Zitao Chen • *zitaoc@ece.ubc.ca*

Education

University of British Columbia — Vancouver, Canada
Master of Applied Science in Electrical and Computer Engineering — *Sept. 2018 - May 2020*
Supervisor: Karthik Pattabiraman
Thesis: Understanding and Improving the Error Resilience of Machine Learning Systems
Thesis Committee: Karthik Pattabiraman, Sathish Gopalakrishnan, Prashant Nair

China University of Geosciences (Wuhan) — Wuhan, China
Bachelor degree in Information Security — *Sept. 2014 – Jun. 2018*

Research Experience

1. Security of Machine Learning (adversarial ML) — Ongoing Research

- *Research question:* How to defend *adversarial patch attack*, i.e., attack by applying adversarial patch into the images to create adversarial samples (AEs)?
  Insight: Please contact for more details.
- *Research question:* How to defend adversarial samples (AEs) with *imperceptible* perturbations?
  Insight: Small and imperceptible perturbations are fragile to transformation (e.g., under different viewpoints), and thus can be defended by *input transformation* (e.g., affine, perspective transformation). Preliminary results show that input transformation is effective against imperceptible AEs (e.g., increase the network accuracy on AEs from CW attack from 1% to 98%), but its efficacy is less pronounced on those perceptible AEs, which are more resilient to transformation (e.g., on PGD attack, the accuracy is increased from 3% to only 58.9%).

2. Reliability of Machine Learning — Master's Research

2.1. Understanding the error resilience of ML

- *Research question:* How to efficiently identify the critical faults in ML systems due to hardware transient faults (i.e., bit-flips)? These are the faults that could corrupt the ML's outputs, e.g., misclassification.
- *Key insight:* Common ML models consist of functions with *monotone* property and thus deviations by faults also propagate monotonically to the output layer through these monotonic functions, i.e., faults at high-order bits are more likely to corrupt the outputs.
- Design a binary fault injector to efficiently identify the critical faults in ML systems.
- *Results:* The proposed method can identify 99%+ of critical faults with 99%+ precision, and significantly outperforms the widely-used random injection approach.
- Code: https://github.com/DependableSystemsLab/TensorFI-BinaryFI

2.2. Improving the error resilience of ML

- *Research question:* How to protect ML systems from hardware transient faults?
- *Key insight:* ML is a statistical process and does not always require inexactness. Propose to selectively restrict the value ranges in different layers of DNNs, which can dampen the large deviations (due to critical faults) into smaller ones. The reduced deviations can be *inherently tolerated* by DNNs to generate correct outputs *without* re-computation.
- Implement an automated transformation to convert the unreliable DNNs into the error-resilient ones in TensorFlow.
- *Results:* The proposed technique significantly improves the ML reliability (e.g., reduce the chance of misclassification due to transient faults from $\sim 15\%$ to $\sim 0.4\%$) with negligible overhead ($\sim 0.5\%$).

3. Wearable Computing Security — Undergraduate Research

- *Research question:* How to leverage the sensory signals in wearable devices for random key generation.
- *Insight:* Selectively exploit the part of sensory data originated from device vibration and/or hardware imprecision as the source for random key generation, without extra feature processing.
- *Result:* This approach enables lightweight and real-time random key generation in wearable devices.

Publications **(SC is a top conference in HPC indexed by csrankings.org)**

- **[IOLTS'20]** Karthik Pattabiraman, Guanpeng Li, Zitao Chen, "Error Resilient Machine Learning for Safety-Critical Systems: Position Paper" *IEEE 26th International Symposium on On-Line Testing and Robust System Design*, 4 pages, 2020. *Invited paper*
- **[ISSRE'20]** Zitao Chen*, Niranjhana Narayanan*, Bo Fang, Guanpeng Li, Karthik Pattabiraman, Nathan DeBardeleben, "TensorFI: A Flexible Fault Injection Framework for TensorFlow Applications" *The 31st International Symposium on Software Reliability Engineering*, 2020 **Acceptance rate: 25.7% (38/148)**
- **[SC'19]** Zitao Chen, Guanpeng Li, Karthik Pattabiraman, Nathan DeBardeleben "BinFI: An Efficient Fault Injector for Safety-Critical Machine Learning Systems", *In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2019 **Acceptance rate: 20.9% (72/344). Finalist for SC reproducibility challenge (one of 3 papers)**
- **[FGCS]** Zitao Chen, Wei Ren, Yi Ren and Kim-Kwang Raymond Choo, "LiReK: A Lightweight and Real-time Key Establishment Scheme for Wearable Embedded Devices by Gestures or Motions", *Future Generation Computer Systems* (2018) [Impact Factor: 6.125]

Preprint

- **[ArXiv]** Zitao Chen, Guanpeng Li, Karthik Pattabiraman"Ranger: Boosting Error-Resilience of Deep Neural Networks through Range Restriction"

Others

Open-source project: https://github.com/DependableSystemsLab/TensorFI Sept. 2018 - present

Teaching Experience: CPEN400A 2019

Programming languages: Python, Java, C, C++

Award: Graduate Student Initiative: $4000; $3000 2019,2020