**TEAM PROFILE**

**ENGINEERING UBER SYSTEMS TO COMBAT FRAUD**

JUNE 30, 2016

BY CONOR MYHRVOLD & MANI PARKHE

Walk into a conference room on the 16th floor of an [Uber Engineering](https://people.uber.com/eng/) building on Market Street in San Francisco. You enter an intense discussion around a table with software and data engineers, data scientists, modeling experts, and even a product manager. The topic? How to determine a fraudulent user.

Fraud prevention is one of the fastest growing areas of research and development at Uber. As our platform has grown, so has the international underworld that tries to undermine it. There’s rider fraud, driver fraud, and sometimes both intertwined. What we see in China one day can arrive in Toronto the next. Soon, identical fraud patterns crop up across six cities on three continents. What does the real-time system that combats and learns from this fraud look like?

Engineering leader Eddie Ma is at the forefront of Uber’s fraud prevention efforts, leading several interconnected engineering teams that work in coordination to reduce fraud across Uber’s software platform. Here’s more from his team on how Uber engineers systems to fight fraud in 2016 and beyond.

**Eddie, Engineering Management**



Eddie Ma: Fraud prevention at Uber started with a small band of engineers and data scientists building specific production rules to detect known fraud patterns. As Uber grows both internationally and into different vertical markets, our platform has evolved rapidly with the latest technologies in big data processing, machine learning modeling and platforms, and distributed real-time transactional systems.

Our success relies on machine learning, data science, and large-scale distributed systems to innovate in these technical domains, and it is this next generation of development that will build our success. Experienced systems engineers will transform our existing infrastructure into highly scalable, online machine learning platforms that can deal with ever-increasing transactional demands in real-time. Data scientists and machine learning experts will build complex models to detect patterns with live and real-time data. Data engineers will build online analytics platforms and pipelines to process our rapidly quadrupling data. Solutions specialists will find the best methodology for diffusing a real-life threat, and UX experts will design and build applications and tools to standardize the anti-fraud process in global operations.

**Raja, Systems Engineer**



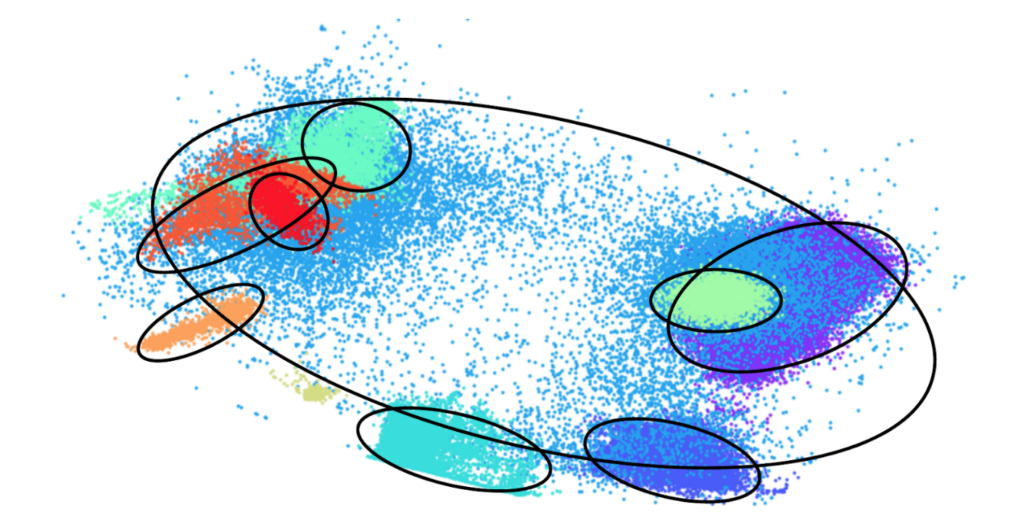
Raja Shekhar Alli: I’m currently focused on the fraud prevention platform, building scalable distributed systems to fight rider and driver fraud in real time. Since people take trips and we charge for trips right after completion, we need to make decisions within milliseconds to stop fraud before it even has a chance to occur.

Fraud requires a lot of creative solutions, especially with machine learning systems, which are critically important to our work. Developing large-scale applications to ferret fraud out is not straightforward. We need to understand how services are tied to each other and how data flows across the Uber infrastructure ecosystem. Our long-term goal is to construct scalable solutions that make it difficult to commit fraud from the start: fraud prevention rather than just detection and mitigation.

**Jinsong, Machine Learning and Modeling Expert**

Jinsong Tan: Uber fraud prevention is a dynamic environment. Challenges arise fast, and so do the solutions we develop to overcome them. We are encouraged to work on problems from end to end. I can fully exercise my skill sets in software engineering, machine learning, data science, and even some game theory.

I studied behavioral economics and game theory in grad school, and here at Uber, fighting fraud means developing insights into the economics of fraudulent behavior. Ideally, that entails designing a system that is, in game theory jargon, [incentive compatible](https://en.wikipedia.org/wiki/Incentive_compatibility). We want to make our platform resilient such that it simply does not make economic sense for frauds to operate. Over the past year, we have had really successful outcomes from tweaking our dispatch algorithms to make our system more incentive compatible—in other words, fraud resistant—and these results will only matter more as the business continues to grow.



Clustering helps us discover fraud rings using [PCA](https://en.wikipedia.org/wiki/Principal_component_analysis) derived independent features.

**Alain, Data Engineer**



Alain Rodriguez: I ensure that data used to make fraud reduction decisions will flow accurately and consistently. Right now, we are in the middle of moving our services to real time—we’re talking seconds—from minutes and hours. The more real-time data, the quicker we can stop fraud and bad actors from undoing the work that the rest of the company does.

People are surprised in general of the many creative ways people try to defraud us. In international markets, the organized fraud rings require us to build a system that has distributed data processing but that can also go real-time across multiple data centers. Fraud is global.

On the zanier side, people have put phones on their dog and taken it for a walk in the street. Or put it on a train. These are real phones with real driver accounts, so it comes down to GPS tracing: Is it a dog? A train? Or a legitimate trip?

**Tara, Risk Management Analyst**

Tara Mitchell: As risk management analysts, we do the digging ourselves and find fraud patterns. A lot of my job is looking through accounts and data. You have to think like a fraud and figure out what’s abnormal. My work is determined by the biggest financial impact to Uber in a given day. I’m responsible for the US, Canada, and Southeast Asia markets, so I work with people in or near those regions. I talk regularly with teams in [Kuala Lumpur](https://www.uber.com/cities/kuala-lumpur/), [Jakarta](https://www.uber.com/cities/jakarta), and [Australian cities](http://love.uber.com/australia/).

We pull from a lot of diverse backgrounds. I come from the ops side, doing manual reviews and learning patterns. We also bring in people with financial and analytical skills who like looking at data and identifying patterns and trends. Not everyone comes with a fraud-focused role previously, but everyone has analytic pattern recognition and an investigative skill set.



When I determine a pattern, I’ll look at 30 days of back data and isolate a behavior into a distribution of different features. With many hundreds of combinations, we test to see how we can stop most of this fraud pattern from occurring. Then, I’ll recommend a fix. I’ll build a set of logic to predict if they’ll engage that pattern in the future. We do the analysis, write the code, put it through review, and then deploy it, all in a couple of days. For example, I wrote something Monday that’ll go live this Thursday. We have to get in front of these patterns, as every moment we’re not deploying the fix means money lost.

End-to-end ownership of pattern identification to solution is a big theme in my job. Since coming here, I’ve needed to become an expert in SQL to test models off of different datasets. I’ve had to learn to program in Python because [that’s what a lot of data scientists use here](https://eng.uber.com/emi-data-science-q-a/). We’re going to migrate to Hadoop, and that’s one more technology I’ll have to learn.

**Ting, Data Scientist**

Ting Chen: As fraud prevention data scientists, we develop models for identifying fraud on large scales. First we have to define fraud-associated metrics: what common traits do fraudulent trips have? What are the telltale signs for a fraudulent driver or rider?

Fraud is an interesting problem because it’s like a chameleon; it tries to blend in over time to avoid being detected. So fraud metrics change over time as the camouflage techniques become more sophisticated. In China, the fraud black market itself is very complete. Fraudulent users put up their own online advertisements to hire data scientists and engineers to perpetrate fraud. They have 24/7 customer service in mobile messaging platforms to help frauds carry out their work undetected.

The major tool I am using to fight fraud at the moment is building a specific machine learning model for the driver signup process. We want to detect the signals of fraudulent behavior before fraudulent trips actually occur. If we want to predict whether a new driver partner signing up is a high fraud risk, we can predict that based on information gleaned from the onboarding process. For example, we can look at the car’s license plate number for registration, the device used to sign up—and whether the same one has been used before—and whether that same person has previously attempted to sign up. It goes on from there, but that is what I can share with you for the time being!

**Gaurav, Solutions Specialist**

Gaurav Agarwal: I work on the fraud program team. We are responsible for all business metrics associated with fraud at Uber, such as chargebacks (customers disputing charges on their credit cards), refunds, uncollectibles (when we can’t collect money at the end of the trip, such as when we attempt the charge, but the charge fails), and finally account takeovers (when someone tries to log into another customer’s account).

Our platform allows us to do near-real-time feature lookups and run machine learning risk engines. What I’m building will be used in other decision points in the life cycle of a rider or driver. I specifically work on risk associated with account logins at Uber. That involves building the end-to-end systems to reduce this from happening.

Typically frauds get passwords from other websites and then try them out on Uber properties. Another company will have a data leak, those logins and passwords will be circulated online afterward, and then frauds will check for similar emails and login credentials across the web, including Uber. My work is focused on effectively blocking these login attempts from untrusted sources.

Uber encourages end-to-end ownership. This means we own login-related risk for account takeovers, period, from building the platform layer and machine learning models to building architecture for this use case that’s extensible enough for other use cases at Uber.



Uber’s fraud prevention engineers try another type of curling for a team outing.

*This article is co-authored with Mani Parkhe, software engineer on the*[*Fraud Prevention Data Infrastructure*](https://www.uber.com/careers/list/19387/)*team.*

*Interested in how we fight fraud at Uber?*[*Learn more about Uber Engineering positions involving fraud prevention, mitigation and detection on the Uber Careers page.*](https://www.uber.com/careers/list/?city=all&country=all&keywords=fraud&subteam=all&team=engineering)

*For additional articles about Business Intelligence engineering, see:*[*Meet Uber Amsterdam Engineering*](https://eng.uber.com/amsterdam-team-profile/)*,*[*Rewriting Uber Engineering: the Opportunities Microservices Provide*](https://eng.uber.com/building-tincup/)*, and*[*Learning as a New Grad on the Uber Engineering Money Team*](https://eng.uber.com/new-grad-money-team/)*.*