Research Statement

In support of his application for PhD in Computer Science for Fall 2015

Mithun Paul
Email: mithunpaul@email.arizona.edu

January 6, 2015

***Abstract*—Here I present the research projects I have worked on in the past, and some ideas of mine, which I want to explore in future.**

I. Introduction

My research interests are primarily in the Security field. In the past I have worked (some published and some patented) on sub fields like group key exchange protocols, security of smart phone communication , data masking,privacy preserving data mining etc. In the grad school (past 2 years) I worked with Dr.Christian Collberg on a privacy preserving data model for cloud. Even though most of my work has been in theory, lately I have been getting fascinated by systems, especially network security. For example I have been wondering about the possibility of having a hook or sniffer into the routers in the control plane of SDN and the data it would bring us for preventing botnet initiated DDoS Attacks. Other ideas I plan to work on are malware injection into distributed systems and virtualization (cloud), censorship monitoring, loopholes in Facebook's haystack etc. My other interest is in Machine Learning. I have worked with Dr. Sandiway Fong in creating some algorithms for capturing contexts in sentiment verification tasks. I invented a tool for Dr. Chris Impey which converts a flat data model, like Wikipedia, into a textbook like material, using machine learning algorithms. For the future I have been wondering about designing algorithms that learn patterns for spam malware and vector analysis. This paper elucidates most of these projects. Hopefully some more course work and a good research mentor will help me publish in these ideas of mine.

II. Previous Work

These are the research projects I have worked in the past.

A. Proof of Data Destruction in Software As A Service

This was inspired from the work of Shacham et al. on proof of retrievability [1]. The problem statement emanated from the thought whether HDFS has an ability to destruct the data completely. Classical file systems, based on Berkeley Fast File System, do not have this ability since data destruction is done by the cleaning/garbage collection mechanism of the file system process. Data is written either in buckets, or log file system models, with the latest edits, written to the end of a log or to a new location. Older locations are earmarked in the iFile or iNode as deleted and hence called a soft delete. This is why even though a file is deleted from a windows or a Linux file system, it can be retrieved until the cleaner hasn't passed through. However, this causes data vulnerability in cloud storage mechanisms. In the scenario of storage as a service, this becomes detrimental. In a SAAS scenario, after a particular customer ends his lease, the file system space gets reallocated to the next customer. As seen in the operating system example the data is not erased completely until the cleaner comes through. Suppose, within this time the file space gets reallocated to the next customer. So there is a possibility of the new user being able to probe into this data that already existed in the drive. Hence there is a necessity of a destruction mechanism incorporated into the operating system. This mechanism will probe into the data and ensures comprehensive data destruction. With respect to the Gutman algorithm, data needs to be rewritten 23 times to avoid deciphering remnants due to voltage left overs. So we opened up the code for the hadoop file system and incorporated the shred functionality. Nuances of this idea was published as [17][18][21] and filed as a patent also.

B. Data Masking

Data masking is a requirement in the scenarios where sensitive data is outsourced to 3rd parties for testing. First we developed a product, MaskIT [31] which sniffs into the network stream, does deep packet analysis, masks the sensitive data and replaces it back into the network. In case of cloud, the data masking gets trickier. Hence we presented an on-demand masking of data as a software service in a distributed environment. In this model first an application hosted on a computing device receives request for access to application data from a user. Credentials of the user are first validated in order to determine whether the user is authorized to access the requested application data. For an authorized user, a category of the user is determined to ascertain whether the user is privileged to obtain full access. In case the user is not a privileged user, application data access request is transferred to a data masking service. Application data is fetched from database utility, masked based on pre-defined masking rules and provided to the user. This was patented as [30]

C. Security in Blackberry Communication

There was a time when national security agencies used to complain big time that blackberry communication channels can be a good channel for covert communication, especially by malicious and terrorist like organizations. We decided to explore this claim and dug deeper into the communication and encryption mechanisms of blackberry, the enterprise service and the Internet service. We analyzed the entire communication channel and protocols and found that the threats and concerns are genuine. Further we suggested certain solutions based on escrow and public key infrastructure, using which the situation can be turned around in favor of both the security agencies and blackberry as a corporation who didn't want to lose their customers. This was published as [19].

D. Privacy Preserving Data Storage as a Service

Despite the economic advantages of cloud data storage, many corporations have not yet migrated to this technology. While corporations in the financial sector cite data security as a reason, corporations in other sectors cite privacy concerns for this reluctance. In this research, we proposed a possible solution for this problem inspired by the HIPAA safe harbor methodology for data anonymization. The proposed technique involves using a hash function that uniquely identifies the data and then splitting data across multiple cloud providers. We proposed that such a "Good Enough" approach to privacy-preserving cloud data storage is both technologically feasible and financially advantageous. Following this approach addresses concerns about privacy harms resulting from accidental or deliberate data spills from cloud providers. The "Good Enough" method will enable firms to move their data into the cloud without incurring privacy risks, enabling them to realize the economic advantages provided by the pay-per-use model of cloud data storage. This work is under review at the time of writing.

E. Non Linear Trust in Group Key Exchange for MANETS

Group Key Exchange allows a set of parties to communicate over a public network and securely establish a secret session key. Thus, it is a critical protocol for emerging network applications that require collaborated output from various peers. Many protocols have been proposed, designed and implemented for group key exchange. However, in all these protocols a scenario of linear trust is assumed i.e. either all the nodes in the group are uniformly considered trustworthy or completely trustless. In this work I presented a novel problem on how to deal with establishing a session key in a scenario where the trust is non-linearly distributed and a solution there of. A protocol is proposed using which every time a session key refresh happens in a group, it recognizes and eliminates the malicious node in the group thus securing every protocol session. This is best suited in highly dynamic scenarios like Mobile Adhoc Networks (MANET). This work was published as [22]

F. Teach Astronomy

This research was done with Dr. Chris Impey in the department of Astronomy, where we were wondering if we can develop some mechanism, which can convert a referential material to text book on the fly. We proposed a clustering approach to presenting information with the goal of giving the user a systematic learning experience. This represents a third approach alongside the "flat world" or undifferentiated model of information represented by Wikipedia and the highly prescriptive results returned by a Google search. Content items are clustered based on keyword overlap and are presented through a dynamic visual interface that lets the user explore closely related items. The method works on any text-based repository of content or an any visual repository (i.e. images, videos) that has ancillary descriptive text. It is also highly scalable and can be used on resources containing million of items or tens of gigabytes of text. This work is currently under review at the time of writing.

G. Ambiguous Movie Reviews

In this work with Dr.Sandiway Fong we tried capturing contexts using a Hilbert Space based model. The inspiration has been the classical work psentiment classification[32]and recent works by Socher et al [8]. Certain movie reviews give the reader a negative feeling about the movie, despite the machine learning algorithm based sentiment analyzer finding it to be a positive review. In this work we sought to find if such an ambiguity exists in some movie reviews, where there is ambiguity between a human review and a machine review. Further we explored the possibility of modelling a sentiment analysis mechanism from a quantum model perspective i.e can a model be prescribed for machine learning algorithms which can understand the underlying feeling and not the superficial feeling of the review. Work was done on training standard machine learning algorithms on positive and negative reviews first. During such a training phase ambiguous reviews were first earmarked as negative and then as positive. Then we made it test on such ambiguous reviews alone. Interestingly the machine learning algorithm goes haywire and classifies some as positive and some as negative in both the cases. Whereas these were based on the actual star based rating of movies which were given very high positive ratings. Further we explore the possibility of modelling based on a quantum mechanics perspective. There have been previous works showing how concepts and their combinations get influenced by the context. Here we try to take this a step ahead and investigate if this model can be used for modelling such specific cases of reviews found positive by machines and negative by humans. This work is under review at the time of writing.

III. Future Work

These are some ideas of mine which I want to work on sometime in the future.

A. Continuous Spaces

I want to extend Klein et.als work on continuous spaces [16]. For example in the colour picking task there is an assumption of spare and linear regression made. I think this is forcefully restricting the predictor into certain compartments and this might be the reason of the error in Japanese indigo-bright purple example. I want to explore if it will help if we consider a continuous aspect here also? Am particularly referring to a model similar to Socher et.al work [8] where they bring in a neural tensor based modeling to capture semantic meaning. I want to explore, in the syntactic local assumption in the predicting path problem in this work; could the meaning be better expressed if we introduce a similar mode? Thus we wont be bound by a preexisting structure assumption.

B. Time Evolving Trajectories

Also another idea I want to explore is the time evolving trajectory used in perceptual grounding conversions [16]. It reminded me of/was drawing parallels to time evolving Schrodinger wave equations in Physics. This is yet another idea I have been wondering whether a Hilbert space based model might be a bit more comprehensive in capturing the deeper meanings as a next step to tensor based modeling mechanisms. This will help in modeling these complex nonlinear interactions. A similar work (but in a different perspective) I bumped into is the one done by Aerts et al [9] who tries to model contexts based on a Hilbert space model and they seem to have been able to arrive at some results. Here they model concepts as a multidimensional wave vector and they say that the context collapses it to a particular axis when it is applied (like a Hermitian operator on a Hilbert space). Further works on this from the same author can be found here [12][13][14]. Similar to the general model projecting into other domains in [16], with this model, we can create a collapse of the multidimensional wave vector based on say, a color picking task, or financial news task. The same can be used for the word2vec [10] work also I think. In a way I think the distance approach for content similarity is flawed until we incorporate context into it. To quote a naive example, "Sun" must be close to "Egg" in the context "Sunny Side Up" but far away in the context of "Scrambled Eggs". So I was just wondering if we can consider this approach as the next level for tensor space modelling. I mean, can we incorporate somehow the context also into the tensor, thus aiding into the accuracy of the projection/collapse operator. This, I think, in a way encompasses/overlaps with the neural tensor networks and the feed forward Neural models [15][12][15].

C. Speculative execution in Distributed File System

This is in reference to Ed Nightingale's 2005 work [2] on speculative execution .A disadvantage that i find in this paper is that the speculations that they do are informed/educated guesses/based on pre-determined empirical values. These are based on pre-determined information that commits will always happen or writes will always happen. Where as if they take it a step further and use machine learning algorithms (Bayesian inference based) they can do an on the fly pattern learning. This methodology I think will enhance their speculations enabling them in much reduced time/much sharper/close to perfect guesses.

D. Finding Patterns in a Haystack

I have been wondering if the haystack [3], Facebook's photo storage, can be enhanced with a learning algorithm. I personally feel, that a weakness is in the protocol that they sell as their USP itself, via, the decreasing hierarchy of look ups. That still is a huge amount of communications/look ups in a way. First the client has to contact a directory, then a CDN, then a Cache and then a Store. I think the protocol where they look up first CDN and then the Store if CDN faults, and then the Cache if Store faults- I personally think this is an unnecessary amount of look ups. The counter argument could be that they complement this look up with a well-structured classification algorithm, using which they increase the chance of finding a photo in the first look up itself viz., the CDN etc. I think this can further be enhanced, and thus an addendum for future work, could be using a probabilistic look ups. As per the new protocol what am suggesting as soon as it gets a request, the directory takes a guess as to where it can find a particular photo. Say for example it takes a guess, based on prior knowledge that the photo is in the store (and it actually is in the store) and not cached anywhere above it, i.e. neither in CDN or Haystack Cache. Suppose it turns out that the photo is actually present only in store and not in any of the layers above (the word above is a big ambiguous, when the actual word should have been something that is before it in look up hierarchy) it that means that 2 look up calls can further be avoided, viz., the CDN first checking whether it has it and then saying a no (along with the overhead for stripping of the url- which I accept is not much, but small drops make an ocean). Then the Haystack Cache could have avoided looking it up and then transferring the control ultimately to the Store. This is similar to many of the earlier work on Speculative execution and we have huge number of works in the research field which proves that probabilistic and random variable based look ups.

E. Censorship

Another interesting topic I have been thinking about is censorship. Some very good reads about current state of the art and its effect on international politics is Paxson et al's work [6] and Feamster et al's Encore paper[7]. In [7] the authors use a twist on cross site scripting to collect censorship details. Though what am wondering is its stability on prefix hijacking/DNS poisoning attacks that could be started by the state. Also this does involve clandestinely incorporating a measuring device into a browser, inviting deployment hurdles. I wonder how much will it be feasible in Google letting you do it in chrome or the open source community in Firefox, left alone publish it only for url requests from sensitive countries to download. (Though I have wondered if having a perennially respawning virtual machine, which recreates itself randomly every time in a different server in a different continent, be a

solution to circumventing international laws on data storage). Though a place where Encore can score (or lose?) is the smartphone deployment. Note that in smartphones the concept of a browser is rarely used, in favor of the tcp connection a news reader app makes to the server-with or without TLS. If only we could incorporate a solution somewhere deeper in the OSI stack (and not int he application layer like the browser does in Encore).

F. Oblivious Computation For Smartphone Communication

So Yao's protocol is like this one stone pillar in 2 party or a multi party computation. However, it requires a huge amount of overhead in terms of network and processor time. As shown by Dr.Traynor in [5] its high time we wonder about low overhead based systems that do multi party computation. However I want to improve his work by exploring to reduce the overhead in his recent work about privacy preservation in cell phones. I want to explore, in place of the partial homomorphic encryption solutions that he uses, if its possible to use a much light weight solution I have suggested in one of my recent works. The argument I had proposed was that, for the privacy of data like social media data, yaos multi party computation might be an overkill. Instead we possibly might want to consider a good enough privacy solution, using a distributed hash table. There have been many recent works suggesting that privacy requires only a good enough protection- by that what I mean is that privacy preserving data is not super secretly sensitive that it requires yao's multi party computation or even homomorphic encryption [22] based solutions. Or if a government agency or an attacker is keen enough to spend millions trying to decrypt/factorize prime numbers trying to read your facebook posts, maybe then you might not want to call it private stuff. Hence a data segregation is paramount, and further privacy preserving data mining techniques can plug the hole very easily.

G. Resource Allocation in Multi-Cloud

While going through [33] I realized that, the solution I had proposed for my privacy preserved cloud work mentioned earlier, can be used for enhancing the algorithm mentioned in [33]. In my work I distribute the data across multiple providers using a distributed hash table approach. I want to use this solution to reduce the cost of the task allocation in multi-cloud scenario. Note that here its not dumb data we are distributing across nodes, but tasks/executables/processes/application allocation. This will be more relevant with a mobile user, where the bandwidth limitation can be solved using the proposed solution.

H. Detection of self spun web pages

I am fascinated by this [34] work of Dr. Voelker. I want to extend this work by bringing in a technique I had invented in one of my previous works [35]. Here the tool we created 'infers' from seed articles what the topic is and uses this knowledge to find more similar articles. This technique can be used for detecting automatically spun web pages. A collection of spun web pages can be used as a training material for this bayesian inference based algorithm, which can then segregate spun web pages from regular pages, thus aiding in the detection of such pseudo search engine optimization techniques.

I. Botnets

I found this recent work [36] on Spam Landscape very interesting. I find botnets to be a very fascinating study. I think they are, like the coral reef, a collective soul,and an ever mutating being. Very interesting a study per se, and further more interesting will be devising techniques to break it. I have been wondering if technologies can be developed where the role of a bot master itself can be voided. For example, dialect fingerprinting is a technique used to identify the host which talks to the mail transfer agent. I have been wondering if could this be replaced with a self modifying code[37] -a random generator which uses a different implementation of SMTP all the time. Techniques of code obfuscation [38] for malicious uses can come in handy here

IV. Implementation Projects

These are the projects where I did more implementation and less research.

A. Log Structured File System

Flash based devices require a log structured file system because of their limited over write capabilities. We created such a complete file system with file layer and directory layers, and their corresponding defragmentation/cleaning mechanisms. It had Fuse in the top most layer interacting with the operating system (unix in this case). The was completely implemented in C (with Gdb to help debugging). At the lowest level data was written onto a flash drive. This was done with Dr. John Hartman as our advisor for the course CS552 Advaced Operating Systems

B. OSPF software routers

In this project which was done as part of the course Computer Networks with Dr. Beichuan Zhang. Here I implemented a fully functional Internet router soft hosts that routes real network traffic. This project was built on Virtual Network Lab (VNL) which intercepts packets on the network, forwards the packets to the soft-hosts, receives packets from the soft-host and injects them back into the network. The soft-hosts were run locally by us as regular user processes and connect to the service via ssh tunnels. Clients, once connected to the server, were forwarded all packets that they are supposed to see in the topology. The soft-hosts were able to manipulate the packets in any way they wish, generate responses based on the packets, or make routing decisions for those packets and send the replies back to the service to place back onto the network. Here we implemented ARP, ICMP, and basic IP forwarding. The protocol used was a simple dynamic routing protocol, PWOSPF. When the soft-host would receive the packet on interface eth0, it will decrement the TTL, recalculate the header checksum, consult the routing table and send the

packet back to the service with directions to inject it back onto the network out of interface eth1. The router was able to create its own forwarding table automatically based on routes learned from other routers in the network. It was able to do this from the link-state advertisements sent from other routers and also was able to route traffic through a topology containing multiple nodes. Also these routers were able to detect when other routers join/leave the topology, and/or when links fail/recover, and correct the forwarding tables accordingly- using a heart beat mechanism. The project was demonstrated by performing trace routes, pings and downloading some files from a web server via your router.

C. NASA Data Crawler

This application I developed for Dr. Impey accesses the astronomy images from APOD (Astronomy picture of the day), a NASA repository [26]. The images are brought across the SOAP protocol in Json format, indexed using Lucene and fed into the textbook generator tool mentioned in section 1. The programming was done in Javascript, Java and C#.

D. Google Web Toolkit

In the website teachastronomy.com I replaced the existing flash based front end with one generated by a google web toolkit. The advantage of this was that the google crawler, which was not able to cross the flash barrier, could read this javascript based web page now and hence increase the page ranking in search engine optimization.

E. Scholarship Universe

I developed an application called Scholarship Universe [27] which dynamically matches students in my university to the scholarships they are eligible for.

F. Implementation of basic machine learning frameworks

As part of the course Introduction to Machine Learning, I developed tools , in Matlab, that implement various machine learning techniques including SVM, regression, clustering etc.

G. Security Attacks

As part of the course Computer Security I implemented Rootkits, RSA algorithm , PGP , buffer overflow attacks , Playfair, Homophonic, Polyalphabetic Ciphers, MATE attacks, SQL Injection attacks, packet sniffing and spoofing. Most of it was written in C, with GDB as the debugger.

H. DARPA BOLT project

I was part of the team which worked on DARPA's Broad Operational Language Translation. We ran machine learning algorithms and human experiments to find when a translated word has to be chosen to be rejected or given to the user for verification.

I. Software Defined Networks

Along with these resources [4] [5] I learned SDN (and cloud) from Dr. Larry Peterson along with learning about distributed storage and cloud. As part of the course Advanced Operating Systems I created Nagios as a service, supported by Cassandra , syndicate and docker.

J. Tokenization

As part of the course 'Natural Language Processing' I developed a tool that does tokenization of the brown corpus with Dr. Sandiway Fong . This was written in PERL and Prolog.

REFERENCES

[1] Shacham, Hovav, and Brent Waters. "Compact proofs of retrievability." Advances in Cryptology-ASIACRYPT 2008. Springer Berlin Heidelberg, 2008. 90-107.

[2] Nightingale, Edmund B., Peter M. Chen, and Jason Flinn. "Speculative execution in a distributed file system." ACM SIGOPS Operating Systems Review. Vol. 39. No. 5. ACM, 2005.

[3] Beaver, Doug, et al. "Finding a Needle in Haystack: Facebook's Photo Storage." OSDI. Vol. 10. 2010.

[4] Feamster, Nick, Jennifer Rexford, and Ellen Zegura. "The road to SDN: an intellectual history of programmable networks." ACM SIGCOMM Computer Communication Review 44.2 (2014): 87-98.

[5] Carter, Henry, et al. "For your phone only: custom protocols for efficient secure function evaluation on mobile devices." Security and Communication Networks(2013).

[6] https://www.youtube.com/watch?v=WVs7Pc99S7w

[7] Marczak, William R., et al. "When governments hack opponents: A look at actors and technology." Proceedings of the 23rd USENIX Security Symposium. 2014.

[8] Burnett, Sam, and Nick Feamster. "Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests." arXiv preprint arXiv:1410.1211 (2014).

[9] Socher, Richard, et al. "Semantic compositionality through recursive matrix-vector spaces." Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012.

[10] Gabora, Liane, and Diederik Aerts. "Contextualizing concepts using a mathematical generalization of the quantum formalism." Journal of Experimental & Theoretical Artificial Intelligence 14.4 (2002): 327-358.

[11] https://code.google.com/p/word2vec/

[12] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013

[13] Bengio, Yoshua, et al. "Neural probabilistic language models." Innovations in Machine Learning. Springer Berlin Heidelberg, 2006. 137-186.

[14] Aerts, Diederik, and Liane Gabora. "A theory of concepts and their combinations II: A Hilbert space representation." Kybernetes 34.1/2 (2005): 192-221.

[15] Aerts, Diederik. "Quantum structure in cognition." Journal of Mathematical Psychology 53.5 (2009): 314-348.

[16] Aerts, Diederik, and Liane Gabora. "A theory of concepts and their combinations I: The structure of the sets of contexts and properties."Kybernetes 34.1/2 (2005): 167-191.

[17] Chen, Danqi, et al. "Learning new facts from knowledge bases with neural tensor networks and semantic word vectors." arXiv preprint arXiv:1301.3618(2013).

[18] Andreas, Jacob, and Dan Klein. "Grounding Language with Points and Paths in Continuous Spaces."

[19] Mithun Paul, Ashutosh Saxena ., Proof Of Erasability For Ensuring Comprehensive Data Deletion In Cloud Computing, CNSA 10:Proceedings of the Third International Conference, Communications in computer and Information Science, Springer, Vol. 89

[20] Mithun Paul, Ashutosh Saxena., Datashredding Service For Cloud.: Proceedings of the 2nd International Conference on Services in Emerging Markets, Mumbai, September 2011

[21] Mithun Paul, Ashutosh Saxena ., Zero Data Remnance Proof in cloud Storage, International Journal of Network Security & Its Applications (IJNSA), Vol.2, No.4, October 2010.
[22] Mithun Paul, M. Choudary Gorantla and Ashutosh Saxena., Group Key Exchange with Non Linear Trust. Proceedings for The IEEE Fifth International Conference on Internet Multimedia Systems Architecture and Applications, Bangalore, Dec 2011
[23] Gentry, Craig. "Fully homomorphic encryption using ideal lattices." STOC. Vol. 9. 2009.
[24] Mithun Paul, Nitin Singh Chauhan, Ashutosh Saxena., A Security Analysis of Smartphone Data Flow and Feasible Solutions for Lawful Interception. Proceedings for the IEEE 7th International Conference on Information Assurance and Security, Malacca, Malaysia, Dec 2011.
[25] Mithun Paul, Nitin Singh Chauhan, Ashutosh Saxena., A Security Analysis of Smartphone Data Flow and Feasible Solutions for Lawful Interception. Proceedings for the IEEE 7th International Conference on Information Assurance and Security, Malacca, Malaysia, Dec 2011
[26] http://apod.nasa.gov/apod/astropix.html
[27] https://scholarshipuniverse.arizona.edu/home/splash.aspx
[28] http://www.darpa.mil/Our_Work/I2O/Programs/Broad_Operational _Language_Translation_(BOLT).aspx
[29] System and Method for Deletion of Data in a Remote Computing Platform. Patent pub no:US8504532 B2/ US20120317083 A1/ US8504532
[30] "Method and system for providing masking services" ;United States Patent : 8,881,224
[31] http://www.infosys.com/products-and-platforms/maskit/Documents/enterprise-data-privacy-product.pdf
[32] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.
[33] Woo, Simon S., and Jelena Mirkovic. "Optimal application allocation on multiple public clouds." Computer Networks (2014).
[34] Zhang, Qing, David Y. Wang, and Geoffrey M. Voelker. "DSpin: Detecting Automatically Spun Content on the Web." (2014).
[35] www.teachastronomy.com., Accesed Nov 2014
[36] Stringhini, Gianluca, et al. "The harvester, the botmaster, and the spammer: on the relations between the different actors in the spam landscape." Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM, 2014.
[37] Cai, Hongxu, Zhong Shao, and Alexander Vaynberg. "Certified self-modifying code." ACM SIGPLAN Notices 42.6 (2007): 66-77.
[38] Collberg, Christian, Clark Thomborson, and Douglas Low. A taxonomy of obfuscating transformations. Department of Computer Science, The University of Auckland, New Zealand, 1997.