ADVERTISING-BASED MEASUREMENT: A PLATFORM OF 7 BILLION MOBILE DEVICES

MOBICOM 2017 OCTOBER 2017

Mark D. Corner, Brian N. Levine, Omar Ismail, Angela Upreti
UMass Amherst

NINE PEOPLE

Nine people. This is what often constitutes the extent of many mobile systems evaluations.

All too often we are building systems, testing them on nine people, and drawing scientific conclusions. We can often get away with this in conference reviews if the system is compelling enough.

DIVERSITY

And where did we get those people from? From our own lab? Are you testing your systems on your own graduate students? How many of them are women? How many of them are older, or younger, or from a disadvantaged household, or non-technical? Would your conclusions hold up if you tested them on a more representative sample?

SCALE

what we really want to do is to scale our experiments and data gathering up to massive scale with plenty of diversity.

We also want to go *global*. We want to be able to conduct experiments in places where we aren't. Other geographies, other countries, with people we haven't met. If you work at google or MS etc, you have access to devices, data, and networks, that other people don't. These things aren't available to your average researcher.

So how are we going to do this? Ads.

ADVERTISING BASED MEASUREMENT

Wait, why ads? What is happening in this ad? A lot more than most people think. This isn't just a picture (which we call the "creative"). Ads typically run in an embedded webview (or an iframe), with the full power of javascript and html. Through some outlets for ads, specifically something called "Real Time Bidding" you get to control what that javascript and html does)

All of this takes place when the impression is *seen*, not when the user clicks.

You have to think of ads as an instant window into every mobile device in the world.

So what, I can run some javascript, this doesn't let me do much.

MORE THAN YOU THINK

DeviceOrientation & DeviceMotion events - CR

Global 1% + 90.19% = 91.2%

API for detecting orientation and motion events from the device running the browser.

Current aligned | Usage relative | Date relative | Show all

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android Browser *	Chrome for Android
			49						
			59					4.4	
	14	54	60			10.2		4.4.4	
11	15	55	61	10.1	47	10.3	all	56	59
	16	56	62	11	48	11			
		57	63	TP	49				
		58	64						

Notes | Known issues (2) | Resources (11) | Feedback

Battery Status API - CR

Global 60.89% + 8.23% = 69.12%

unprefixed: 60.89% + 8.2% = 69.09%

Method to provide information about the battery status of the hosting device.

Current aligned | Usage relative | Date relative | Show all

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android Browser *	Chrome for Android
			49						
			59					4.4	
	14	54	60			10.2		4.4.4	
11	15	55	61	10.1	47	10.3	all	56	59
	16	56	62	11	48	11			
		57	63	TP	49				
		58	64						

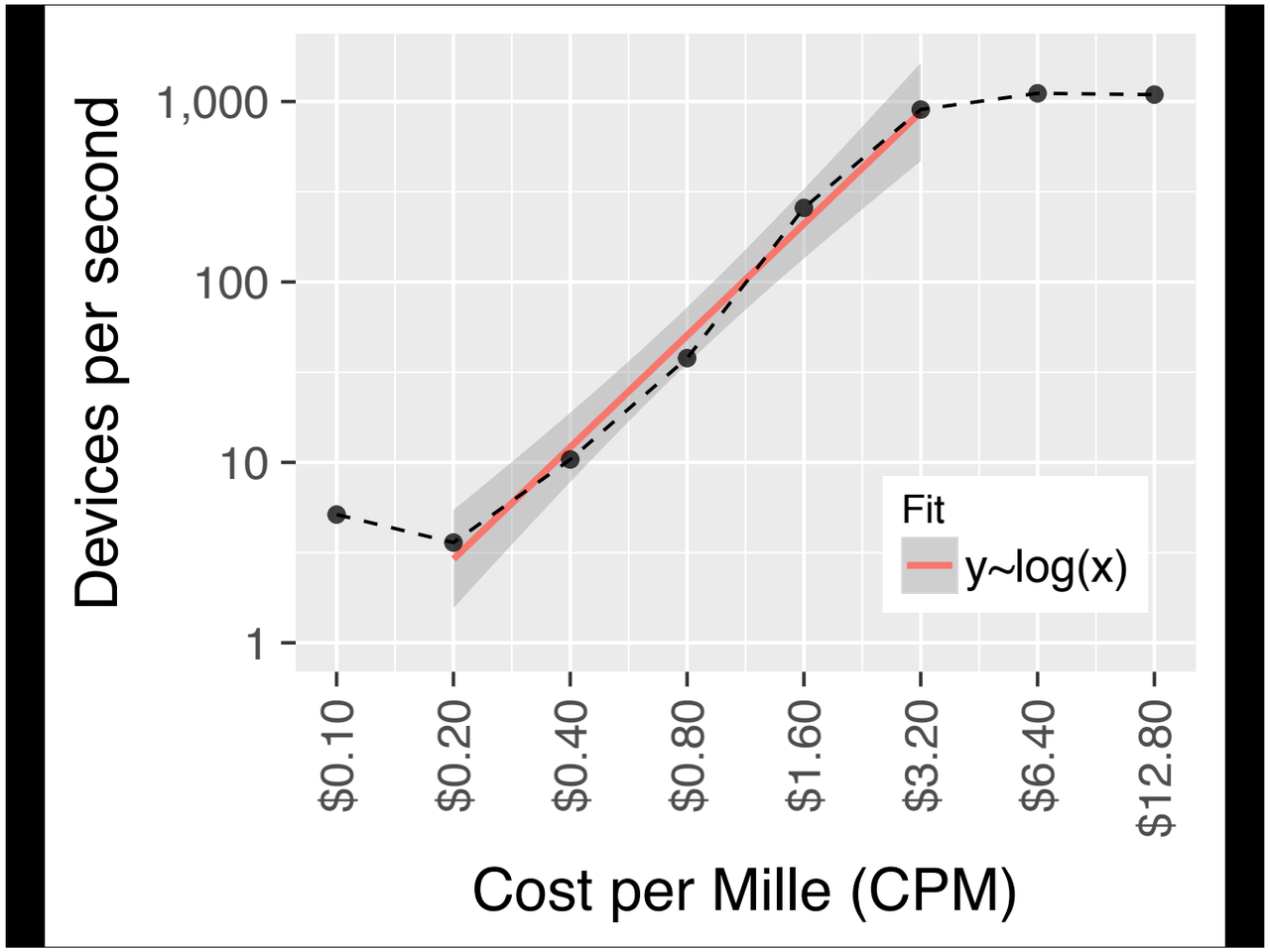
As two examples, you can get measurements from the accelerometer. You can get the status of the battery and if the device is plugged in (only on Android). But there is more…. You can get the internal IP address of the user (android only), even behind a NAT, you can measure bandwidth from the device to other servers. You can fetch more images to see if those images are blocked. Some stuff is off limits because it requires an extra permission to get (like the camera). But we feel like given the creativity of people here, there is a lot more.

You can also leverage data that is passed into the ad via the app and the advertising SDK. These apps don’t really have a whole lot in common, but both rely very heavily on accurate location. And each of them share that location with the advertisement. So as a researcher, in many cases you can get at extremely accurate locations.

WHAT IS THIS GOING TO COST ME?

Ads are very inexpensive. The cheapest ads are the smallest ones (320x50) and you can buy them for approximately 10 cents per 1000 ads (this is known as a CPM). To place this in perspective, this is the Vietnamese Dong (VND) which is the least valuable currency at 22k VND to the USD. One ad impression costs 2.24 VND. That bill up there can buy you 450 data points (4.5 cents).

Everything you will see in this paper, all totaled up, cost us about $450 in advertising.

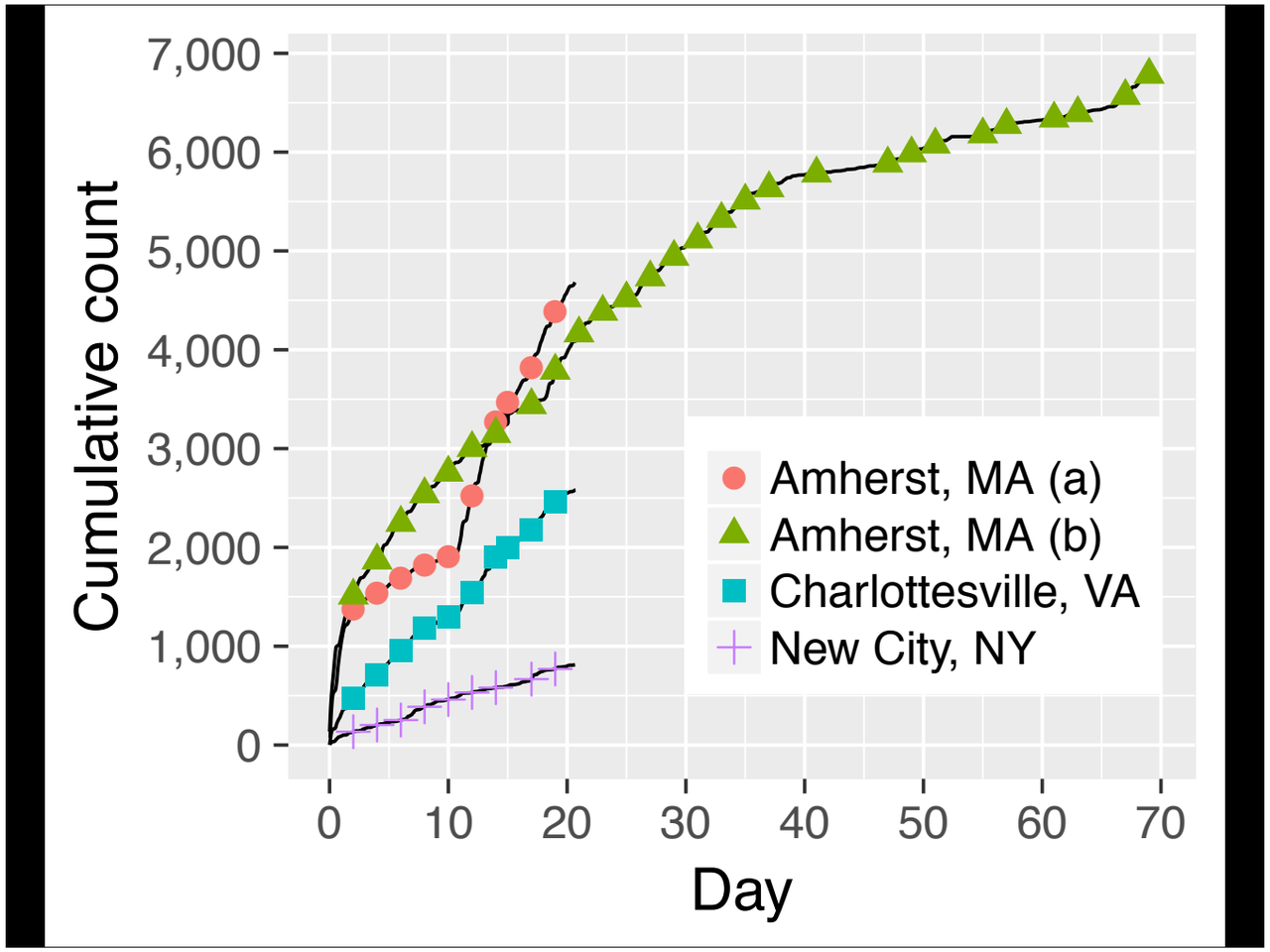


It is a bit more complicated than that. What happens in ads is a miniature “second-priced” auction. You bid the maximum you are willing to pay per mille (which is 1k ad impressions). You don’t pay what you bid, you pay the second highest price.

Just to characterize this system, we wanted to see how many unique devices (in the US) that we could show an impression to, on a log scale. At several dollars per thousand ads you win all of the auctions and can discover about 1k new devices per second.

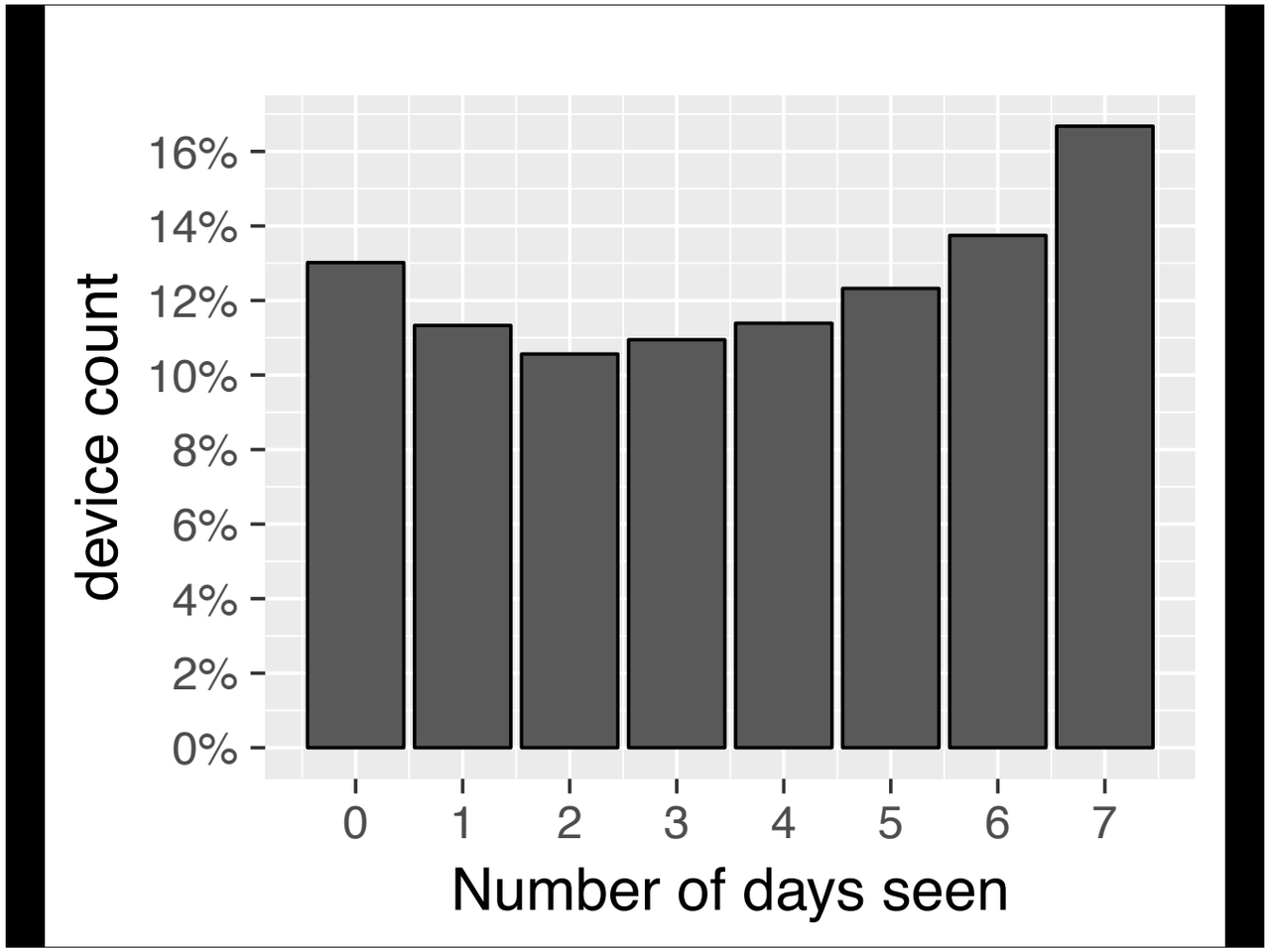
If anything bothers you about the graph, it should. At the right side of the graph we are spending about $3.20 per second or about $280k / day. So we sampled random sets of IPs and targeted with ads them to estimate the total. Details are in the paper.

And there is plenty to work with: the platform we used had access to 34 billion ads per day.



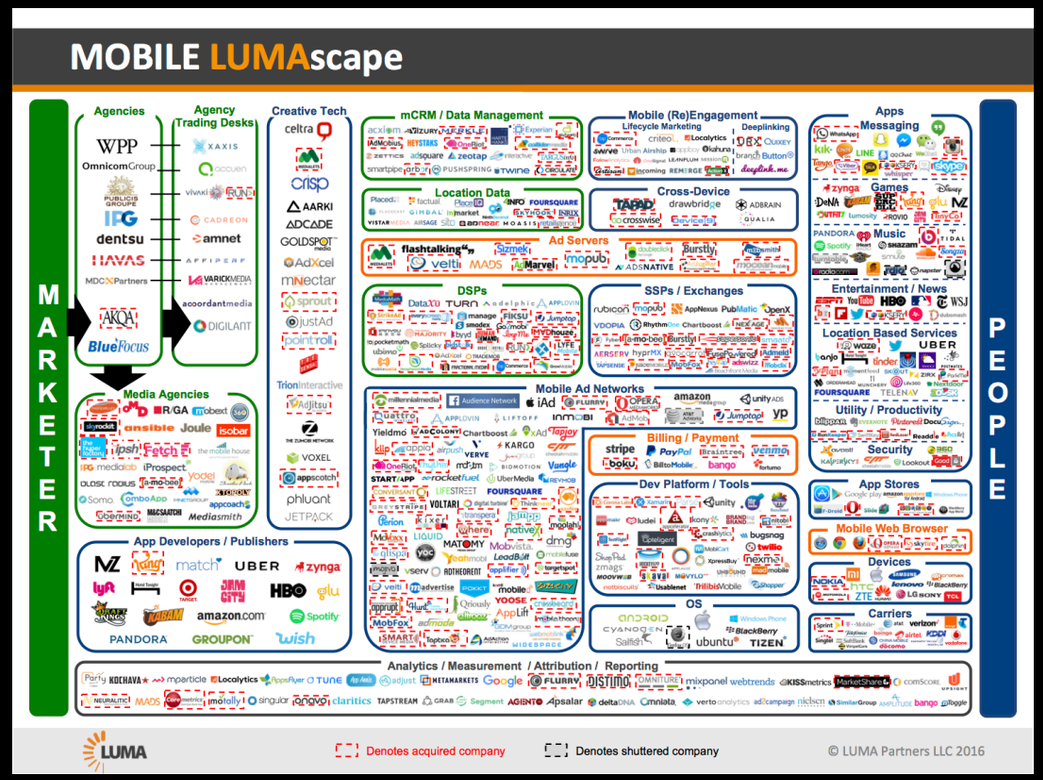
We also wanted to see how well we could target a defined geographic area. We set up campaigns to target three different cities (we did Amherst twice to see if the results were repeatable) and counted the number of unique devices we reached by each day.

Based on some assumptions about the number of mobile devices in Amherst, we reached approximately 6% of the devices after 20 days and 12% of the devices after 60 days. This is pretty impressive considering how little effort is involved. Imagine asking, and getting, some kind of data point from 12% of the devices, and what that might involve.



Ok, but what about longitudinal experiments? What if we need a data point from devices several days? In this experiment we targeted ~8k devices we had seen previously. We measured for a week and limited impressions to once per day per device. This graph shows the fraction of devices we saw for each number of days.

13% of the devices we never saw during the week, but more than half of the devices were seen on at least 4 of the seven days.



So there is a virtually unlimited opportunity for gathering data, so why don't research scientists do much work on ads? After all ads are essentially the economic engine of a large portion of the internet. There is some work on ad optimization, and a bit of work on privacy in ads (typically with the bent of saying that ads are evil)

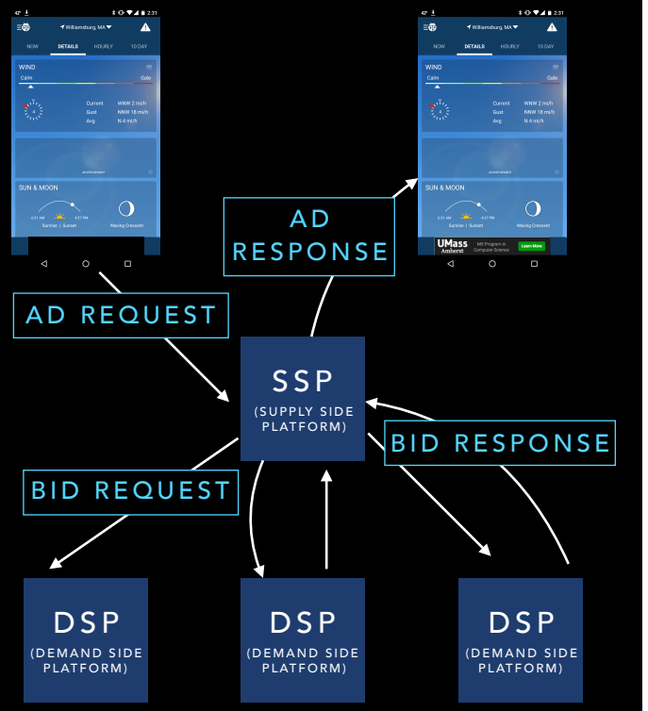
But there are lots of places to buy ads and the lingo is complicated and not really set up for hobbyists. Part of my goal here is to teach you all about advertising.

I worked as the CTO of a company called Fiksu. I will buy 10 beers for anyone who can find it on that chart.

There are lots of places to buy ads, but we concentrate on buying ads through "Real Time Bidding". Real Time Bidding is the "bare metal" of advertising and you can gain a lot of control over the ads itself that is typically not available in ads in Facebook etc.

REAL TIME BIDDING

- Every single impression is the result of auction involving (dozens, hundreds of) DSPs
- An SSP controls the process and the DSPs bid
- Highest bid wins
- Cost of impression is the second-highest bid
- Total time is around 200ms



What does this process look like when you peel back the layers?

The three main parties are the user+device, a Supply Side Platform, and a Demand Side Platform.

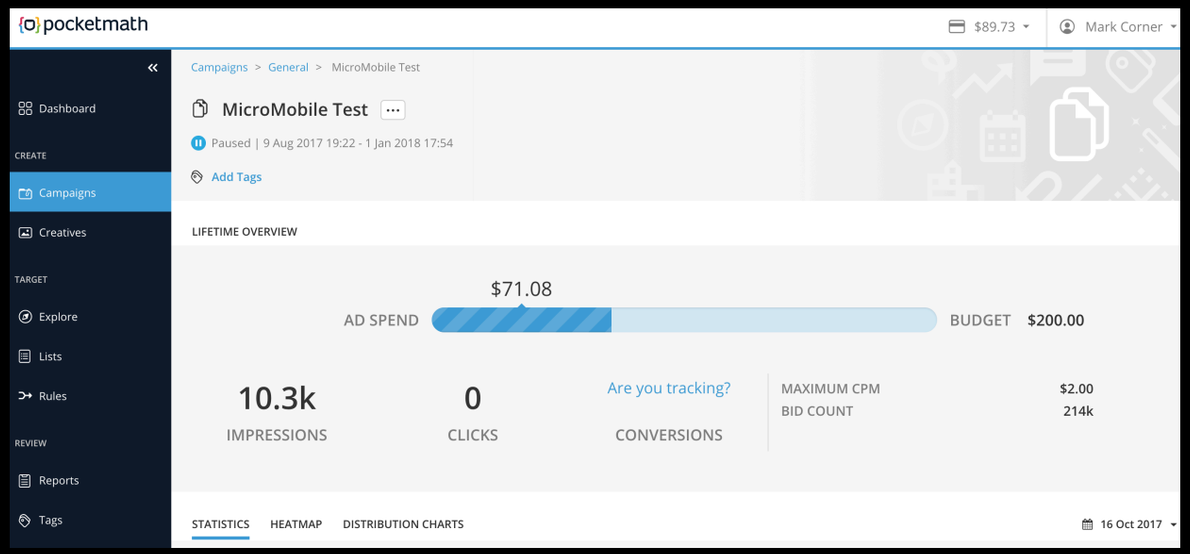
(GO THROUGH PROCESS)

We can't be an SSP since that involves getting code into a large number of apps. So we just need to be a DSP.

Unfortunately, answering 500k+ bid requests per second is pretty expensive and complicated.

This is the HERMÈS ALLIGATOR CONSTANCE 18 which costs $68k. That will let you run a DSP for only a couple of months. Hilariously I now see a lot of ads for this handbag.

SELF SERVICE



Instead we use a “self-service” DSP, which does the bidding our our behalf. They have an API and allow us to target lots of parameters, change bids, etc.

"AD TAGS" AKA JAVASCRIPT

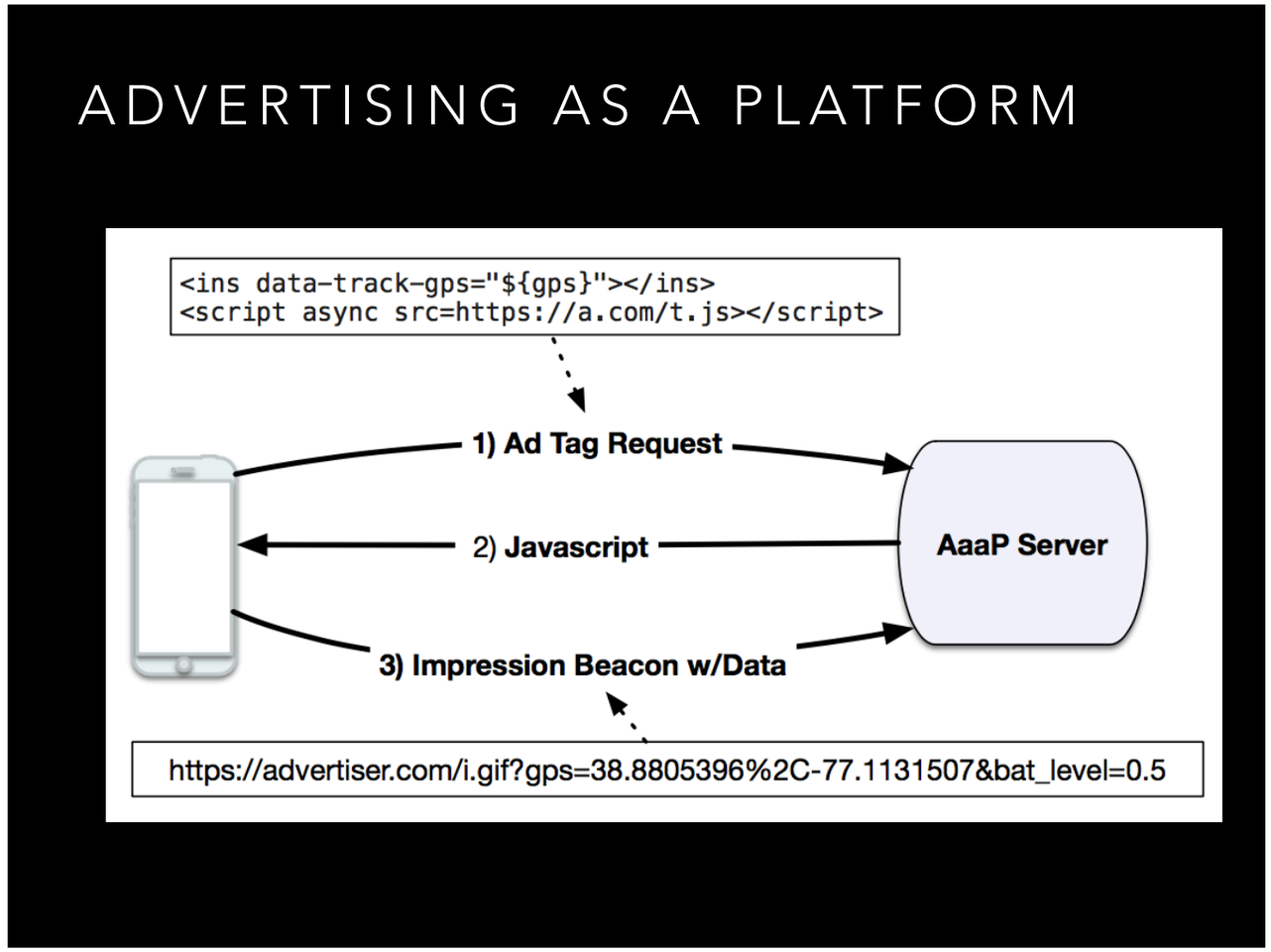
```
<script async src=https://
advertiser.com/t.js</script>
```

But instead of a simple html document with an image, you can also ask the DSP to insert an theme javascript tag. This will then fetch javascript from your server and run it in the embedded webview. What can that javascript contain? Anything you like.

```
<ins data-track-gps="${gps}"></ins>
<script async src=https://advertiser.com/t.js</script>
```

```
<ins data-track-gps=42.395317,-72.531336></ins>
<script async src=https://advertiser.com/t.js</script>
```

You can also use “macros” that get filled in by the DSP. This allow you to access data, such as the name of the app where the ad appears and other information only available to the native app, such as GPS.



So we built a system called Advertising as a Platform to take advantage of advertising based measurement. It has a lot of pieces, but chiefly it interacts with a DSP to buy advertisements that contain javascript. That javascript measures things and then reports that data back to our servers for analysis.

But what are the challenges here to overcome?

FIVE CHALLENGES

- Ads provide inaccurate location information
  - Improve MaxMind
- Macros often provide abstracted data
  - Measure bandwidth from iPhone models
- Ads only provide samples
  - Measure user-battery interactions
- Ground truth identities are difficult to obtain
  - Identifying users using digital fingerprinting
- Devices are held by humans, so what are the ethics?

We looked at five specific challenges and developed generalizable techniques to deal with them.

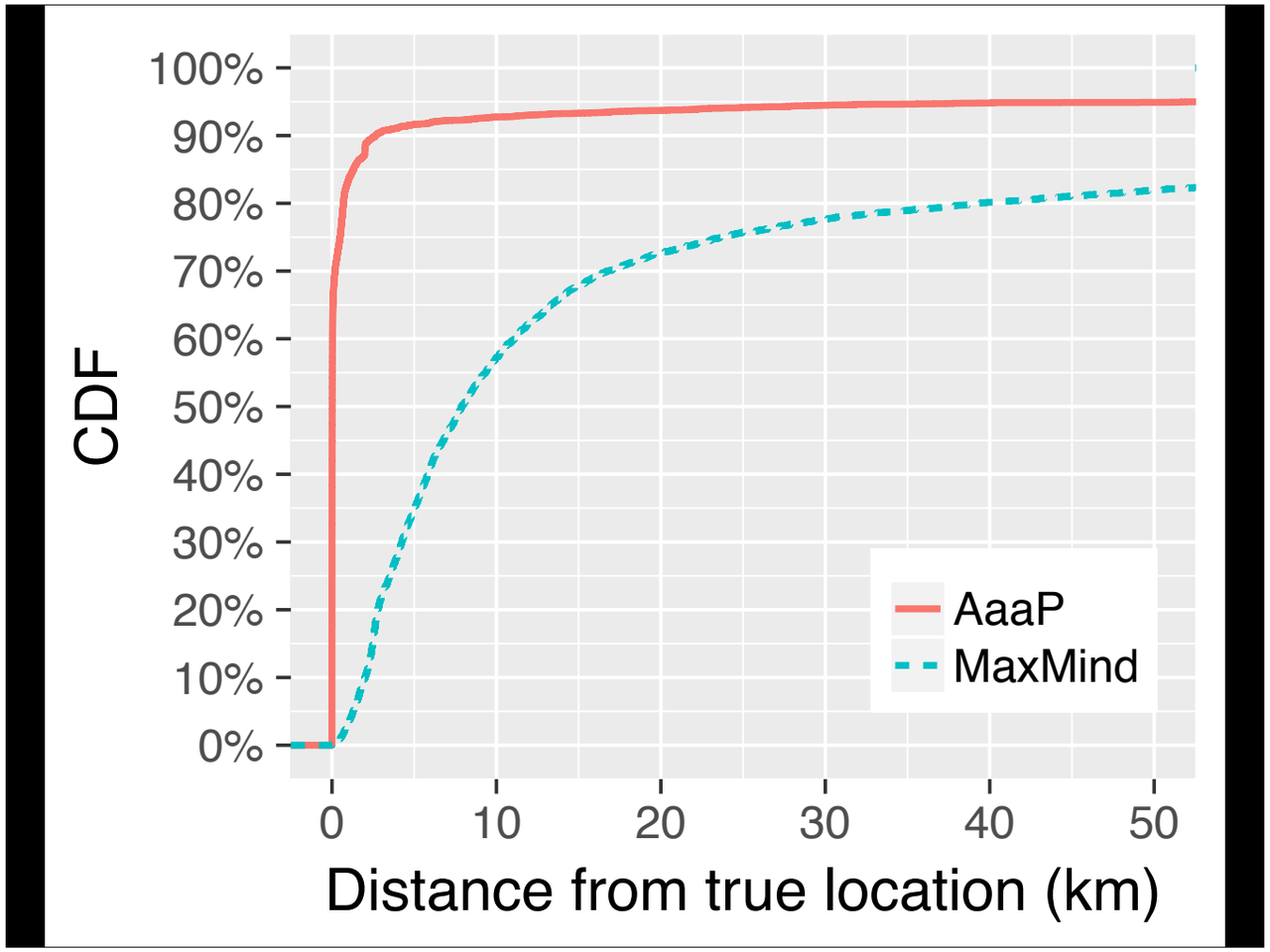
Each is used in a specific data collection experiment to demonstrate the technique and the utility of the overall system.

INACCURATE LOCATION

Almost all of the impressions come with something labeled as GPS. But if you dig into it, many of those apps never asked for permission to access location. And what you find is that most of them are derived from the IP Address. Especially in cellular networks this information is pretty much unusable. However, if you dig into the data, you can find apps that do provide accurate and identifiable location information. What kinds of apps? Primarily two kinds: location-based dating and the weather.

App Name	Perc.	Notes
GROWLr Android	99%	Required
GROWLr iOS	99	Required
Grindr iOS	98	Required
Weather Mood	97	For weather
Grindr Android	96	Required
The Weather Channel iOS	84	For weather
My Clock Android	81	For weather
The Weather Channel Android	78	For weather
MeetMe: Chat Android	76	Encouraged
TextNow Android	71	Asks at install
My Clock Free Android	67	For weather
TuneIn Radio iOS	61	Local stations
TuneIn Android	61	Local stations
MB3: Mixer Box	61	Asks at install

And how often do users grant the app permission? It depends, but they grant it a lot. Some apps are unusable if you don't give it permission, but even with weather apps, the percentage of people that grant access is very, very high. And these apps have tens of millions of users so you can get highly accurate location information tens of millions of devices.



What can you do with the data? For one, you can improve the location information for those highly inaccurate GeoIP addresses. So we conducted an experiment where we used IP addresses with at least two different devices. We then compared the accuracy of predicting the other location with the location predicted by the popular MaxMind GeoIP database. What we find is that using the advertising-based measurements improved accuracy by 53x over MaxMind.

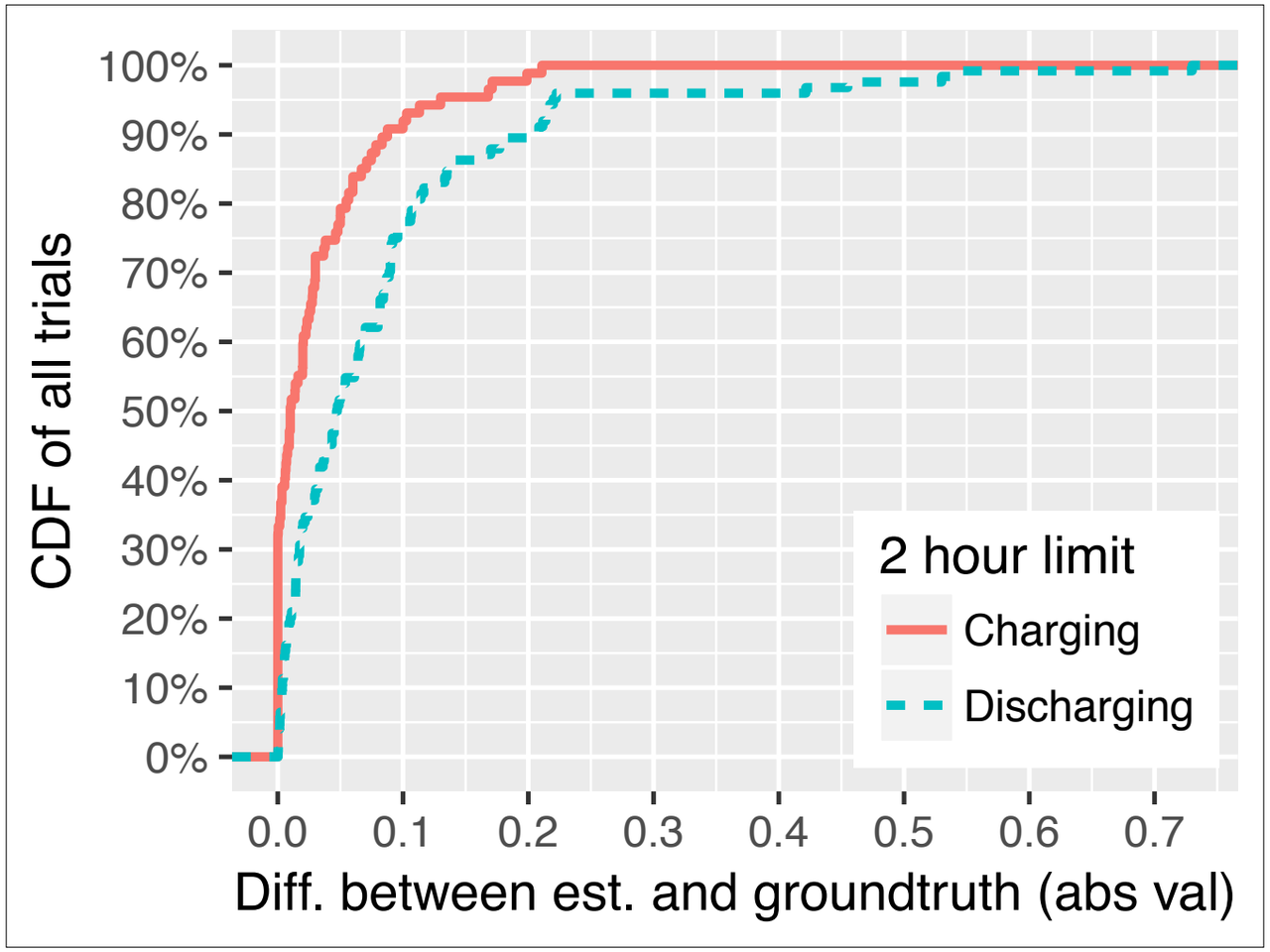
Nilanjan Banerjee, Ahmad Rahmati, Mark D Corner, Sami Rollins, and Lin Zhong. Users and Batteries: Interactions and Adaptive Energy Management in Mobile Systems.

18 users, measured continuously. Almost all of them were computer scientists.

Took a while to recruit users and get monitoring software installed on their machines. A bit easier now to do, but manual recruitment is still necessary.

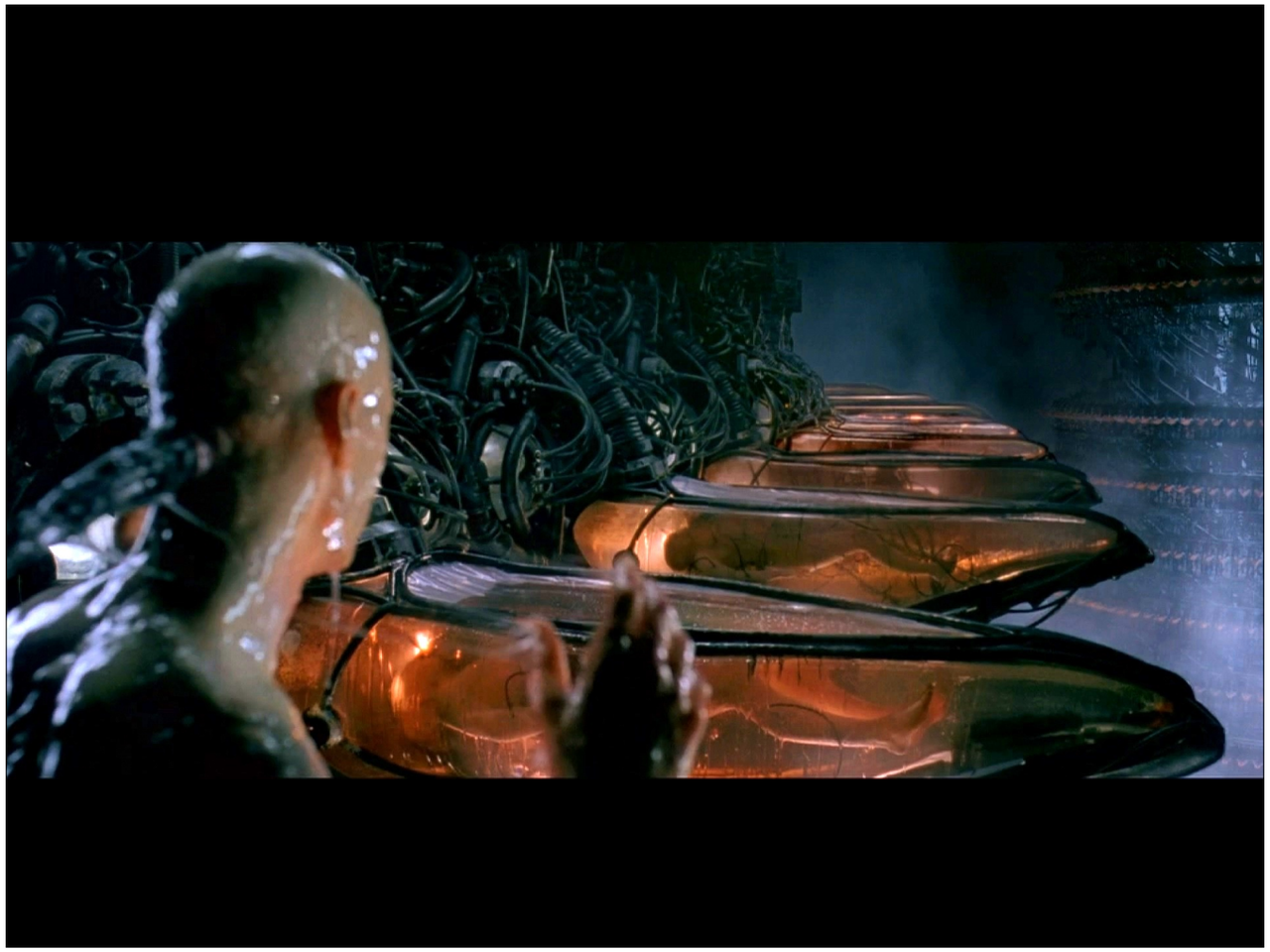
What if we sample user-battery interactions, and then predict or interpolate between samples?

We targeted 38k devices with ads to gather information about the state of their battery. We can't directly pace ads to devices, so we used 4 six hour campaigns that would gather up to ten readings from each device. We recorded what percent the battery was full and if it was plugged in or not.



Then for a series of charging or discharging points for a device, we leave out the last point as ground truth and try to predict it using a simple linear model.

Using this we can predict the last point in the sequence. Such a model is accurate to within 10% of the battery capacity 90% of the time within 2 hours. After that prediction accuracy degrades rapidly.



What about human subjects concerns?

All of the studies that we conducted as part of this were vetted and approved by the UMass IRB. They had lots of questions and I had to appear before the board.

One of the critical issues is that of “Informed Consent”. Normally one gets a description of a research experiment and has to agree to participate. You may notice that we never did that. It is permissible to do this as long as there is “no more than minimal risk”, it is “impractical” to obtain consent, and “the probability of and magnitude of harm” is not “greater than that from daily life”. This research meets all of those criteria and was thus allowed.

ADVERTISING-BASED MEASUREMENT: A PLATFORM OF 7 BILLION MOBILE DEVICES

MOBICOM 2017 OCTOBER 2017

Mark D. Corner, Brian N. Levine, Omar Ismail, Angela Upreti
UMass Amherst

Statement from the SIGCOMM 2015 Program Committee: The SIGCOMM 2015 PC appreciated the technical contributions made in this paper, but found the paper controversial because some of the experiments the authors conducted raise ethical concerns. The controversy arose in large part because the networking research community does not yet have widely accepted guidelines or rules for the ethics of experiments that measure online censorship. In accordance with the published submission guidelines for SIGCOMM 2015, had the authors not engaged with their Institutional Review Boards (IRBs) or had their IRBs determined that their research was unethical, the PC would have rejected the paper without review. But the authors did engage with their IRBs, which did not flag the research as unethical. The PC hopes that discussion of the ethical concerns these experiments raise will advance the development of ethical guidelines in this area. It is the PC's view that future guidelines should include as a core principle that researchers should not engage in experiments that subject users to an appreciable risk of substantial harm absent informed consent. The PC endorses neither the use of the experimental techniques this paper describes nor the experiments the authors conducted.

Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests

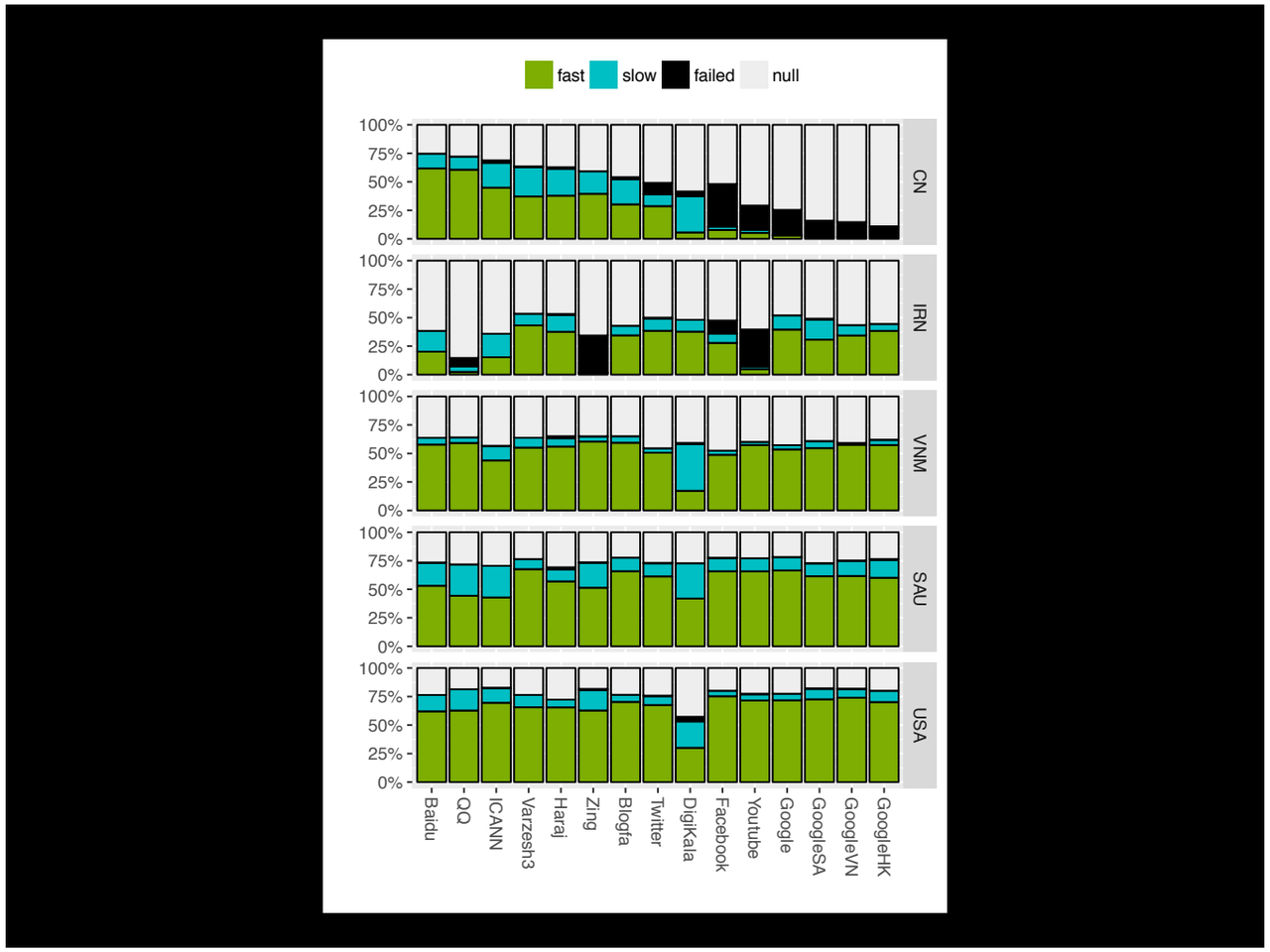
Sam Burnett
School of Computer Science, Georgia Tech
sam.burnett@gatech.edu

Nick Feamster
Department of Computer Science, Princeton
feamster@cs.princeton.edu

But at the boundary is doing experiments on censorship. This is a brilliant paper from Sam Burnett and Nick Feamster on measuring censorship from web pages. You essentially put a hidden image into a web page and measure if it loads or not. This is the only paper I have read that came with a warning label.

Early on we thought this was quite brilliant and used a similar technique, but inside of ads. But the advantage is that we can target very specific places, and we don't have to attract any users. I think I built this system in about a day and could measure censorship in any country I wanted (perhaps not NK)

This went into our MobiSys paper. It got great reviews, except one reviewer who thought this was completely unethical. There was a 40 minute discussion at the PC meeting who eventually rejected paper because of this problem. So the following slide wasn't in our Mobicom submission.



We can conducted a very similar experiment, by buying you can buy advertisements in China, Iran, Vietnam, Saudi Arabia etc. We put icon loads into the background and measured if they failed or never finished loading. You can see heavy censorship of certain domains in countries such as Iran and China. Interestingly you can dig into the data and find that censorship in China is not even, with the province surrounding HK and Macau allowing traffic to sites like FB.