

Yao Pan

CONTACT INFORMATION *Mobile:* +1-615-852-9853
E-mail: yao.pan@vanderbilt.edu
Website: <https://sites.google.com/view/yaopan>

EDUCATION **Vanderbilt University**, Nashville, TN, United States

Ph.D. Computer Science December 2017

- Dissertation: [Gray Computing: A Framework for Distributed Computing with Web Browsers](#)
- Advisor: Dr. Jules White

Zhejiang University, Hangzhou, China

B.S., Computer Science, GPA: 3.9/4.0 August.2008 – June.2012

EXPERIENCE **Instacart Next Basket Product Prediction Competition on Kaggle**

Research Assistant, Vanderbilt University Jun. 2017 – Aug. 2017

- Participated in the Instacart prediction competition hosted on Kaggle. Developed models to predict which products a user will buy again from 3 million historical grocery orders.
- Ranked in the top 7% in prediction accuracy among 2600 participants.
- Performed exploratory data analysis on the dataset and extracted over 80 features including word embedding. Used LightGBM and XGBoost to train gradient boosted tree-based models.
- Implemented a RNN-based deep learning model in TensorFlow without manual feature extraction.

Real-time Cryptocurrency Price Prediction with Big Data

Research Assistant, Vanderbilt University Dec. 2017 – Jan. 2018

- Implemented a real-time cryptocurrency price prediction pipeline with Kafka, Spark Streaming and Cassandra.
- Collected real-time and historical price data from Coindesk and social media data from Twitter and Reddit. Persisted the data with Cassandra.
- Performed sentiment analysis for real-time social media data as an indicator for price movement.
- Implemented a LSTM-based deep learning prediction engine and a ARIMA-based time series prediction engine.

Automatically Generating Fake Yelp Reviews with Deep Learning

Research Assistant, Vanderbilt University Aug. 2017 – Oct. 2017

- Trained a LSTM network with 100K user reviews from Yelp Restaurant Review Dataset. The trained network was able to automatically generate restaurant reviews that are indistinguishable from human reviews.
- Improved the review quality by replacing the food related nouns (finding with WordNet) in the generated reviews with the ones appeared in the restaurant's reviews.
- Evaluated the quality of machine-generated reviews with plagiarism test and SVM classifier. More than 90% of the generated reviews passed the test.

Anomaly Detection with Deep Learning to Defend against Web Attacks

Research Assistant, Vanderbilt University Jun. 2016 – Aug. 2017

- Developed tools using Java Agents for server-side method call instrumentation to detect cyber-attacks with deep learning. Recorded call traces to transform call sequences into a matrix as signatures to differentiate between legitimate web requests and malicious requests.
- Implemented a stacked denoising autoencoder in Keras and applied it to the call traces for anomaly detection. Compared the performance with PCA and SVM. The detection engine achieved a 0.96 F1 score on automatically generated legitimate traffic and abnormal traffic (SQL injection, deserialization attacks).

A Framework for Distributed Computing with Web Browsers

Research Assistant, Vanderbilt University Mar. 2014 – Jun. 2016

- Proposed a framework utilizing Web Workers in HTML5 for distributed computing with heterogeneous and volatile browsers' computing power.
- Evaluated the feasibility of this framework in terms of task distribution/scheduling, JavaScript performance, user experience impact, security and cost-effectiveness.
- Ran experiments on applications including face detection, sentimental analysis, rainbow table generation implemented in JavaScript running in browsers and C++ running in AWS to compare the cost-effectiveness.
- Published work as 1st author in ICSE, the top 1 conference in software engineering.

Large-scale Taxi Traces Data Mining and Passenger Waiting Time Prediction

Undergraduate Researcher, Zhejiang University

Aug. 2011 – Jun. 2012

- Extracted get-on/off hotspots using clustering analysis from a large-scale taxi trace dataset which consists of 300 million GPS data records from 4000 taxis.
- Modeled the taxi waiting process and derive the PDF and CDF of passenger waiting time. Applied maximum likelihood estimation to estimate parameters to predict passenger's waiting time for a given location/time.
- Achieved an average prediction error of 4.5 minutes, which outperforms approaches using autoregression integrated moving average (ARIMA) and Naive Bayes. Visualized results using Google Map API and heatmap.

JOURNAL PUBLICATIONS

Yao Pan, Jules White, Yu Sun, Jeff Gray, Gray Computing: A Framework for Computing with Background JavaScript Tasks, *IEEE Transactions on Software Engineering*, 2017.

Yao Pan, Jules White, Douglas C. Schmidt, Ahmad Elhabashy, Logan Sturm, Jaime Camelio, Christopher Williams, Taxonomies for Reasoning About Cyber-physical Attacks in Advanced Manufacturing Systems, *Special Issue on Advances and Applications in the Internet of Things, International Journal of Interactive Multimedia and Artificial Intelligence*, 2017.

Hyojoon Bae, Michael Walker, Jules White, **Yao Pan**, Yu Sun, Mani Golparvar-Fard, Fast and Scalable Structure-from-Motion based Localization for High-precision Mobile Augmented Reality Systems, *Journal of Mobile User Experience*, 2016.

Hyojoon Bae, Jules White, Mani Golparvar-Fard, **Yao Pan**, Yu Sun, Fast and Scalable 3D Cyber-physical Modeling for High-precision Mobile Augmented Reality Systems, *Journal of Personal and Ubiquitous Computing*, 2015.

CONFERENCE PUBLICATIONS

Yao Pan, Jules White, Yu Sun, Jeff Gray, Gray Computing- An Analysis of Computing with Background JavaScript Tasks, *37th International Conference on Software Engineering (ICSE)*, May 16-21, 2015, Florence, Italy.

Yao Pan, Jules White, Yu Sun, Assessing the Threat of Web Worker Distributed Attacks, *IEEE Conference on Communications and Network Security*, October 17-19, 2016, Philadelphia, PA.

Fangzhou Sun, **Yao Pan**, Jules White and Abhishek Dubey, Real-time and Predictive Analytics for Smart Public Transportation Decision Support System, *IEEE International Conference on Smart Computing*, 2016, St.Louis, USA.

Jules White, **Yao Pan**, Zack McCormic, Addressing the Challenges of HTTP-based Mobile/Cloud Interaction, *IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, Oxford, UK, April 8-11, 2014.

SKILLS

Programming Languages:

- Java, C++, Python, JavaScript, Scala, SQL, C, Matlab, HTML.

Technologies:

- Scikit-learn, TensorFlow, Keras, Hadoop, Spark, MLlib, Kafka, Cassandra, Docker, Unix/Linux, pandas, AWS, Google Cloud Platform, MySQL, MongoDB, Spring, Node.js, Git, Android.