June 24, 2010

Dear Colleague:

I am pleased to inform you that the Air Force Office of Scientific Research (AFOSR) published an article last week on their web site of our research at the University of Texas at Dallas on secure cloud computing http://www.wpafb.af.mil/news/story.asp?id=123209377.

In preparing for this article, I answered several questions posed by AFOSR.

I have attached the questions and my answers.

Please let me know if you have any questions on our work.

Sincerely

Bhavani Thuraisingham

Dr. Bhavani Thuraisingham
Louis A. Beecherl, Jr. I, Distinguished Professor
Department of Computer Science
Director of the Cyber Security Research Center
Erik Jonsson School of Engineering and Computer Science
Box 830688, EC 31
The University of Texas at Dallas
Richardson, TX 75083-0688
Email: bhavani.thuraisingham@utdallas.edu
Voice: (972) 883-4738
Fax: (972) 883-2349
http://www.utdallas.edu/~bxt043000/
1. What was achieved by your AFOSR-funded research?

First let me say that AFOSR took the lead on assured information sharing even before the 9/11 report was published. The report states that we have to migrate from a need to know to a need to share paradigm. This is easier said than done. Sharing is not something that comes naturally to most humans as we are competitive by nature. Our challenge is how do we share securely and yet maintain our autonomy? Our first 3 year project was a small one funded in 2005 by AFOSR. At that time the DoD had not yet published its information sharing strategy document. Therefore, we took a rather simplified approach by assuming that we have only three types of partners: trustworthy, semi trustworthy and untrustworthy. Subsequently we developed three sets of solutions: policy-based information sharing, game theoretic strategies for semi-honest partners, and data mining strategies for malicious partners.

I call this first project “the mother of all projects” as it led us to several other projects funded by multiple agencies (e.g., AFOSR, NASA, NGA, IARPA and even NSF and NIH). It also gave us a lot of insights into the MURI BAA that came out in 2007 about the time Hon. Mr. Grimes, the then Assistant Secretary of Defense, had published the DoD’s information sharing strategies. We assembled a six university team and put a proposal together that aimed to develop breakthrough technologies to address the strategies. We were fortunate to win this MURI award, and we at the University of Texas at Dallas (UTD) are focusing on solving two main problems in this project. The first one is on incentives for information sharing. We believe that unless there are sufficient incentives, it will be difficult to get people to share information securely. I am pleased that we are tackling this hard problem and we have brought in an interdisciplinary team of researchers both from our School of Management and our School of Economics, Policy and Political Sciences. Technology alone is not sufficient. We truly need an interdisciplinary approach for this. For example, proper economic incentives need to be combined with secure tools to enable assured information sharing.

A second area we are focusing on, and this one jointly with Purdue University, is what is called confidentiality and privacy preserving ontology alignment. Information integration is a major challenge when sharing. Different agencies may use different representations of the same data. This is especially a problem in terrorist watch lists. For example, the late Senator Edward Kennedy was stopped several times at the DCA airport as his name was probably in some watch list. This problem is called semantic mismatch. This has been around a very long time, but today with the use of ontologies (e.g. Common representation of an entity by a community), partial solutions to the problem are possible. However, different communities have different ontologies for the same entity. Therefore, aligning the ontologies so that the differences between them are resolved is a major challenge. In fact, we are completing a project funded by IARPA to solve this problem, but it does not take into consideration security and privacy issues. Therefore, for AFOSR, we are developing algorithms for confidentiality and privacy preservation for ontology alignment. Our approach leverages what we have done for one agency to another agency.

Under a different but complementary AFOSR project we are investigating security for cloud computing and here again AFOSR is taking a leadership position. Since this is such a complex and vast topic, we are taking a bottom up approach and solving manageable problems. We are using Apache Hadoop and Google’s MapReduce to build the cloud and investigating security for (i) storage (ii) Hadoop and (iii) query processing. For example, we are applying in-line reference monitoring techniques to efficiently and reliably enforce XACML security policies for Hadoop, and on top of that building secure federated query processing strategies. Our secure query processing strategies will be demonstrated in September of this year and the XACML implementation will be demonstrated next year.

A lot has happened during the past 3 years in assured information sharing. Different agencies now have strategies for information sharing. Well known are the efforts of the Department of Justice. There is even a Justice information sharing initiative. Therefore, we are also keeping up with all the developments and conducting surveys. While the area is extremely challenging, the good news is that this is truly a multiple-use technology. Once we have the framework and incentives in place, then we can tailor the solutions for multiple agencies. I am really glad that AFOSR had the vision to fund such an initiative.
2. How are you accomplishing your research goals? For example, are you using any special tools or cutting-edge methods?

Incentives and Information Sharing: We developed a framework to understand how security and incentives play a part in assured information sharing. Our results indicate that proper incentives combined with good security mechanisms allow different organizations with potentially conflicting goals to securely share critical information. For example, monetary incentives based on each organization’s contribution to the final success of some common task can be very useful in encouraging assured information sharing.

We have developed an evolutionary game theoretic framework to simulate different assured information sharing scenarios. With respect to incentives and risk in assured information sharing, we deploy different tools from the fields of risk management and economics. For example, we are using credit risk modeling tools to understand and measure risks involved in sharing sensitive information. In addition, we are using mechanism design tools from the field of economics to understand incentive issues in assured information sharing. Working with social scientists, risk analysts and economists has helped us a great deal to understand how people think and behave in the real world.

The work we have carried out with our multi-disciplinary team for AFOSR has also supported us in a number of projects for other agencies. We recently had a small grant from ONR and we used our knowledge of risk modeling and analysis for botnet detection and management. We also prepared a white paper on adversarial modeling with game theoretic strategies and ARO has invited us to submit a full proposal. We believe that in order to understand humans including terrorists and hackers we need to bring social scientists and psychologists into our project.

Confidentiality preserving ontology alignment: Ontology alignment is important for information integration and sharing. The problem is, given a set of ontologies, ontology alignment will identify matching concepts across these ontologies. Under our IARPA funded project, we have developed algorithms for both manual as well as automated ontology alignment and have conducted extensive experimentation. In particular, we have integrated our algorithms into an Intelligence Community Platform called Black Book. Using such a platform we can demonstrate the algorithms to our customers.

However, our IARPA research does not focus on confidentiality and privacy. Therefore, under our MURI project, together with Purdue, we are investigating an approach based on the path difference among concepts in the ontology combined with approximate confidentiality-preserving matching techniques that the team has developed. Path difference consists of comparing the path leading from each concept to the root of the ontology. The more similar these paths are, the more semantically similar the concepts are. As these paths can be compactly coded as strings, by adopting, for example, some numeric identifiers for concepts, they can replace the corresponding values in the original records. These modified records can then be matched by using the approximate confidentiality-preserving matching techniques. This matching technique securely computes the distances between pairs of records and returns only the records whose distance is below a certain threshold.

Secure Cloud Computing: In our work, we focus mainly on designing and implementing fine-grained access control for managing large amounts of data securely in a cloud environment. To achieve our goals, we are combining technologies such as inline reference monitoring, query re-writing, and encrypted data storage in a novel manner. For example, query rewriting is used to enforce access control policies when the data is processed using high level cloud query languages. For programs written in languages such as Java, we use inline reference monitoring to enforce fine-grained, history-based security policies, such as Chinese Wall policies, which control access to information based on a history of information already accessed. We use encrypted data storage for securing data at rest. We have also introduced the notion of secure-Co-DBMS in a cloud that can carry out security critical functions even if the data management system may get corrupted. Finally, we have implemented some prototype systems that can store sensitive data on Amazon cloud infrastructure in an encrypted form and query it securely using Black Book.
3. What are the challenges of your research and how do you plan to meet them?

**Incentives for Information Sharing:** One of the major challenges in conducting interdisciplinary research is to learn the assumptions and terminology used in other fields. Therefore, we spent considerable amount of time describing our research problem as well as learning from others. We tried our best to abstract the problems in such a way that our colleagues in management and economics sciences can relate to the problems. Once we understood where the other party was coming from, we managed to develop a comprehensive framework for understanding risk and incentive issues in assured information sharing. One important challenge that is remaining is to see how our framework operates in practice. Basically, we plan to run some experiments using online social network applications such as Facebook to see how various security and incentive measures affect information sharing. We are expecting to demonstrate our tool in September to AFOSR.

**Confidentiality-preserving Ontology Alignment:** In order to ensure confidentiality, we have to first examine the various ontology alignment algorithms. We have developed algorithms for manual as well as automated ontology alignment. The manual ontology alignment algorithms need the human to align the ontologies. While there have been some efforts on this, we were the first to integrate these algorithms into Black Book. Our automated algorithms, as well as the algorithms of others, have low accuracy. Therefore, our challenge is to come up with algorithms with more accuracy. We are convinced now that we need semi-automatic algorithms where we have a combination of both automation and the human for ontology alignment. In summary, our challenge is to develop tools that can align ontologies semi-automatically with high accuracy and at the same time ensure confidentiality.

**Secure Cloud Computing:** One of the main challenges in implementing secure clouds is to integrate multiple security technologies, both existing as well as the new technologies, and yet maintain security of the overall system. In addition, cloud computing is increasingly being used to process large amounts of semantic web data. Some of the existing techniques need to be modified to be applicable for semantic web data as well as to ensure security. We are addressing these challenges by carefully considering potential use cases for multiple domains.
4. **How will this technology improve current AF operations or the AF's ability to achieve its mission?**

Daniel Wolfe (formerly of the National Security Agency) defined assured information sharing (AIS) as “providing the ability to dynamically and securely share information at multiple classification levels among U.S., allied and coalition forces.” The DoD’s vision for AIS is to “deliver the power of information to ensure mission success through an agile enterprise with freedom of maneuverability across the information environment.” Our research attempts to address this challenge.

What we believe that our technology will provide is a framework for information sharing and policy enforcement and address some of the major challenges including secure information integration and also give the managers an understanding of the incentives for sharing. Even if we have the technology, how can we promote the culture of sharing? This is a major challenge that we hope our work on incentives will give some insights. Tim Berners Lee, who is the inventor of the semantic web, has stated that his vision is to develop a web with little or no human intervention. By using semantic web technologies such as RDF (Resource Description Framework) and Ontologies, we believe that the framework we are developing will be evolvable and flexible. Therefore, we hope to plug and play different types of policies and incentives as needed for information sharing. The solutions we are implementing for a secure cloud can also be integrated into this framework. This will enable the coalition partners to share information in a cloud securely.

We are monitoring closely what the DoD and AF are doing with respect to mission assurance, network centric enterprise services (NCES) and the global information grid (GIG). At present, there is a lot of emphasis on attribute-based access control which our framework can support. If the access control mechanism changes to some sort of Usage control then our flexible framework can accommodate such changes. Furthermore, NCES is using XACML for policies. Should they decide to use RDF-like languages, our flexible framework can support this as well. Essentially the various functions (e.g., knowledge management, policy management, incentives management) are implemented as services that can be turned on or off depending on the need. Furthermore, we are discussing opportunities to present our work to AFRL and DISA so that the technologies such as XACML implementation in Hadoop can be transferred to the GIG.
5. Who else was involved in getting you to the particular milestone? (collaborators, grantees, other DoD agencies)?

We are a team of four professors working closely with universities both in US and Europe, commercial industry, government labs, other agencies, AFCEA, Standards organizations and our students to accomplish the vision of Daniel Wolfe on assured information sharing. The professors are myself, Latifur Khan who joined us from the University of Southern California, Murat Kantarcioglu who joined us from Purdue and Kevin Hamlen who joined us from Cornell. Our MURI project is carried out in collaboration with the University of Maryland Baltimore County (UMBC) who is the lead, Purdue University, University of Illinois at Urbana Champaign (UIUC), University of Michigan (UM) and the University of Texas at San Antonio (UTSA). While UTD’s main focus is on incentives and secure ontology alignment, we are collaborating with all of our partner universities on some aspect of the information sharing problem. For example, with UMBC we are working on policy management with semantic web-based infrastructures. With Purdue, we are collaborating on a number of topics including confidentiality-preserving ontology alignment. With UIUC, we are working on examining data quality and provenance for integrating information. With UM, we are working on secure social network aspects. With UTSA, we are working on gathering requirements from agencies. In our work on secure cloud, we are collaborating with Prof. Elisa Bertino of Purdue on a piece that deals with accountability for cloud computing which we will then integrate into our XACML implementation of Hadoop.

Our goal is to select pieces of the technology and have an initial prototype demonstration for the Air Force in September of this year and more robust demonstrations in subsequent years until the project ends in April 2013. We are also encouraged by the fact that AFOSR (EOARD) has funded our collaborators at Kings College, University of London and the University of Insubria in Italy. I am hoping that we can have a joint demo between US, Italy and UK sometime next year. As I have mentioned, we have projects with multiple agencies on different topics. Therefore, if we can use something we have developed for one agency to enhance the work for another agency then we can save money and not reinvent the wheel. I gave an example earlier about the work we have carried out for IARPA which we are using for AFOSR. NGA has also benefitted from some of our research for IARPA. Similarly the work we are doing for AFOSR is being used for our work with ONR and ARO.

Working with AFRL is a major goal for us. I visited AFRL researchers in 2006 during the mid-term of our previous grant from AFOSR to present the results to the researchers. We are planning to do the same for the MURI project this summer. We would like to get their feedback and also install our systems at their labs. We are especially interested on getting feedback on our research for incentives which we believe will be a major breakthrough in our MURI project. We have also formed close partnerships with corporations such as Raytheon, IBM, Rockwell and Lockheed and are trying our best to transfer our research to their operational programs. We have been quite successful in the work we have done for IARPA on semantic web. Some of this research is being used by Raytheon on their Intelligence programs. The program manager from Raytheon visits us every three weeks to discuss our progress. Later I will mention what we are doing in the DFW area to promote collaboration among the commercial industry.

I teach at AFCEA (Armed Forces Communications and Electronics and Association) two classes: one on Knowledge Management through semantic web and social networking and the other on Data Mining for counter-terrorism. Information sharing is a unit in both my classes. The students are mainly from the DoD. Therefore, I get valuable feedback from them. Through AFCEA, I have also taught at various AF bases including Offut, Lackland, Edwards, Kirkland and Eglin. I get feedback from the AF personnel on their requirements which helps our work a great deal. We also give seminars at several universities and keynote addresses at various conferences. The past few years I have mainly discussed assured information sharing at these meetings to make the research community aware and to get their feedback. Last but not least, our students are our most precious asset. Later on I will discuss how we use our education programs as well as standards efforts to advance the vision of the DoD.
6. What is the vision for this individual program?

I have two visions: one for the work we are doing at the University and one for the MURI project. For the MURI project, if we can get an organization like AFRL or DARPA to issue a BAA and have a system integrator to build the Information Sharing System for the DoD with pieces of our research inserted, then I will be very pleased. In fact, I spent 13 years at MITRE before joining NSF for 3 years and then came to the University in October 2004. At MITRE I learned that while fundamental research and making breakthroughs are important, research also has to be useful to the customer. For example, I proved that the general inference problem was unsolvable back in 1990 and then examined solutions for special classes of the problem for SPAWAR and CECOM. I was also involved in AWACS experimental research program at MITRE and some of our research went into the systems built by Boeing and Lockheed for the AF. We are attempting to do something similar at the University. The systems we build are based on fundamental principles such as mechanism design for assured information sharing and formally certified in-line reference monitors for provably secure, high-assurance XACML/Hadoop implementation.

For our university research on this particular program, I have two major objectives. One is to solve the challenging secure ontology alignment problem so that we do not have semantic mismatch. If we can accomplish this, it will be a major breakthrough. Regarding the incentives, I really want to make an impact on AF managers so that they can understand what incentives to provide to their personnel for assured information sharing. I believe that secure cloud will be the next big thing in computing, but securing the cloud is such a complex and challenging problem, and it is difficult to do it all. I believe that what we are doing will help solve some aspects of the problem that relate to assured information sharing in a cloud.

When I teach classes at AFCEA, students have mentioned the following: “we have no incentives to share. In fact, we are not penalized if we don’t share. But if we share the wrong information we are punished”. So what I would like to accomplish with the students who attend my AFCEA classes as well as others in the DoD is that information sharing is a good thing and if done carefully with proper incentives and policies in place then we can accomplish a great deal.
7. What are the next steps in advancing this technology?

To understand the practical implications and effectiveness of various assured information sharing technologies that we are developing, we plan to conduct several experiments over the next year. In addition to experimentation with Facebook, we plan to ask students to share information to solve puzzles and see the effect of various security measures on sharing. We believe that such experiments will enable us to better modify our tools for real world deployment. This is where we can benefit greatly by working with social scientists. As computer scientists we generally do not conduct surveys and develop systems based on surveys. This is an area that social scientists are very good at. Therefore, our partners in the School of Economics, Policy and Political Sciences would help us in devising the experiments as well as analyzing the experimental results.

Confidentiality-preserving ontology alignment is a new area. Therefore, what we need to do here is not only get our algorithms to work with high accuracy and maintain confidentiality, but we also need real world examples to test our algorithms. Organizations like IARPA are preparing data sets for researchers to use. We hope that we can have access to these data sets to test our algorithms. In addition, we would like to work with AFRL to enforce appropriate policies for the ontologies and see if our algorithms produce meaningful results.

For secure cloud computing, we are planning to contribute our code to the open source community to increase the adoption of our technologies. In addition, we plan to explore how to use public cloud computing services offered by corporations such as Amazon, with the private clouds used for storing sensitive data. We plan to explore how to outsource as much work as possible to public clouds without sacrificing the security of the sensitive data.
8. Why is this milestone significant, both in terms of Science and Technology (i.e. never been done before) and capability development (i.e. new or improved capability etc.)?

Assured information sharing solutions require various important components that deal with security, incentives and risk management to be integrated carefully. In our project, we are developing such an interdisciplinary solution that could potentially enable information sharing not only in DOD applications but also in other domains such as healthcare where security, privacy and conflicting interests can hamper information sharing. Second, information sharing cannot be carried out effectively without information integration. Ontology alignment with security is important for secure information integration. Our research on this topic will also be a significant accomplishment.

Educating our students both in terms of research and technology is critical for us as a university. Recently we received the NSF Scholarship for Service award which will start in the Fall of 2010. We are now recruiting students to take our MS track in information assurance. As part of this track, we teach Data and Applications Security and Assured Information Sharing is covered in several lectures. We include a lot of our research results from the MURI project in these lectures. These students will be US citizens and therefore, when they graduate they will work for federal agencies.

I am also very interested in our students joining organizations like AFRL to conduct research. I am encouraging my US citizen students to join government research labs. In fact, one of my colleagues will be applying for the summer program at AFRL for next year (2011) to understand their needs and this way we will be in a better position to send our students to AFRL. I strongly believe that understanding the customer requirements is very important for our research. This was instilled into me by Mr. Henry Bayard, my mentor at MITRE. Therefore, I take courses myself to improve my knowledge on areas important to my customer. For example, at present I am taking the Certificate for Terrorism Studies Course online at St. Andrews University in Scotland. This course has given me a very good understanding of how terrorists function and their ideologies. This would I hope enable me to provide more realistic solutions to our customers.
9. Why didn't we know/have this technology before (i.e. what obstacle was overcome to achieve this milestone)?

We knew about the problems for a long time. For example, back in the early 1990s when I was at MITRE, we discussed the semantic mismatch problem. At that time, we provided solutions that were hardcoded. That is, A in one system means B in another. However, with the advent of the semantic web by Tim Berners Lee in the late 1990s and the maturing of ontologies, we started using variable ontologies for representing information. Then when we started integrating information from multiple systems we still had to solve the semantic mismatch problem and so we started working on ontology alignment for IARPA around 2007. We felt that to share information securely we needed the secure ontology alignment problem solved. Therefore, I would say the obstacle that was overcome for this effort is the invention of the semantic web by Tim Berners Lee.

With respect to the incentives, I think that back in the old days we had stove piped systems and there was not much emphasis on sharing information. After the cold war and especially after 9/11 there was emphasis on information sharing. We realized that sharing information is not easy. We can have all the policies in place but if the culture does not permit one to share then there will be no sharing. Therefore, we brought together a group of interdisciplinary researchers. As we have stated earlier, it is not easy to work with people in other disciplines. Therefore, we had to make a huge effort to understand what the others were saying. Therefore, the second obstacle I would say we overcame was to understand how social scientists and economists think and work.

With respect to secure cloud, I believe that the development of the middleware by Apache and Google have been major breakthroughs. This has enabled us to do some interesting things with the cloud such as secure query processing, secure storage and policy design and implementation. If not for the work of Apache and Google, it would have been difficult for us to build a cloud in the first place.

Therefore, each aspect of our research has benefitted a great deal from what the others have done and are doing including Tim Berners Lee’s vision of the semantic web, the products of Apache and Google and the improved collaboration between interdisciplinary researchers.
10. What are the possible private sector uses of this technology?

The good news is that information sharing is not limited to just the DoD, DHS and the Intelligence Community. It is a problem for healthcare organizations, the Justice Department and also for financial organizations. As I mentioned, AFOSR was far-thinking in funding this type of work back in the early to mid 2000s. These days, we see calls for proposals from NIH as well as the DoJ to develop technologies for information sharing. For healthcare applications, privacy is critical. DoJ has come up with reference architecture and a model for Justice Information sharing and I am studying this effort to see if our AF project can benefit from this.

The private sector is also forming partnerships between corporations and information sharing would be quite useful. However, the private sector is also extremely concerned about company sensitive data. I see a role for us as a university to bring the companies together. Therefore, we were instrumental in forming a Defense SIG in the DFW area together with the Metroplex Technology Business Council and hosted the first meeting in April 2010. Major defense contractors such as Raytheon, Lockheed, General Dynamics and Rockwell are part of this group. We are hoping that we can enable these companies to share information for their mutual benefit and to benefit our customers.

We have also started a university spin-off company called Infosec Analytics so that we can transfer key pieces of the technology we develop to products. In fact, we are having discussions with the university to license the data mining technologies we have developed for malicious code detection under our prior AFOSR project.

Since I have come to academia after a career in the commercial industry, MITRE and the government (as IPA), I strongly believe that we have to combine theory with practice. Therefore, I place a lot of emphasis on developing tools and prototypes that are based on fundamental principles. We have developed a lot of software including open source software such as submissions to HP Lab’s JENA systems as well as developed multiple toolkits and repositories. We demonstrate our tools and prototypes to our sponsors and commercial industry partners and get their comments for improvement. Therefore, by continuing to do this I am hoping that our technology can be transferred not only to the DoD and the Intelligence Community, but also to other organizations (e.g., DHHS), commercial corporations including the financial industry.

Standards are also a means to transfer our technology to products. I had quite a bit of experience on standards when I was at MITRE. In the early 1990s, I was part of a Navy standards group specifying interfaces. Later in the mid to late 1990s, I was instrumental in founding the C4I working group with OMG (Object Management Group) and influenced the Real-time Special Interest Group at OMG with our work on the AWACS experimental research program. At the university, we are members of the Open Geospatial Consortium and we have given presentations on building a geospatial semantic web based on our research for NGA to OGC as well as at the World Wide Web Consortium meetings. This way I am hoping that the corporations will develop products based on the various standards so that our customers can benefit.

Finally, we get tremendous support from the Erik Jonsson School of Engineering at the University of Texas at Dallas. Through the university we have close collaborations with corporations like Texas Instruments, UT Southwestern Medical Center and Collin County Homeland Security and Local Law Enforcement. This enables us to explore new applications for our research in RFID-based information sharing, healthcare and law enforcement.