Last Updated on 6th March 2017
Written using LaTeX

Steven H. H. Ding

steven.h.ding@mail.mcgill.ca | +1(514)915-8881
stevending.net | Github:// **steven-hh-ding** | LinkedIn:// **steven-hh-ding**

EDUCATION

MCGILL UNIVERSITY
PH.D. IN INFORMATION STUDIES
School of Information Studies
Sept 2014 - present | Montreal, CA
Supervisor: Benjamin C. M. Fung
Topic: Assembly Code Data Mining for Reverse Engineering
Cum. GPA: 3.95/4.0

CONCORDIA UNIVERSITY
M.A.SC. IN INFORMATION SYSTEMS SECURITY
Concordia Institute for Information Systems Engineering
Sept 2012 - Apr 2014 | Montreal, CA
Supervisor: Benjamin C. M. Fung and Murad Debbabi
Thesis: Visualizable Authorship Analysis
Cum. GPA: 4.24/4.3

UNIVERSITY OF THE FRASER VALLEY
B.S. IN INFORMATION SYSTEM
Sept 2011 - Aug 2012 | Abbotsford, CA
Dean's List
Institutional GPA: 3.64/4.3

UNIVERSITY OF SHANGHAI FOR SCIENCE AND TECHNOLOGY
B.S. IN COMPUTER SCIENCE
Sept 2008 - Jun 2012 | Shanghai, CN
Excellent Graduate (top 3%)
Thesis: Location-based Collaborative Recommendation System
Institutional GPA: 3.24/4.0

SERVICES

PROGRAM COMMITTEE (PC) MEMBER AND REVIEWER
IDEAS 2016 • IEEE BigData Congress 2017

EXTERNAL REVIEWER
CIKM 2017 • IJCAI 2017 • WWW 2017 • IEEE ICDM 2016 • ACM SIGKDD 2016 • ACM SIGKDD 2015 • ACM SIGKDD 2014 • IEEE ICDE 2014

REVIEWER
IEEE Big Data 2014 • IEEE ICDM 2013

RESEARCH INTERESTS

DATA MINING AND MACHINE LEARNING
•Nearest-neighbor search • Information retrieval • Graph retrieval • Heterogeneous information networks mining • Text mining • Representation learning

APPLICATIONS
•Reverse engineering •Malware analysis •Authorship analysis •Information management • Cyber investigation

HIGHLIGHTED PROJECTS

KAM1N0: ASSEMBLY CODE DATA MINING FOR REVERSE ENGINEERING
FULL STACK DEVELOPER • ALGORITHM DESIGNER • SYSTEM ARCHITECT
Collaborate with Defense Research and Development Canada (**DRDC**)
Jan 2015 – present | McGill University, Montreal, CA

- Developed a scalable assembly code clone search engine Kam1n0. It significantly reduces the manual effort of reverse engineers.
- The search engine is based on Cassandra and Spark. Available at Github.
- Won the second price in the Hex-Rays Plug-in Contest 2015.
- Publications: SIGKDD16, SIGIR17 and USENIX-SEC17 (under review)

AUTHORMINER: AUTHORSHIP ANALYSIS FOR CYBER CRIME INVESTIGATION
FULL STACK DEVELOPER • ALGORITHM DESIGNER • SYSTEM ARCHITECT
Collaborate with National Cyber Forensic and Training Alliance Canada (**NCFTA Canada**)
Sept 2012 – present | McGill University, Montreal, CA

- Developed accurate models and algorithms to identify, verify and characterize the author of a given anonymous text snippet based on the reflected writing style. Developed a software to visualize the identified writing styles.
- Collaborated with Zayed University, UAE starting from 2014.
- Publications: ACM TISSEC, IEEE CYBERNETICS (under review).

OWNERMINER: FILE SYSTEM OWNERSHIP ANALYSIS BASED ON INFORMATION MANAGEMENT BEHAVIOR
DATA ANALYST • ALGORITHM DESIGNER
May 2016 – present | McGill University, Montreal, CA

- Analyze one's information management behaviors based on his/her file system.
- Develop new data mining techniques to identify one's characteristics such as age, gender, educational levels, etc., based on the reflected information management behavior.
- Propose new data mining techniques to identify the ownership of a file system given a set of candidates.

AWARDS

2016 Best Poster Award, The SERENE-RISC (funded by NCE) Fall 2016 Workshop
2016 ACM SIGKDD 2016 Student Travel Award
2015 Second prize in Hex-Rays International Plug-in Contest, International
2014 The Dean's Graduate Award (4 years), McGill University, CA
2013 Concordia Partial Tuition Scholarship, Concordia University, CA
2012 GSSP funding (2 years), Concordia University, CA
2012 Excellent Bachelor Graduate (top 3%), Shanghai, China
2012 Dean's List, University of the Fraser Valley, CA
2010 1st Prize Excellent Student Scholarship, Univ. of Shanghai for Sci. & Tech., CN

SKILLS

DATA MINING FRAMEWORKS

Scikit-learn • tensorflow • spark • theano • weka

PROGRAMMING

- JAVA > 80,000 lines
- C# > 20,000 lines
- LaTeX > 12,000 lines
- HTML > 10,000 lines
- JavaScript > 10,000 lines
- Python > 5,000 lines
- C++ > 5,000 lines
- R > 1,000 lines
- Scala Familiar
- Shell Familiar

DEVELOPMENT FRAMEWORKS

ASP.Net • ASP MVC 3 • WPF • WF • WCF • Maven • JSP • SSH2 MVP • Apache/Google common libs

OTHER ACTIVITIES

•2016
SERENE-RISC Spring 2016 Workshop Volunteer

•2012-2016
Volunteer Website Developer for Grace Ministries International of Montreal (a non-denominational bible believing church)

•2011
Expo. Shanghai 2011 Volunteer, CN.

•2010
TORY International Marathon Volunteer, Shanghai, CN

RESEARCH EXPERIENCE

MCGILL DATA MINING AND SECURITY LAB | Research Assistant

Sept 2014 – Present | Montreal, CA

- Create and improve Kam1n0, a clone search engine for reverse engineers.
- Participate in the research plan for the Kam1n0 project.
- Closely work with Benjamin C. M. Fung to improve our previous work on the authorship analysis problem. Propose three joint learning models for this task.
- Collaborate with Benjamin C. M. Fung and Jesse David Dinneen on the OwnerMiner project.
- Manage and maintain lab equipments and cloud resources. Create quotes and design lab network topology for equipment purchasing and grant application.

EBTIC RESEARCH CENTRE ESTABLISHED BY BRITISH TELECOMMUNICATIONS (BT) | Research Internship

Sep 2016 – Sep 2016 | Abu Dhabi, UAE

- Developed a new Heterogeneous Information Network (HIN)-based data mining approach for source code repository mining and visualization.

NATIONAL CYBER-FORENSICS AND TRAINING ALLIANCE CANADA | Research Consultant

Jun 2014 – Aug 2014 | Montreal, CA

- Developed a real-time monitoring system to detect DNS anomalies. Supervised by Murad Debbabi.

CONCORDIA INSTITUTE FOR INFORMATION SYSTEMS ENGINEERING | Research Assistant

Sept 2012 – Jun 2014 | Montreal, CA

Created new data mining algorithms to identify the actual author of a given anonymous snippet from a set of candidates.

- The proposed classification algorithm achieves a higher identification accuracy than the traditional stylometric-based approaches.
- Created a software for authorship analysis. It is able to present the visualized writeprint as evidence supporting the attribution results.

LABORATORY OF NETWORK ENGINEERING | Research Assistant

May 2010 – May 2011 | Univ. of Shanghai for Sci. and Tech., Shanghai, CN

- Assisted in evaluating and developing clustering techniques for uncertain data. In uncertain data, a data point is represented as a standalone probabilistic distribution.

PUBLICATIONS

[1] **S. H. H. Ding**, B. C. M. Fung, and P. Charland. Kam1n0: Mapreduce-based assembly clone search for reverse engineering. In *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (***SIGKDD16**), San Francisco, CA, August 2016. ACM Press.

[2] **S. H. H. Ding**, B. C. M. Fung, and M. Debbabi. A visualizable evidence-driven approach for authorship attribution. *ACM Transactions on Information and System Security* (**TISSEC**), 17(3):12:1–12:30, March 2015.

[3] **S. H. H. Ding**, B. C. M. Fung, and K. Zhao. Assembly code data mining: design and prototyping (part 3). Technical Report W7701-155902/001/QCL (TA02), *Defence Research and Development Canada*, March 2016.

[4] **S. H. H. Ding**, B. C. M. Fung, and K. Zhao. Assembly code data mining: porting to other platforms. Technical Report W7701-155902/001/QCL (TA03), *Defence Research and Development Canada*, February 2016.

[5] **S. H. H. Ding**, Benjamin C. M. Fung, Farkhund Iqbal, and W. K. Cheung. Learning stylometric representations for authorship analysis. *arXiv preprint arXiv:1606.01219* **(ACM TOIS under review)**, 2016.

[6] B. C. M. Fung, **S. H. H. Ding**, M. R. Farhadi, and K. Zhao. Assembly code data mining: design and prototyping (part 1). Technical Report W7701-155902/001/QCL (Firm Portion), *Defence Research and Development Canada*, March 2015.

[7] B. C. M. Fung, **S. H. H. Ding**, and K. Zhao. Assembly code data mining: design and prototyping (part 2). Technical Report W7701-155902/001/QCL (TA01), *Defence Research and Development Canada*, January 2016.