Jianfeng Wang

- 941 Bloom Walk, Salvatori Computer Science Center (SAL) 224, Los Angeles, CA 90089
- Email: jianfenw@usc.edu • Phone: (+1)213-709-5021

Education

University of Southern California — Los Angeles, CA, USA

Ph.D. Candidate in Electrical and Computer Engineering; GPA: 3.92/4.00 *09/ 2016 – present*

Thesis: Performant, Scalable and Efficient Deployment of Network Function Virtualization
Advised by Prof. Ramesh Govindan and Prof. Barath Raghavan
Focus: **Network Function Virtualization (NFV), multi-core CPU scheduling, hardware offloading**
Courses: Advanced Computer Networks (CS 551), Advanced Analysis of Algorithms (CS 670), Operating Systems (CS 402), Artificial Intelligence (CS 561), Database Systems (CS 585), Design and Analysis of Computer Communications (EE 550)

Peking University — Beijing, China

Bachelor of Science in Electrical Engineering; GPA: 3.73/4.00 *09/ 2012 – 07/ 2016*
Key words: mobile computing, wireless communication; Rank: top 2/71

Bachelor of Art in Economics *09/ 2013 – 07/ 2016*
Key words: game theory, optimization theory

Employment

INVISV Inc. - Private Relay Group, Founding Engineer *05/ 2022 – present*

- Tech lead on INVISV private relay, a multi-party relay service, to develop an Android VPN client and a Layer-4 traffic relay service. Collaborated with Fastly's H2O (HTTP proxy) team
- Designed and implemented INVISV Layer-4 traffic relay service. The service includes a control plane for monitoring, scaling and handing failures, and a proxy data plane using DPDK and docker container. It is deployed across 6 global cloud sites. Also implemented INVISV Android VPN client
- Thousands of users in less than a month. Wired News had a report on our service and Android app

Google Inc. - Congestion Control (CC) Group, Intern *05/ 2020 – 08/ 2020*

- Worked on benchmarking and optimizing TCP Congestion Control (CC) for GCP Cloud VMs
- Ported Swift CC [SIGCOMM '20] to the upstream *Linux kernel*. Proposed the design for evaluating CC performance in GCP Cloud and developed the benchmarking framework
- Discovered several performance problems in TCP and BBR. Fixed them in Google's Linux kernel and the upstream Linux kernel. Improved BBR's streaming throughput by 70% compared to SOTA
- Demonstrated that Swift reduces the network delay by 60% compared to TCP Cubic (Linux's default CC) and benefits latency-sensitive apps, including Redis database

Google Inc. - Chrome Performance Team, Intern *05/ 2019 – 08/ 2019*

- Worked on improving Chromium loading performance by detecting and preventing network congestion
- Designed and implemented Chromium *Network Congestion Analyzer* that detects network congestion events by analyzing the transport-level packet traces in real time (open-sourced)
- Refactored Chromium's resource scheduler by making its algorithm aware of network congestion

Experience

Microsecond-scale Tail Latency SLOs for Network Functions *09/ 2021 – 11/2022*

- Advisors: Prof. Ramesh Govindan, Prof. Barath Raghavan, USC
- Designed and implemented Ironside, a rack-scale cluster task scheduler, that enables μs-scale tail (p99) latency SLOs for latency-sensitive flow/session-based workloads, such as NFV (e.g., L4 Load Balancer)
- Designed and Implemented Ironside's monitoring system and its auto-scaling algorithm that utilizes hardware timestamps, and uses flow tables at ToR switches and NICs for scalability, and software packet forwarding for detecting and handling traffic bursts
- Ironside is able to deliver tail latency SLOs at 100s μs-scale, 10-100$\times$ better than SOTA NFV systems

Elastic Network Function Executions In Rack-scale Clusters *09/ 2019 – 12/2020*

- Advisors: Prof. Ramesh Govindan, Prof. Barath Raghavan, USC
- Existing serverless platforms cannot support latency-sensitive workloads. Designed and prototyped Quadrant, a serverless platform for rack-scale cluster. Quadrant deploys NFs as containers. It uses DPDK, and a novel isolation mechanism for low-overhead inter-container message passing
- Designed and Implemented Quadrant Controller that controls the system ingress, per-worker CPU core schedulers. The controller scales NF chains to serve dynamic traffic under latency SLOs
- Quadrant delivers up to 2.3x the performance per core compared to SOTA NFV platforms

Optimizing Network Function Performance With Hardware-offloading *10/ 2017 – 08/ 2019*

- Advisors: Prof. Ramesh Govindan, Prof. Barath Raghavan, USC
- Hardware techniques, e.g., SmartNICs, P4 switches, are options for accelerating network data plane
- Designed and developed Lemur, a NFV framework that places and executes NF chains across heterogeneous platforms (SmartNICs, P4 and OpenFlow switches, multi-core servers). Lemur's algorithm produces a near-optimal NF placement while considering hardware resource constraints. Lemur's meta-compiler automatically generates NF code and codes for coordinating NFs across platforms
- Lemur outperforms other alternative placement algorithms while ensuring low performance overhead

TCP Trace Compression for Network Diagnosis *11/ 2016 – 07/ 2017*

- Advisor: Prof. Ramesh Govindan, USC
- Packet traces are critical for diagnosing network anomalies. Existing diagnosis systems collect sampled packet traces, which leads to a low hit-rate for critical network events, and wastes storage and CPUs
- Designed and implemented a TCP trace compression system to store traces using only 20% of that traditional methodology requires. Our system uses an online amnesic algorithm that preserves great details of most-recent traces and reduces the amount of details for historical traces
- Measured that reconstructed TCP traces retain high accuracy (>90%) when examining with popular network diagnosis tools

Publications

Jianfeng Wang, Siddhant Gupta, Marcos A. M. Vieira, Barath Raghavan, Ramesh Govindan, "Ironside: Microsecond-scale Latency SLOs for Network Function Chains", submitted to USENIX ATC 2023

Zongyin Hao, Quanfeng Huang, Chengpeng Wang, Jianfeng Wang, Yushan Zhang, Rongxin Wu, Charles Zhang, "Pinolo: Detecting Logical Bugs in Database Management Systems with Approximate Query Synthesis", submitted to USENIX ATC 2023

Jianfeng Wang, Zhuojin Li, Tamás Lévai, Marcos A. M. Vieira, Barath Raghavan, Ramesh Govindan, "Quadrant: A Cloud-Deployable NF Virtualization Platform", published at ACM SoCC 2022

Jianfeng Wang, Zhuojin Li, Tamás Lévai, Marcos A. M. Vieira, Barath Raghavan, Ramesh Govindan, "Galleon: Reshaping the Square Peg of NFV", arXiv CoRR 2021

Jane Yen*, Jianfeng Wang*, Sucha Supittayapornpong, Marcos A. M. Vieira, Ramesh Govindan, Barath Raghavan, "Meeting SLOs in Cross-Platform NFV", published at ACM CoNEXT 2020 (* = co-first author)

Professional Experience

- External reviewer for ACM SoCC 2021, USENIX NSDI 2022
- Reviewer for IEEE INFOCOM 2020, IEEE Transaction of Information Theory, and IEEE Wireless Communications and Networking Conference (WCNC) 2020

Skills

Programming: Go, Python, C, C++, Shell Scripting, HTML/CSS, Git

Specialties (Cloud Computing): Multi-core CPU Scheduling, Docker and Kubernetes, Container Networking, Cloud Networking, Serverless Lambda Services, Prometheus (monitoring)

Specialties (Networking): Network Protocols (TCP BBR), SDN (P4, OpenFlow), Fast Packet Processing Techniques (DPDK, SmartNIC, P4 and programmable switch)

Others: Linux Kernel programming/tracing/debugging, Mininet, Wireshark, MATLAB and LaTeX

Awards & Honors

Student travel grants for ACM SIGCOMM, USENIX NSDI

USC Annenberg Graduate Fellowship (awarded to top 5% of PhD applicants) 08/ 2016

Carku Scholarship, Peking University (top 5%) 11/ 2015

Excellent Researcher Award, Peking University 10/ 2015

Peking University SK Scholarship (highest award in PKU, top 5/3,300) 10/ 2014

Baidu Scholarship 10/ 2013

Silver Medal, Chinese Physics Olympiad (CPHO) Final 11/ 2011