FTC PrivacyCon
January 14, 2016
Segment 4
Transcript

KEVIN MORIARTY: OK. Welcome back, everyone. My name's Kevin Moriarty. I'm with the Federal Trade Commission, and this is Session Four on the Economics of Privacy and security. First, we have Jens Grossklags from Penn State University presenting an empirical study of web vulnerability discovery ecosystems.

JENS GROSSKLAGS: So welcome to the first of two talks in the session that are actually about security. This is joint work with Mingyi Zhao and Peng Liu. We are all at the College of Information Science and Technology at Penn State University.

So my talk about a topic of bug bounties and vulnerability discovery that is mostly conducted by external researchers that are often called white hats. In 1995, the first bug bounty program was founded by Netscape that invited external security researchers who scrutinize its services.

Since then, we had a number of other company-sponsored programs emerging that were run in an independent fashion. However, more recently, we actually observed the merchants of so-called back bounty platforms-- two of them, HackerOne and Wooyun, which are the focus of our study. Wooyun was founded in 2010, and it's mostly focus on the Chinese market. HackerOne operates in Europe and in the United States mostly, and was founded in 2017.

So the motivation for our study is to better understand how these web vulnerability ecosystems actually operate, and whether they make a significant contribution to web security. We also want to provide useful data for the policy [INAUDIBLE], for example, on the limits of vulnerability research and practice.

Our approach is to do an in-depth empirical study of these two ecosystems, and in our paper, we take a very broad approach in the sense that we try to understand how organizations, white hats, black hats, the public interacts on these third-party vulnerability platforms. But in the presentations, I will mostly focus on the perspective of companies and organizations.

So the two programs that we look at have a couple of common aspects, mostly that they're very popular, and lots of white hats are interacting on them, and also a lot of vulnerability reports are

irrespective of the wishes of the company, after 45 days, the whole technical details of the discovered vulnerability will be communicated to the public.

So there are some differences in the type of data we have about the platform, so we cannot always directly contrast and compare the two. But what we can do is in five broad categories, provide somewhat of a comparison on how these platforms actually operate. The first one is participation. And what we observe here is that on HackerOne, the number of public programs that are run is limited to about 100. And all of those are IT companies.

In contrast, on Wooyun, we see a much broader portfolio of companies that are more or less coerced to participate on to platform. And interestingly, you see here, a lot of organizations are typically are not known to run bounty programs by themselves, like government institutions, education institutions, and also financial institutions.

So the first takeaway that we have is that the white hat-initiated model allows for a much broader participation, and which may be good in the sense of web security. The more limited participation model of platforms such as HackerOne, of course, raises then the question of how these platforms can actually encourage more companies to participate.

A second issue that you want to then explore is the quality of the submissions of what you observe here, in particular on the platform of Wooyun, is that you have a very broad range of types of vulnerabilities that are submitted. And in 44% of these cases, these are actually classified as high severity vulnerabilities.

On HackerOne, this a little bit harder to determine from publicly available data. However, if you actually peruse the bounty amounts of that are paid through the white hat hackers, and also look into the policy statements, by combining these two data points, we can actually also then infer how many vulnerabilities are of high and medium severity, which is plotted on the slide.

So here, we can also conclude that across these two programs, white hats actually make significant contributions to the security of these websites by contributing high severity vulnerabilities. But more broadly speaking, the white hat initiated model that we see on Wooyun seems to harvest more of these vulnerabilities in an efficient fashion.

Now the question arises how well actually these different platforms, and the particular companies associated with them, can actually respond to these submitted vulnerabilities. And here, we see some interesting differences. When we look at Wooyun, we actually that, in particular, those very popular companies as measured by Alexa rank, you see that most of them can actually adequately respond to the submitted vulnerabilities and handle them.

In contrast, less popular and smaller website very often are actually not capable to do so. So in fact, about 25% of the submitted vulnerabilities remain entirely unhandled by the organizations to which they are targeted.

On HackerOne, in contrast, since these are company-initiated programs, we see a very quick response time. Within four and a half hours, we see the first response to submit vulnerabilities,

So when you take a first look at the data, then we see that it's actually rather spiky. So it's not

kind of discovery patterns they actually have in place. For example, are they focusing on specific programs, or are they applying the same kind of technique across very different website?

So there are lots of interesting additional results, if you have already accumulated our papers, and I encourage you to take a look at them. In total, I believe that the jury is still out about which of these two participation models, the white hat initiated model or the company-initiated model, are really giving us the best advantages.

On the first glance, it seems that the white hat initiated model really has strong benefits in terms of participation. So we see many more whites hats, many more organizations that are involved in these kind of ecosystems. But on the other hand, a lot of these participating organizations are not very well-prepared when it comes to receiving these kind of vulnerability reports, and actually then improving also the security on their website.

So there is kind of pros and cons that we can observe. One issue that's clear is we can jumpstart and further engage in the discussion what kind of contributions overall these bounty programs make to the security office websites. Our initial assessment is positive, but I think we can go into further detail during the discussion, and that brings me to the end of my talk. Thank you very much.

[APPLAUSE]

KEVIN MORIARTY: Thank you. Thank you again. Next up, we have Veronica Marotta and Alessandro Acquisti from the Carnegie Mellon University.

ALESSANDRO ACQUISTI: Thank you, and good afternoon. This is a joint work. We have Veronica, [INAUDIBLE], and myself. If some of our previous work, you will know that often, we use behavior economics to try and understand how people make decisions about the personal information.

The study represented today is actually about traditional microeconomics. And it is about understanding the allocative and welfare impact of a targeted advertising. However, there is still a behavioral angle, at least in the motivations behind our work.

In the behavior decision research, it is very well-known that how you frame a certain problem influences the way people will think about this problem, and we make decisions about that. And currently, we live not only in the age of real data, but also under a very powerful frame, the frame that personal data is the new oil And we are all going to benefit, perhaps in the equal parts, from the collection and sharing [INAUDIBLE] of our personal information.

More specifically, there are a number although frames which are quite common in the family debate over privacy. For instance, personal information is the lifeblood of the internet. So the increasingly sophisticated collection of data is necessary for us to have free services online. Or lots of privacy is the price to pay to extract the benefits of the data. Or sharing personal information is an economic win-win, which benefits equally the data [INAUDIBLE] and data subjects.

target consumers directly. They need to rely on an intermediary that facilitates the allocation's advertisements.

We assume that the intermediary itself is a profit-maximizing agent that receives a payment every time you hold the auction for the advertisement's allocation. Finally, consumers have product preferences, but they need to know which seller is selling which product. So in this sense, advertising plays an informative role to consumers.

Now, the different colors correspond to one of the different informational scenarios that we consider. Specifically, each region captures under which scenarios the consumers are better off. So we have two predominant colors here. The green regions captures all the combinations of the model parameters for which consumers are better off, when already the whole [INAUDIBLE] information is available during the targeting process.

So what's an intuition there? In their region, consumers are more heterogeneous in their product preferences. Therefore, revealing the horizontal information actually ensures the consumers see the advertisements for their products they like the most. So there is a better matching between consumers and companies.

The yellow region instead captures all the combinations of model parameters under which the consumers are better off when no information about them is revealed. So in their region, consumers tend to be more homogeneous, so brands don't matter as much, and so the targeting is not as valuable to consumers.

Now, we can construct a single graph for the intermediary's profit. Again, we have two main regions. The yellow region, again is the combination of model parameters for which the intermediary's profit now is highest when no information is revealed about the consumer.

So we said in the region, consumers tend to be more homogeneous. So what happens is that if advertiser had that information, they will tend to bid lower to show the advertisement, lower than the intermediary's profit. But if the information is not revealed, then the advertisers have to beat an expectation, so they may overbid, increasing the intermediary's profit.

The red region instead is the combination of model parameters for which the intermediary's profit is highest when the vertical information about the consumers is available. In the regions, consumers are more heterogeneous, and so revealing the vertical information during the targeting process intensifies the competition among the bidders. They may tend to bid more aggressively

So if we put together these two pictures, we see that we have situations in which the interest of these two players are actually aligned to the yellow region. But there are also situations in which they have contrasting interests. So we may think of a situation of an intermediary that may have power over the information about a consumer, and may decide to act strategically, either by revealing the wrong type of information, sea green versus red region, or revealing too much information, when instead consumers would've been better off with less information being revealed.

Now finally, we can use the simulations to understand and analyze how the allocation of the benefits among the different players changes under different scenarios. So we can construct a pie

a decent amount of the benefits in all the cases, with the vertical information one being by far the best case.

For firms instead intuitively, it's always better off to at least some of the information about the consumers, with the complete information case being in this case the best scenario. So if you want to summarize those findings, we find that consumers are generally better either when a specific type of information about them are available, or in general, when less information are available, and that there exist situations where the interest of the players, the intermediary and consumers, may be misaligned. And therefore, a selected intermediary may choose to selectively share consumer data in order to maximize its profits.

So I'll leave Alessandro to some final red.on c ru the 2e

KEVIN MORIARTY: Thank you Veronica and Alessandro. Next is Catherine Tucker from MIT to present privacy protection, personalized medicine, and genetic test.

CATHERINE TUCKER: OK. Thank you very much for having me. So I'm Catherine Tucker, and I am an economist who studies the economic effects of different types of privacy regulation using real-life data. And what I'm going to be presenting today is joint work from Amalia Miller, where we investigate how different forms of privacy protections affect consumer take-up of genetic testing.

And because I know that a lot of you are here to think about advertising and more mainstream issues, I want to make a pitch for why this is interesting before you all go to your electronics. The first reason-- so why we think it was interesting is that, first of all, this is a technology with a huge upside, as I'll get to later.

Secondly, it's also a technology where I think even the most cynical person about privacy would say there are potential privacy consequences of this data being created. Sometimes, when you're thinking about targeted advertising, it's hard to actually articulate the privacy harm, which is why we often think about health and financial examples.

But if you think about genetic data, it's not hard to come up with examples of harm. So for example, I took a 23andMe test. I will share with you, I find out rather depressingly that I got a three times normal than average chance of getting macular degeneration later in life. That means I won't be able to see too well.

Now the reason I feel comfortable announcing it in this audience is because ultimately, I have tenure at MIT. I probably have the least potential consequences of anyone in the world of releasing that kind of data because I have a job and I have health insurance. But there are potential-- and you cannot have to go far to think of potential negative consequences of that data.

And as the previous presentation on genetic privacy articulated I think very well, there also issues to do with identifiable, the fact this data is persistent, and the fact that potentially, this data has spill over to family members. So it's really quite important privacy consequences.

The other reason I think this paper setting is useful is simply because there's been a lot of experimentation about different kinds of regulation, which allows us to have more of a horse race than we usually do when trying to evaluate how well privacy protections work.

Now, I said there's an upside to this day, and I just talked about the downside to it being created, but there's a huge upside. And the upside is the promise of personalized medicine. And the typical statement made in favor of the personalized medicine is that for the average drug, based on your genetic makeup, it won't work 25% of the time.

So we can imagine if we actually had genetic data, we'd be able to identify effective

So Angelina Jolie did genetic testing. She found out that she unfortunate

Now, I realize this is not an economist audience, so what I want you to think of this is a statistical relationship that we do, where we're controlling for just about everything that you might think of going on in the background. We're controlling for the year, we're controlling for

So what is really going on? I've ruled out hospitals. I've ruled out just it being some spurious correlation to do with the state. And I think what we're going to argue is that ultimately, it makes sense when you understand how this privacy information is delivered.

Now, we found that in general, informed consent-- that is, giving people information about how their data will be use, but without giving them corresponding control-- just deters patients, both patients and hospitals from having genetic tests and offering genetic tests.

Lastly, we found that data usage policies have absolutely really little effect. And so it's either

To give you a sense of the overall totals, we see that data breaches have been, in fact, increasing over the past few years. So these claims by others that there are more breaches now than there were before do seem to be true. However, we find that they're increasing at a decreasing rate, as opposed to security incidents, privacy violations, and these phishing, and skimming attacks, which represent a much smaller proportion of the overall incident.

So we see the first takeaway from this is that data breaches really represent the majority of these events. Interestingly, security incidents seem to be increasing at an increasing rate over the past few years. Now as far as I know, there have been no changes in regulation requiring disclosure-- an increase in disclosure of security incidents. And so, conditional on the same level of reporting and of detection, what this might suggest is that firms are being attacked more now than they were before.

In regard to the insurance industry in trying to understand the risk of their insured, one way to understand that is to look at analysis by industry. So we might want to understand what kinds of industries suffer the greatest number of attacks, or pose the greatest risk. And of course, there are many ways to think of this. We could look at total number of events by industry.

But that gives us an incomplete picture. And so we might look at the incident rate, the proportion, the percentage, of firms within a given industry that suffer the greatest number of attacks. And then, we could also look at lawsuits, just as an aggregate, and a litigation rate. We could also look at cost of events. I won't go through all of these in the interest of time.

But I'll show you the-- so as a function of total incidents, the finance and insurance industry

of about 3% or 4%. What we also show here is that you'll notice the litigation rate for privacy violation is very quite high, 95%.

And I think this is really just more of an artifact of the data. I think while for the data breaches, we can understand a sample of breaches, and identify which of those have been litigated, because the breach notification laws. But for privacy violations, we don't really have that same denominator. We don't really understand the total number of violations, and therefore, the percentage of which would lead to litigation. I think in our data set, all we're really finding is we're only observing a privacy violation when a lawsuit is occurring.

Now the next question, we're going to look at some cost data. So I will couch this by saying that these are estimates of cost. They certainly don't include lots of other information. They're all firm-based. So typically, first party losses and third party losses. So all the costs that a firm would incur because of the data breach that you could imagine.

So the cost in notification, the costs of forensics, the cost of repairing any IT systems. In some cases, they represent a dollar figure loss, a financial loss. The(t)-2(i)-2(n)esrty losse-1(s)-11(e-1(s)-11)-4( )ncan

quite extraordinary, in fact. And indeed, in the information and financial insurance sectors, almost 50% of them are repeat players. I think that is quite interesting also.

reduce vulnerabilities. Veronica and Alessandro proposed an economic model for advertisers, platforms, and consumers, and concluded that the allocation of the benefits of sharing consumer information tends to benefit the platform and the advertiser.

And if I'm wrong about any of these recaps, you can tell me in just a second. . Catherine presented an evaluation of the rate of genetic testing in states with privacy laws that fall into three different general categories, and concluded that states where redisclosure restricted have the highest testing rates, and that states with informed consent decreases the rate of genetic testing.

And finally, Sasha looked at one set of data and offered conclusions about the median cost of cyber events, putting it around 200,000 and less than what other studies have found about the cost of cyber events. So to start, I just want to turn it over to Siona to offer some thoughts and start the questions.

SIONA LISTOKIN: So Kevin had asked me to talk about themes in this panel. And I would note that the title of the panel is the Economics of Privacy and Security. And I think that's about as close as we'll get to a theme. Lots of variation here.

Papers covered some of the most important or touchstone topics in privacy-- so health data, online advertising, and of course, security. I'd also point out that the 14(f)-1( )-10(g) ond ofhat t orod an(i)-2(e)

empirical work. So that is on my wish list, is to see even more empirical work. And in order to have more empirical work, sometimes we need data from the industry.

And I think this a very critical issue when it comes to genetic privacy, but also to consumer privacy. A similar issue also rises in the context of security, where actually the most interesting things might happen in the context of what we do not observe. And so you saw it in Sasha's chart.

We could only analyze the data that was detected. So what about all the security breaches that we do not observe and we know nothing about? Similar, with respect to my presentation, there is the behavior of white hats, which we can now analyze in a reasonable fashion, even though this was one of the first works doing that. But what we do not observe is the behavior of black hats.

And there, we still have a lots of work to be done in terms of investigating them, and getting maybe qualitative data, but also tying together data sets such Sasha's. This, for example, analysis that we have done to kind of be able to infer where vulnerabilities abilities have been known by the black hat community that have not been discovered by the white hats.

ALESSANDRO ACQUISTI: May I add something. Jens said something really important about long-term effects. And here is the dilemma that, as a field, economic privacy faces. In my belief, the most interesting applications of data sharing under the protection are long-term and indirect.

But generally, as economists, we can publish and do rigorous work when we have short-term and

horse race to make me think, well, perhaps there is something more generalizable we can say about the effectiveness of different privacy regimes.

DOUG SMITH: Thanks. And then the question I have for the group is actually a little bit of a follow-up on one of the things Siona pointed out, which is, you guys are looking a lot at how firms' choices are happening in this arena. So what do these papers in general suggest about what the private sector is getting right, what it's getting wrong? What can this improve on our understanding of what kind of market failures you might be most concerned about in this area? Probably start this side, I guess.

JENS GROSSKLAGS: What's the private sector getting right? I think one observation also Alessandro and I have made over the time is that we see a lot of entities, private entities, entering the market, with privacy enhancing offers, but they're not really picked up in the marketplace to a sufficient degree. And well, the good news is that we do see these offers. We see a lot of technological solutions that are eventually picked up by start-ups.

But what we see less is an adoption by the big players because of the lack of incentives. Targeted marketing or advertisement is just too enticing to give it up in exchange for more privacy-friendly practical solution. So there's a fundamental conundrum that we are presented with that is very hard to sidestep. Nevertheless, I think it's very important that we see these new offers in the marketplace, and I hope more of them are actually picked up in practice.

SIONA LISTOKIN: What are they getting right?

ALESSANDRO ACQUISTI: Well, getting back to Jens' point about offers in the marketplace, one reason for optimism is the existence of privacy enhancing technologies, PET. So almost every time I'm invited here at the FTC, I end my talk about the PET, because I really strongly believe the technology is not just the problem, it can be the solution. But obviously, [INAUDIBLE] technologies do not stop altogether the flow of data, but rather modulate the sharing the protection.

So the reason for [INAUDIBLE] is that private sector firms can actually-- this may be wishful thinking, but may be proactive in deploying PETs, anticipating otherwise regulatory intervention so that they can still do much of what they're doing now, but in a more privacy-preserving manner. Now truth to be told, some of these technologies are still in their infancy. For instance, homomorphic encryption is very promising, but we still don't know how efficient and practical it will be. But the promising is enough for the moment, and I do believe that in the space of privacy, we can actually have the cake and eat it too, because of these technologies.

SASHA ROMANOSKY: In terms of what are firms getting right, god, that's such a good question. And I wish I had a better answer than the one I'm about to give. So I think what we can rely on is that firms will operate based on incentives. And of course, the goal, then, is to tweak the incentives such that they become aligned for all of the players.

Right? So that's not new. And what that means is that, look, if privacy really is a big deal, then consumers should really act like it's a big deal. And if and only until they do, will firms have

incentive to take it seriously. So I guess I would say that consumers should take it seriously, and act like it, then firms will take it seriously.

Now, if there are market failures for which consumers can't impose any kind of effect on the firm, then that's where regulation or policy or FTC actions could come into play. Go ahead.

CATHERINE TUCKER: No, I just want to build on that. Because I think what I often see in the

So from their perspective, it was worth their while. And certainly, one of the main selling points is that it provides a different perspective in addition to running software office security tools, having internal security researchers, in the sense that white hat researchers have perhaps somewhat more of a view like a black hat in an organization. They are more creative. They poke holes in places where other the security researchers would not look.

And this is certainly a big selling point to inch the security of your website even a couple of steps further. Also, I think that's a lot of criticism about bet ratios between the reports and the data that is actually then useful. And I think when you actually look very closely at this, a lot to do with the matter of duplicate reports.

And well, I mean, this is actually white hat researchers doing their job. If the reports have not yet been disclosed, then well, they have will report oftentimes the same kind of security weaknesses to the particular entity, and well, taking this into account, then actually, the error rate is not that high.

My last point here is that here, actually, the involvement of bug bounty platforms can really have a positive impact, because they can introduce measures such as reputation mechanisms, coordinate [INAUDIBLE], and so on that actually then also instill some part of competition between the white hat community participants, so that they are more inclined to actually provide high quality data to the participating companies.