) and submitted it for the competition.

As this was ODU’s first participation in the competition and we had a very limited time frame to work on, I expected more benefit from the feedback rather than to receive any award. The results of the competition were returned to the participants in late July, 2011. Three reviews of various categories were qualified. Both my and Kalpesh’s results were:

|                          | <b>Matthew Kelly</b>       | <b>Kalpesh Padia</b> |
|--------------------------|----------------------------|----------------------|
| Clarity of Explanation-1 | Average                    | Average              |
| Clarity of Explanation-2 | Good                       | Average              |
| Clarity of Explanation-3 | Good                       | Good                 |
| Threat Accuracy-1        | Inaccurate                 | Inaccurate           |
| Threat Accuracy-2        | Inaccurate                 | Accurate             |
| Threat Accuracy-3        | Accurate                   | Inaccurate           |
| ThreatDetail-1           | Little Detail              | Inaccurate           |
| ThreatDetail-2           | Detailed                   | Inaccurate           |
| ThreatDetail-3           | Detailed                   | Inaccurate           |
| SupportingDocuments-1    | A Few Correct Documents    | Inaccurate           |
| SupportingDocuments-2    | Majority Correct Documents | Inaccurate           |
| SupportingDocuments-3    | Majority Correct Documents | Not Supplied         |
| Visualizations-1         | Marginal                   | Marginal             |
| Visualizations-2         | Good                       | Marginal             |
| Visualizations-3         | Marginal                   | Average              |
| Interactions-1           | Marginal                   | N/A                  |
| Interactions-2           | Average                    | Marginal             |
| Interactions-3           | Average                    | Marginal             |
| Novelty-1                | Marginal                   | Marginal             |
| Novelty-2                | Moderate                   | Marginal             |
| Novelty-3                | Marginal                   | Marginal             |
| Overall Rating-1         | Marginal                   | Marginal             |

|                  |         |          |
|------------------|---------|----------|
| Overall Rating-2 | Average | Marginal |
| Overall Rating-3 | Average | Marginal |

Though absolute scores for this competition were not released, I believe that given the abbreviated timeframe as compared to a student with much more time allowed, likely more proficiency in the topic area and VAST being his primary project, the above results were sufficient for a first-time entry. It should be noted that many of the other submissions had as many as five people on a single team where, because Kalpesh and I worked alone, were essentially two separate one-person teams.

The final project in which I was involved in did not get started until nearly the end of the independent study term due to the late arrival of required hardware, the *USRP Investigation*. The investigation, as guided by Dr. Tamer Nadeem, began around early to mid July with the exploratory mission of "Figure out how they work, what their capability is and how we go about using them for further research projects." Results for this project were very limited due to the large quantity of other project (i.e. the above) that were being completed at the same time; however, the intention of this project is for further work with Dr. Nadeem and the devices in the Fall. I was able to setup the environment to interact with the USRPs, which is appeared to be a blockade for many given the documented experience of others. Further, I was able to configure the devices, actively communicate with them and obtain basic readings of the airwaves for very low frequencies. As this is an on-going project, the devices are currently being explored for potential research applications.