

Mishari Almishari

Department of Computer Science

University of California, Irvine

Phone: (213) 446 2724

Email: malmisha@uci.edu

Objective

- Seeking a research internship position for summer 2010

Research Interests

- Web Mining, Security, Privacy, Applied Cryptography, Cloud Computing

Education

- **Ph.D. in Computer Science** **2006 –present**
University of California, Irvine, GPA: 3.99/4
Advisor: Gene Tsudik
Course Work:
Machine Learning, Randomized Algorithms, Data Base Management Systems, Advanced Network Topics, Information Retrieval, Internet, Introduction to Cryptography, Advanced Cryptography
- **M.S. in Computer Science** **2004 – 2006**
University of Southern California, Los Angeles, GPA: 4/4
Course Work:
Introduction to Network, Advanced Data Base, Advanced Networks, Computer Security, Algorithm Analysis, Web Development, Introduction to Artificial Intelligence
- **B.S. in Computer Science** **1996 – 2001**
King Saud University, Riyadh, Saudi Arabia, GPA: 4.81/5

Employment

- **Research Assistant, UC Irvine** **Summer 2009**
 - Worked on devising a technique to detect “Ads-Portal” domains, websites that have no real content and only shows ads content. The goals were (1) To devise accurate detection technique that identifies ads-portal domains, and (2) To use this technique to measure the prevalence of such domains in the Internet.
- **Teaching Assistant, King Saud University, Saudi Arabia** **2003 - 2004**
 - Helped (for 3 semesters) in tutoring the following undergraduate computer science courses: Introduction to Programming in C, Structural programming in C, Data Structures, and Discrete Mathematics.
- **Systems Engineer, Medunet, Saudi Arabia** **2001 - 2003**
 - Worked in network administration, administering CISCO routers and Marconi ATM switches
 - Worked in designing Internet solutions for several customers
- **Summer Intern, International Systems Engineering Company, Saudi Arabia** **2000**
 - Developed a Q&A website, with email notifications and administration facilities (in a team of 4 persons)
 - Developed Company Society web service that includes a bulletin board, a photo album, and personal information about the employees in the company (in a team of 3 persons)

Awards

- First Class Honor Award from King Saudi University for graduating with outstanding GPA, 2001
- Two-year Scholarship from the Ministry of Higher Education, Saudi Arabia for M.S. study, 2004
- Academic Achievement from USC Office of International Services, 2006
- Five-year Scholarship from Ministry of Higher Education, Saudi Arabia for Ph.D. study, 2006

Publication

- M. Almishari , X. Yang, “Text-Based Ads-portal Domain: Measurement And Identification”, To appear in ACM Transaction on the Web.
- M. Almishari, E. De Cristofaro, K. El Defrawy, G. Tsudik, ” Harvesting SSL Certificates to Identify Web-Fraud”, Under Review.

Research Projects

- **“Text-Based Ads-Portal Domains: Identification & Measurements”**

A text-based ads-portal domain refers to a web domain that only shows advertisements, served by a third-party advertisement syndication service, in the form of ads listing. The goals of this project are (1) To devise accurate detection technique that identifies ads-portal domains, and (2) To use this technique to measure the prevalence of such domains in the Internet. We analyze the content-based properties of text-based ads-portal domains and based on these properties, we develop a machine learning-based classifier to identify text-based ads-portal domains, which has 96% accuracy. Then, we use this classifier to measure the prevalence of ads-portal domains in the Internet. Surprisingly, 28.3/25% of the (two-level) *.com/*.net web domains are ads-portal domains. Also, 41/39.8% of *.com/*.net ads-portal domains are typos of well-known domains, also known as typo-squatting domains. In addition, we use the classifier along with DNS trace files (UCI DNS traces) to estimate how often Internet users visit ads-portal domains. It turns out that around 5% of the two-level *.com, *.net, *.org, *.biz and *.info web domains in the traces are ads-portal domains and around 50% of these accessed ads-portal domains are typos. These numbers show that ads-portal domains and typosquatting ads-portal domains are prevalent in the Internet and successful in attracting many visits. Our classifier represents a step towards better categorizing the web documents. It can also be helpful to search engines ranking algorithms, helpful in identifying web spams that redirects to ads-portal domains, and used to discourage access to typo-squatting ads-portal domains. **To Appear in ACM Transaction on the Web.**

- **“Harvesting SSL Certificates to Identify Web-Fraud”**

The goal of this project is explore the information embedded in the SSL (Secure Sockets Layer) certificates and see if it can be possibly used to identify phishing or typosquatting domains. To address our goal, we first conduct a comprehensive study of SSL certificates for legitimate, popular domains, and web-fraud domains. The study shows that there are significant differences in the content-based properties of SSL certificates of legitimate/popular domains as opposed to those used for web-fraud. We leverage these differences and design a machine-learning based classifier that identify web-fraud domains with high accuracy. **The work is under review.**

- **“Prove of Data Possession Revisited”**

In this project, we have an owner of a file outsourcing it to a server that is not trusted. The goal of this project is devise a cryptographic-based technique that enables the owner of the file to publicly verify that its file is still possessed by the server and has not been changed. Public verification means that the client does not need to keep any credential at the machine to verify the possession of its outsourced file. Despite the enormous amount of work done in the file outsourcing, none of them proposed a solution the enables a super-efficient public verifiability. The best solutions offer a super-constant number of exponentiations which makes their solutions unpractical, especially with client machines that have limited computation capabilities. Our goal is to devise a solution that eliminates the exponentiations or drastically minimizes them. **The work is still under progress.**

Computer Skills

- **Programming Languages**
Java, C, C++, Prolog, Python, BASH Scripting, Pascal
- **Web Technologies**
HTML, CSS, Javascript, XML, XSLT, Java Server Pages, Web Services
- **Environments**
Windows, Linux

Languages

- Arabic (Native)
- English (fluent)

References

- **Gene Tsudik**
Computer Science Department
University of California, Irvine
Bren Hall, 3rd Floor, Irvine, CA, 92697
Phone: (949) 824 3410
Fax: (949) 824 4056
Email: gts@ics.uci.edu
- **Xiaowei Yang**
Computer Science Department

Duke University
450 Research Drive, D336 LSRC Building, Durham, NC 27708
Phone: (919) 660 6536
Fax: (919) 660 6519
Email: xwy@cs.duke.edu

- **Contact Information**

Mishari Almishari
8111 Palo Verde Rd
Irvine, CA, 92617
Phone: (949) 737 2006
Mobile: (213) 446 2724
Email: malmisha@uci.edu; malmisha@ics.uci.edu; malmishari@gmail.com