KAPIL AGRAWAL

kapilagrawal95.github.io kapila1@uci.edu Google Scholar

INTERESTS

Interested in Cloud Computing and Datacenter Networking with a special focus on large-scale failure mitigation and disaster management techniques for cloud-scale services.

EDUCATION

University of California, Irvine Present
3rd year Graduate Student, Computer Science *Irvine, CA*

- CGPA: 4.00
- Graduate Research Assistant, Prof. Sangeetha Abdu Jyothi
- Graduate Teaching Assistant, CS 256 System and ML *(Grad Course)*

Delhi Technological University Jun 2017
B.Tech (Hons.) in Mathematics and Computing Engineering *Delhi, India*

EXPERIENCE

Microsoft Research Bengaluru, India
Research Fellow *Jul 2018 - Jul 2021*

Indian Institute of Technology Himachal Pradesh, India
Research Assistant *Jun 2017 - Jul 2018*

TECHNICAL SKILLS

Languages: Python, C/C++, LaTeX, HTML/CSS, JavaScript, SQL, Bash, PHP
Frameworks & Services: Kubernetes, Helm, Ansible, Docker, Flask, MongoDB, SQLite, Azure, Jaeger, Redis
Toolkits and Packages: TensorFlow, Selenium, Gurobi, RLLib, NetworkX, Sklearn, D3js, lifelines, plotly

PUBLICATIONS & PATENTS

Gecko: Automated Feature Degradation for Cloud Resilience
Kapil Agrawal, Sangeetha Abdu Jyothi
(In Submission) OSDI '24

Monitoring Cloud Service Unreachability at Scale
Kapil Agrawal, Viral Mehta, Sundararajan Renganathan, Sreangsu Acharyya, Venkat Padmanabhan, Chakri Kotipalli, Liting Zhao
Proceedings of IEEE INFOCOM 2021

A Comparison of Class Imbalance Techniques for Real-World Landslide Predictions
Kapil Agrawal, Yashasvi Baweja, Ritwik Saha, Shubham Agrawal, Pratik Chaturvedi, Naresh Mali, Venkata Uday Kala, Varun Dutt
IEEE Conference on Machine Learning and Data Science, 2017

Low-cost sensor-based system for landslide monitoring and alerts
Varun Dutt, **Kapil Agrawal**, Shubham Agrawal, Pratik Chaturvedi, Naresh Mali, Venkata Uday Kala
Patent Application PCT/IN2018/050217. Geneva, Switzerland, 16/04/2018

PROJECTS

Multi-Cloud Resilience UC Irvine, IBM
Advisors: Prof. Sangeetha Abdu Jyothi, Dr. Saurabh Jha (IBM) *Apr 2023 - Dec 2023*

- The rise of multi-cloud architecture while significant does not yet answer the problem of resilience fully such as performing automatic failovers, cloud bursting, disaster recovery, etc.
- Currently, dissecting individual control plane operations to understand and build a theory for developing a new control plane for the multi-cloud for resilience.
- Early results show that while most control plane operations such as PodScheduler, ReplicaSetController of Kubernetes do not require a strong co-ordination across two clusters running on separate clouds, some controllers such as traffic managements, ResourceQuotas, require strong consistency for correct functioning.

Automating Resilience Management UC Irvine

Advisor: Prof. Sangeetha Abdu Jyothi *Sept 2021 - Dec 2023*

* Proposed a new resilience strategy for cloud termed as *Diagonal Scaling* to coordinate graceful degradation at datacenter (DC)-scale so critical workloads can continue running with high availability at the cost of turning off non-critical workloads as opposed to expensive redudancy based geo-replication approaches.
* Pioneered a concise representation, *Criticality Tags*, that allow third-party applications to specify their resilience goals to the datacenter in a simple yet expressive manner making it practical in third-party settings such as public clouds.
* Built an automated controller in Python (1200 lines of code), Phoenix, that interfaces with Kubernetes and can perform diagonal scaling in an automated manner by ensuring that average critical service availability remains high while maximizing cloud provider's goals such as fairness, revenues at DC-scale settings when disasters strike.
* Prototyped and validated the idea of **Diagonal Scaling** on 5 microservice-based applications running on a 100 node Kubernetes cluster. Our results show that applications can continue serving their critical services by surviving up to 60% of cluster failures.

Monitoring Last-Mile Network Unreachability for Office365 Microsoft

Advisors: Dr. Venkat Padmanabhan & Dr. Sreangsu Acharyya *Jul 2019 - Mar 2020*

* Carried out a Microsoft-scale study of first-party services and observed that users' requests were often not directed to the nearest service point causing a bump in latency consequently prompting interest to actuate the first planet-scale measurement platform, Odin, for Microsoft's 2 billion+ users.
* Built a highly scalable ML-based time-series detection tool, NetDetector implemented in Python without library overload, that monitors availability globally atop Odin, a client-side javascript agent, that probes Outlook by mimicking users. NetDetector reduced customer complaints from over 100+ complaints per month to zero in 3 months after deployment.
* Developed a drill-down mechanism that performed distributed search over the vast space of possible reasons to explain the root cause of failures as a one-click auto-debug feature for on-call engineers thereby reducing the time from 2+ hours down to less than a minute.

Data-driven Networking for Exchange Datacenters Microsoft

Advisors: Dr. Venkat Padmanabhan & Dr. Sreangsu Acharyya *Jan 2020 - Jan 2021*

* Co-led a Microsoft-scale study to understand which workloads consume the most network bandwidth and found that Migration contributes the most with 40%, of which Mailbox Migration alone contributes 70%.
* Streamlined SQL pipelines to generate daily reports on Microsoft's cold storage, CosmosDB, that processed billions of mailboxes to study access patterns on mailboxes. Used smoothing techniques such as sliding window, moving average.
* Modeled access patterns of individual workloads as a function of the age of emails by applying the Kaplan-Meier filter implemented using python's lifelines package on a representative mailbox population and found three buggy clients that brought about bulk accesses burdening the Office Exchange Store.

SSD Cache Modelling for Outlook Mail Microsoft

Advisors: Dr. Venkat Padmanabhan & Dr. Sreangsu Acharyya *Jan 2020 - Jul 2021*

* Partnered with Exchange Store Team to model access patterns to devise a new cache eviction policy. Developed an open gym on SQL and python to test several caching policies and analyze performance overheads.
* Launched a monitoring tool to track access patterns of 50+ workloads by building an efficient pipeline that implemented downstream survival analysis tasks on access pattern logs for more than 300K mailboxes in minutes every day.
* Pioneered a testbed that replays accesses on mailboxes to compare cache eviction policies. Engineered a data-driven mechanism to auto-tune the time parameter in TTL-based cache eviction policy that outperformed LRU's 60% cache hit rate with an improved cache hit rate of 95%.

Focused Extraction from Structured Data Microsoft

Advisor: Dr. Arun Iyer *Jul 2018 - Dec 2018*

* Designed a highly accurate ML-based feature for Outlook that extracts information from organized content such as order summaries, flight itineraries, etc. This feature allowed Outlook to extract information on unseen templates which was previously not possible.
* Curated a large data repository with >10000 labeled examples as a testbed to examine ML models on information retrieval tasks such as email summarization, and autofill feature in Excel with Microsoft-wide adoption.
* Devised a technique that leverages visual cues in the structured text to form segmentation trees to generate embeddings for entire text blocks. Assembled a semi-supervised classifier which improved the overall precision of the two-staged ML model from 90% to 99% with <100 data points.