# James C. Davis

davisjam@purdue.edu

https://davisjam.github.io/

## RESEARCH INTERESTS

My research focuses on **software quality**. I use qualitative and quantitative methods to identify socio-technical defects and shortcomings. I identify and resolve problems using systems and security principles. For example:

- **Troublesome tools**: I have given a strong empirical foundation to the study of regular expressions (FSE'18, FSE'19, ASE'19a, ASE'19b).

- **Emerging paradigms**: I have studied the systems and security issues that arise in event-driven programming as embodied in the Node.js framework (EuroSys'17, USENIX Security'18). I have also examined security issues in the use of the query language GraphQL (ICSOC'19, FSE'20).

- **Distributed systems**: I hold several patents on testing distributed storage systems, and have researched performance issues in stream processing engines (USENIX ATC'19).

## EMPLOYMENT

**Assistant Professor, Electrical and Computer Engineering**          Fall 2020-present
*Purdue University*

**Graduate Research Assistant**          2016-2020
*Virginia Tech*
- **Practice-motivated systems design**: Re-designing regex engines based on practitioners' perspectives.
- **Security in emerging paradigms**: Denial of service in event-driven frameworks and GraphQL.

**Intern, Microsoft Research (RiSE group: Cloud Security)**          Summer 2019
*Microsoft Research, Redmond, WA*
   Project sponsor: Dr. Patrice Godefroid
- Techniques and tools for web API security testing.

**Intern, IBM Research (Storage)**          Summer 2018
*IBM Research, Almaden, CA*
   Project sponsor: Dr. Deepavali Bhagwat
- Provenance collection system for machine learning applications (VLDB'20).

**Software Engineer, IBM (GPFS)**          2012-2015, Summer 2016, Summer 2017
*IBM, Poughkeepsie, NY*
- Developed distributed applications and tooling for cluster management and file system testing.
- Worked with and trained test teams in the US, the UK, Mexico, Germany, India, and China.
- Created five patents on techniques for file system testing.

## EDUCATION

**Ph.D Computer Science**          2015-2020
*Virginia Tech*
   Advisor: Dr. Dongyoon Lee
   Dissertation: *On the Impact and Defeat of Regular Expression Denial of Service*

**B.S. Computer Science, B.S. Mathematics**          2008-2012
*Clarkson University*
   Research mentor: Dr. Takashi Nishikawa
   Honors Thesis: *Relating Synchronizability to the Topological Properties of Networks Using a Linear Classifier*

## CONFERENCE PAPERS

Cha, Wittern, Baudart, **Davis**, Mandel, Laredo. *A Principled Approach to GraphQL Query Cost Analysis.* Proceedings of the 28th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (**ESEC/FSE'20**). *ACM Distinguished Paper Award.*

Rupprecht, **Davis**, Arnold, Gur, Bhagwat. *Improving Reproducibility of Data Science Pipelines through Transparent Provenance Capture.* Proceedings of the 46th International Conference on Very Large Data Bases (**VLDB'20 Industry track**).

**Davis**, Moyer, Kazerouni, and Lee. *Testing Regex Generalizability And Its Implications: A Large-Scale Many-Language Measurement Study.* Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering (**ASE'19**).

Michael, Donohue, **Davis**, Lee, and Servant. *Regexes are Hard: Decision-making, Difficulties, and Risks in Programming Regular Expressions.* Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering (**ASE'19**). *ACM Distinguished Paper Award.*

**Davis**, Michael, Coghlan, Servant, and Lee. *Are Regular Expressions a Lingua Franca? An Empirical Study on the Re-use and Portability of Regular Expressions.* Proceedings of the 27th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (**ESEC/FSE'19**).

Wittern, Cha, **Davis**, Baudart, Mandel. *An Empirical Study of GraphQL Schemas.* Proceedings of the 17th International Conference on Service-Oriented Computing (**ICSOC'19**).

Fu, Ghaffar, **Davis**, and Lee. *EdgeWise: A Better Stream Processing Engine for the Edge.* 2019 USENIX Annual Technical Conference (**USENIX ATC'19**).

**Davis**, Coghlan, Servant, and Lee. *The Impact of Regular Expression Denial of Service (REDOS) in Practice: an Empirical Study at the Ecosystem Scale.* Proceedings of the 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (**ESEC/FSE'18**). *ACM Distinguished Paper Award.*

**Davis**, Williamson, and Lee. *A Sense of Time for JavaScript and Node.js: First-Class Timeouts as a Cure for Event Handler Poisoning.* Proceedings of the 27th USENIX Security Symposium (**USENIX Security'18**).

**Davis**, Thekumparampil, and Lee. *Node.fz: Fuzzing the Server-Side Event-Driven Architecture.* Proceedings of the Twelfth European Conference on Computer Systems (**EuroSys'17**).

## JOURNAL ARTICLES

Ozkan, Davis, **Davis**, James, Murzi, Knight. *Expectations and Experiences of Short-Term Study Abroad Leadership Teams.* Journal of International Engineering Education (**JIEE**).

## SHORT PAPERS

**Davis**. *On the Impact and Defeat of Regex DoS.* **ACM SRC – Grand Finals**. *Second place, graduate student division*.

Rupprecht, **Davis**, Arnold, Lubbock, Tyson, and Bhagwat. *Ursprung: Provenance for Large-Scale Analytics Environments.* Proceedings of the 2019 International Conference on Management of Data (**SIGMOD'19 Demo**).

**Davis**, Kildow, and Lee. *The Case of the Poisoned Event Handler: Weaknesses in the Node.js Event-Driven Architecture.* Proceedings of the 10th European Workshop on Systems Security (**EuroSec'17**).

## POSTERS

Vivek, Chinnakotla, Banna, Vegesana, Yan, **Davis**, Lu, Thiruvathukal. *Exemplars for Machine Learning: Towards Software Engineering & Reproducibility.* SIAM Conference on Computational Science and Engineering (**CSE'20**).

## PATENTS

Davis, **Davis**. *Injection of Simulated Hardware Failure(s) in a File System for Establishing File System Tolerance-to-Storage-Failure(s)*. IBM, U.S. patent application 20200264961. U.S. patent pending.

Davis, **Davis**. *Verification of the integrity of data files stored in copy-on-write (CoW) based file system snapshots*. IBM, U.S. patent application 20200242075. U.S. patent pending.

**Davis**, Davis. *File metadata verification in a distributed file system*. IBM, U.S. patent 10,678,755B2, granted Jun. 9, 2020..

Davis, **Davis**. *Testing of lock managers in computing environments*. IBM, U.S patent 10,061,777 B1, granted Aug. 28, 2018.

**Davis**, Davis, Knop. *Detection of file corruption in a distributed file system*. IBM, U.S. patent 10,025,788, granted Jul. 17, 2018.

## TEACHING

| | |
|---|---|
| **Instructor, ECE 368 Data Structures** | Fall 2020 |
| *Purdue University* | |
| **Co-advisor, VIP: Open-Source TensorFlow Software** | Fall 2020 |
| *Purdue University* | |
| **Instructor, Data Structures and Algorithms** | Fall 2019 |
| *Virginia Tech* | |
| **Instructor, Introduction to Programming in Python** | Spring 2019 |
| *Virginia Tech* | |
| **Track Leader, Rising Sophomore Abroad Program** | Spring 2018, Spring 2019 |
| *Virginia Tech* | |

## STUDENTS

*Undergraduates*

| | |
|---|---|
| Albert Sun (BS, Purdue University) | 2020-present |
| Jitesh Motati (BS, Purdue University) | 2020-present |
| Tuhin Sarkar (BS, Purdue University) | 2020-present |
| Jenna Ryan (BS, Clarkson University) | 2018-present |
| Jonathan Alexander (BS, Virginia Tech) | 2018-2019 |
| Christy Coghlan (BS, Virginia Tech) | 2018 |
| Celine Stewart (BS, Virginia Tech | 2017-2019 |

## INVITED TALKS

**Regexes Awry: Characterizing and Defeating Regex-based Denial of Service**     2020
*Clarkson University CS department colloquium*

**Regexes are Hard: Qualitative and Quantitative Perspectives**     2019
*NC State CS department colloquium*

**The Dangers of Copy/Pasting Code**     2019
*Episode of the Podcast "The Secure Developer": `https://tinyurl.com/DavisResearchPodcast`*

**Regexes in the Wild**     2019
*Virginia Tech department seminar*

**Academic Perspectives on Node.js**     2018
*Node.js Collaborator Summit, Vancouver*

**International Engineering**     Annual, 2015-2019
*Rising Sophomore Abroad Program, Virginia Tech*

## OTHER NOTABLE ACTIVITY

**Disclosed DoS vectors in Python core and Node.js core**
*Python: CVE-2018-1060, CVE-2018-1061; Node.js: CVE-2018-7158*

**Guide: "Don't Block the Event Loop (or the Worker Pool)"**
*`https://nodejs.org/en/docs/guides/dont-block-the-event-loop/`*

**VT Intramural Racquetball Champion (Singles, Doubles)**     Spring, Fall

## AWARDS AND RECOGNITION

ACM Distinguished Paper Award, ESEC/FSE 2020     2020

Outstanding Graduate Student Service Award, CS@VT     2020

Second place, Grand Finals of the ACM Graduate Student Research Competition     2020

First place, Graduate Student Research Competition, ESEC/FSE 2019     2019

ACM Distinguished Paper Award, ASE 2019     2019

ACM Distinguished Paper Award, ESEC/FSE 2018     2018

Microsoft Security Researcher Acknowledgments (Regex DoS)     2018

Graduate Fellow, VT Academy for Global Engineering     2019-2020

IBM Significant Contributor Award (Node.js)     2018

IBM Poughkeepsie's New hire of the month     2014

Frederica Clarkson Award     2012

Clarkson University's Outstanding Senior (x2): Mathematics, Computer science     2012

Clarkson University Phalanx Commendable Leadership     2011

## EXTERNAL SERVICE

| | |
|---|---|
| Reviewer, ACM Transactions on Software Engineering (TSE) | 2020-present |
| Member, ICSE Demonstrations Track | ICSE 2021 |
| Member, ESEC/FSE Artifact Evaluation Committee | ESEC/FSE 2020 |
| Member, CGO Artifact Evaluation Committee | CGO 2019 |
| Sub-reviewer: ASPLOS'18, EuroSys'18, MASCOTS'18, HPCA'19, CGO'19, ISMM'19 | 2016-2019 |
| Regional Judge, ACM ICPC | Fall 2015 |

## INTERNAL SERVICE

| | |
|---|---|
| Committee member, Purdue ECE Undergraduate Curriculum Committee | 2020-present |
| President, VT CS Graduate Student Council | 2018-2019 |
| Organizer, VT Systems Reading Group | Spring 2017-2020 |

## SELECTED OPEN-SOURCE PROJECTS

| | |
|---|---|
| **safe-regex** | Check if your regex is super-linear. 4.7M dependents. |
| **regexp-tree** | Analysis tools for regular expressions. |
| **marked** | Regex-based Markdown parser (I consult on ReDoS). 354K dependents. |

## PROFESSIONAL MEMBERSHIPS

Member, Association for Computing Machinery