Watch Script

Facilitator: Welcome and thank you for standing by. At this time all participants will be on listen only until the question and answer session in today's conference that which time you may press star 1 to ask a question. Today's conference is being recorded and if you object you make disconnect at this time. Thank you.

Keith Marzullo: So, welcome everyone to the 6th in our new series of our Watch lectures. I am very pleased to be able to introduce Stefan Savage from the University of California-San Diego. Stefan has done some just remarkable work in a couple of areas and as a byproduct occasionally has created some interesting discussions inside NSF about who should buy cars and things like that. He graduated from the University of California—Sorry, University of Washington got his PhD there and uh-- joined San Diego Jacob's School of Computer Science and Engineering in January 2001. And since then he has just been uh—done a whole lot of interesting things. He is not actually going to talk about any of those things today as far as I know but rather a topic that I think is very provocative and particularly appropriate for this series which is why the hard problem of computer security needs the soft sciences—Stefan.

Stefan Savage: Thank you.

Stefan Savage: Thank you very much. [Pause] Thank you. So first off, I want to be clear that this is—the work that I describe today will be work that I was present for but was actually done by the amazing cast of characters listed here-- graduate students staff and faculty—have been working together for a bit over a decade. To give people a little bit of context about myself first because there are a number of faces in the audience that I don't recognize. So, I assume not just my lack of memory and that I haven't met you before. Let me give you some background about who I am and how it relates to the talk I am going to give today. So one—I co-direct a-an effort called Collaborative Center for Internet Epidemiology and Defenses which is a mouthful-- which is one of the NSF cyber trust centers-just winding down. And we were created in response to the worm threat you may remember-turn of the century we had the first big worm outbreaks on the internet, code red and slammer and so forth. And so to address that, actually our group started to get together and eventually funded by NSF in 2004 and through that actually we were able to build a fairly sizable activity. We get support also now from ONR for aspects of this work and a lot of industrial support primary in kind in terms of data and so forth but it allows us to do quite a bit. And so the context here is when we got into this-we had a number of goals for our work. One is what we call internet epidemiology, we wanted to be able measure and understand the kinds of attacks that were confronting internet users. And we wanted to come up with a quantitative methodology that would allow us to describe the growth and viruses and worms and so forth. Then based on that understanding, we wanted to build automated defenses because now we wanted to protect—understand and protect—that's what we thought computer security was about. In writing the proposal, there was—there is this section about broader impact. And so in-I will be honest and pure good service for our

broader impact- we also said that we wanted to look at a bunch of other things—the economic, social and legal issues and we had our lawyer on staff and so we figured she would write some legal stuff and that would be a checkbox-I impress. And part of the reason why I'm here today is because I think a number of years through the effort, we decided that in fact that thing was a glorified item for us was in fact by far the most important aspect of the work and that this other stuff was perhaps not as important as we thought it was—the technical aspect going in. [Pause] So, just to give you some context, after a few years doing this, we had a lot of successes by traditional metrics of success-right, so we had a lot of papers and journals and conferences. We had built big systems just a couple of examples—So, UCSD runs this thing called the network telescope, 1% of all writeable addresses actually come through there. So anything that happens on large scale using random addresses on the internet like worm spreading or denial of service attacks, we're able to observe a lot of them pretty much all of the measurements-worm growth that anyone did either used either the state or the methodology. We had for a long time the largest active honey farm I think that was at least public, there are 250,000 live virtual hosts-so things that were out there to get infected. So, we could find out what was new. We built a system that could recognize worm signature in under a millisecond that got spent out to Cisco it's in the catalyst series devices. We did a lot—by all these metrics we were a successful—we were a successful enterprise-. [Pause/break] But right [break in audio, then laughter heard in background]—we reflected on our time doing this and it was not the case that things were getting better. Alright? And so we did not in fact stop worms from spreading and we didn't stop malware and we didn't stop e crimes. And it wasn't that we were lousy—I mean we might have been—that might have been better people but it wasn't just that we were lousy. No one else was accomplishing this goal either, right? In fact if you were to ask the late person on the street are you more secure than you were 5 years ago...in the reprising previous presidential debate they would say "no of course not, I've been reading about all this terrible stuff that's happened." And I think anyone giving you an unvarnished independent opinion would say, "yes, that the situation has gotten worse, not better." And so, the mistakes—the conclusion that we came to which is what I am really here to talk to you about today is that the key mistake here that we made and that in fact most of the security community both in industry and in research-I think makes is looking at this as a purely technical problem. That's there are a set of holes and dikes that if only that they could be plugged up, that we would have a solution. [Break in audio]Sorry-- So this—this is somewhat unfair but I think does characterize some aspect of the state of the practice today which is that we find new problems and then we try to beat them down as quickly as we can and we eventually lose. Alright? Its just—Its just like the game of Wack a Mole. And why you might ask is it this way? Why are we stuck in this kind of game? And I think in part is because of the way we have set the game up for ourselves. So, I am going to talk about a number of A-symmetries that exist in the way the computer security game--where game is meant in the more serious sense of the word game operates. So the very first one that I think most people are familiar with is the a-symmetry of initiative. So as defenders we are fundamentally reactive. We find out about a new attack or a new class

of attack or a new kind of attack. Someone invents a new kind of a fraud, we have to react. Alright? By contrast, attackers can attack precisely when they want to. And this manifests in a number of significant advantages for attackers. So the simplest ones of these is advantages is to describe one that goes to malware. So, if you were to go to an anti virus company and ask them what percentage of malware do you detect? They will tell you 99 point whatever percent and there is a measure in under which that is true but from the standpoint of does it detect the new malware that someone has released the answer is absolutely not. The answer for anyone or for any kind of malware written by someone who doesn't totally suck. Alright? [laughter in the background] The answer is zero percent is detected. And the reason why zero percent is detected is definitional—it is that they all have access to consumer over the counter software to the anti-virus product and there are services that they use and catered to their criminal underground that will test their malware to see if it gets detected or not. And they do not release it until it is not detected. And so by definition they only choose to release when it is not detected. The detection rate has to be zero just because of the way we set up the game. [Break in audio] Another a-symmetry we have is on the innovation side. New defensives tend to be much more expensive than new attacks some of that is technical it tends to be a lot harder to build recognizers than to build obfuscators but some of it as well as economics. So when you are building a company in the space, you have a lot of ...cost in a particular business model. So, if you are selling a particular kind of product and you can see this from people who in response to big worm out breaks, a lot of people built software to try to defend against these things. Then Microsoft added the firewall to Windows in EXPSP2 in 2002 and in a very short period of time, in a matter of months, the bad guys figured out that really quickly that well we could probably get around this but it's much easier just to shift our base of operations to the web and we will come up with exploits that will attack your browser and we will just have you come to them instead of sending them to you. And if you were to talk to people in the companies who were building this other kind of software, they would say, that's not our problem. Right? That's web security which apparently is a new kind of security that had been invented right then and there. So about two years in which where start ups get created in web security before they get bought and absorbed and that you can expect the product that you buy would actually do something about this class of threat. So they can respond very quickly. And it's harder for us institutionally to change- to change this way for-for lots of reasons. [Break in audio] We have an a-symmetry incentives so generally speaking there is low risk to attack some of this is kind of the old thought that because there is a lot of anonymity that there's not much deterrence. But there's another aspect to this as well which is that security is generally speaking not a key competitive feature. Alright? This is even true of security products. And the reason for this is the last one which is asymmetry and evaluation. [Break in audio] We do not have any meaningful way to talk about the security that a product offers. Alright? How many quatloos of security does Symantec offer versus McAfee? Alright? Which one offers more security? We just don't know. It's not to say they don't work really hard on improving the security of their product But it is not built into the way this works that we can tell. As a result, you know, it's a dirty secret of the security

industry. They compete on every aspect of the product that they sell, except that the security they offer, because they can't compete on security. It's not possible to compete on security . So the end result is that you--You compete on performance and marketing and things like that. On the flip side for the attackers, particularly the economically motivated attackers is very easy to measure how well you are doing. Do you make more money? They use the exactly same criteria that uses Amazon for redesigning their web page. Will we swing at this a little bit?...Did we make more money? And that's a good UI feature. They do the same thing. I did this to my malware, I made more money. Darwinian. It just gets better – It just gets better and better because they have a very easy way to evaluate. [Break in the audio] So if we were to go back and reflect given this, we spend depending on whose made up numbers you believe somewhere in the $50 to $75 billion a year worldwide on IT security primarily spent on the goal of securing the end host. Right? Making sure that my laptop and your server are not compromised in some way that's everything from A.V and firewalls to security reviews to programming languages. You know all this kind of stuff. The down side which our focus being there is this is the probably single most expensive front to secure because it's all of the 2 billion PCs that exist. And they are being administered by individuals who have incentives that have nothing to do with your definition of security as an outside third party. So, you are counting on 2 billion people that are administered by like my dad to actually stay secure. And so we can turn this question around and say well, how valuable are these hosts that we're spending all this money to secure to the bad guys? Let's suppose they want to buy some hosts. Well, how would you know that? Well it turns out that there is a great market at any given time there are about 25 organizations that are selling compromised hosts and what's called the paper install market. This is one now out of business called Installed For Sale. And they will list prices here for hosts in different countries. These prices are somewhat out of date but ballpark about $100 per U.S. host and now sounds about-$5 to $8 for Asian hosts. And oh to be absolutely clear that is per thousand hosts. Per-thousand hosts. Alright? Everything is in the underground is in units of a thousand. So the value of a host is in fact at most a dime to them. And probably under a penny for a large fraction of them. So now that's not to say that aren't there's not individual hosts that have enormous value but the huge disparity in how much we spend on generally every host versus how much it's worth suggest that in fact there's something going on here we are not fully understanding. [Break in audio] Just to be clear this goes—and this is true for everything. So, these are prices per thousands like hotmail accounts or gmail accounts. We in working with a mail provider, two weeks ago we bought [break in audio] I think it was 35 million accounts for $350. So it's a total free for all out there, and this stuff is not worth very much. And these weren't even like new accounts. These were like people with accounts. [break in audio] Alright. So what should we do? Alright? If people know these things are bad and so one refrain that I've often heard particularly in DC is you know what we need? You know what we really need? We need science. We need [break in audio]. Time to apply some sciences. So we know that science is good and it can fix everything because this is you know true justice and science is how we—How we solve these problems. And I don't disagree with this

actually. Alright? But then the question is what kind of science? Alright? What do we mean? Science is part of the reason this is funny is because everyone says let's use science and they don't know what the hell they mean. [laughter in the background] Alright. So let's reflect on this. So one version, and I want to be absolutely clear—That I think all of these are actually good approaches. I just think that the one I'm going to talk about today has gotten short drifted. So one approach is, let's look at this like math. Alright? That there is--Think about security as having an axiomatic basis. There's some things we're going to define a policy. And once we define a policy, once we have an axiomatic basis for how computers operate then we'll divide principles and implement them and we will be able to prove that the system is secure. Alright? And there are great examples of this work. Among them are cryptography kind of in this realm, type theory is a huge success story in this realm and the beauty of this is once you actually define the set of policies you want to enforce and push it through your formal system, you actually subject to those particular things you care about, you have some kind of strong guarantee. [Break in audio/indiscernible phrase] Another approach is—Alright let's think about this like a physical science. Alright? It's not like this is axiomatic but there is some kind of platonic truth. There's some truth out there and we are going to discover it through experimental measurement and then we are going to generalize and this will tell us how well this particular property holds. I think a fairly good example of this is biometrics. Alright. You go in and you say we believe there's this distinguishing characteristic and we if measure well enough how strong that distinguishing characteristic is across a certain set of populations and so forth and then that tells us how well we can expect to how well work in a that's going to work in a statistical sense and feel comfortable with it. The third one which is what I'm going to talk about is like a social science. Alright? And that is it is not the case these are platonic or axiomatic truths. They are contextual truths this is a socially constructed system that we live in. We can discover those conceptual truths experimentally through both controlled and experiment and through field work. And some of them are in fact generalizable in a way you use them to make predictions or to make interventions so to improve your life. Alright. So let's take a step back and try to justify why it is that I think the social science aspect is one that's I think is particularly so fruitful. Whenever you corner a computer security researcher and ask them what is security is about? Well it's about how do you provide some kind of functionality in the presence of an adversary. So implicitly we have an adversary and we are in conflict with the adversary. Alright? Absent conflict with an adversary, we don't care about security. We just don't do it and then we are fine. So for us to care, we need complex and we need an adversary. So kind of to paraphrase CHARL son HEPTON, it's people. Security is people. Alright? Because all of the people who are involved, we tend to think as the adversary as this—The computer security community inheriting from crypto tends to think about these things very informally. We have—You know Alice and Bob and Eve and they have arbitrary or abstractly defined powers. But in fact our adversaries are real people and they have real motivation and capabilities. They have organizations that show up the ways they operate and what context that they do. Our victims are victims in part because they have particular behaviors and limitations. Are defenders are limited on

what they can do because they haven't set their structures that guide what they do. And so while this particular situation, computer security has a conflict that is mediated by computers, it goes through computers. It is not driven by computers. It's social and political and economic forces. And so in the end, why would we ever believe that we're going to solve this entirely in the computer domain? Alright, in fact, the hypothesis is that we're going to advance is that you are going to do much much better in addressing security problems if you actually understand what these forces are. I'm going to give some examples today of how is that so. Just so you know--So to make some analogies, it would be so bizarre if computing was the one place in which you can solve in what really is a problem of human conflict entirely in a technical way. We don't expect that some technology is going to solve our—Our need to have wars, our criminal justice problems. I mean the analogy I made at times is—Yeah, let's solve crime by making better highways because you know criminals drive on highways and therefore. You know it is clear that technology is a critical piece but it's not going to solve a problem anymore than building a good fence is going to solve the drug problem. Alright, there are some deeper issues here. [Break in audio] Alright. So the remainder of the talk--I'm going to focus on one particular area that we know a fair amount about which is the spam echo system. Although in the very end I will talk about how it's applied in a whole lot of cases. I'm going to first give you some background for how the under groundwork and what the economic and business structure for the spam you get in your e-mail. And then talk about how one goes about reasoning and what one might do about this problem.[Break in audio] Sorry. So, just to give you guys a little bit of background. The worm thing was actually a big turning point in computer security. So, in the 21st century we had three big technical changes that then in gents that allowed these other economic changes to take place. So one is we get efficient large scale compromise things like worms and viruses, becomes possible for you to takeover large numbers of hosts in short period of time. We get centralized control. We pull back access to all those hosts to a central point so now you can control them as a single entity, as like the cloud as it were. But the kinds of dark cloud. And then the probably the most important version is we get applications that are profit driven. Alright? So we get applications that either try to monetize the commodity resources on the PC like their band width or their addresses and so forth. Or unique credentials like your bank account. [Break in audio]The big change comes about—This actually, I would argue the single biggest moments in as these pragmatic computer security happened in 2004 when these guys who are running this so big in my doom- Which were basically these worms. They didn't do anything. Just they just took over machines. Cut a deal with a spammer. And the problem was at this time that we were successful in keeping spammers from delivering mail by creating blacklists of the hosts that they were using to send mail. And having done that they now needed a way to wander the origin from which they were sending this mail. They cut a deal with these guys. They say when you take over the host, how about you put in a little proxy so I can bounce mail through it? And so it allowed them to wander that point of origin. This changes absolutely everything because all of a sudden-- Now these guys are getting paid for doing what before just a joy riding activity and they can get paid more.

The more hosts they takeover and the longer period which is what I have been told are kind of what I call is a virtuous economic cycle which I find is very ironic considering the context. In a very very short period of time this creates a commodity market for compromised hosts. By commodity market, I mean like what I showed before in lots of thousands. You can pay differential pricing based on where they are what that mail hosts or whatever. I want high up time hosts, high band width host their value added tiers. You know let a thousand flowers boom you get a very dynamic market. And so what happens is you get innovation in the substrate which is these compromised hosts controlled from a central entity and this really creates a platform economy. And so from the stand point of online economically motivated criminals, it is totally reasonable from them to think about just buying hundreds of thousands of hosts as a platform for them to base some kind of fraudulent activity on. And then all of these basically, every bad thing you know about on the internet, whether its phishing or spam or info*- These are all vertical market applications that are built on top of the same platform. [Break in audio] I'm not going to go through all this. We spent a lot of time studying this underground market and the ones in green are the ones are the ones we've have actual projects studying but basically they fall into these aspects of this category fall into, one acquiring elicit goods whether they are your bank account or computer or what have you. A market for trading with others so you can expand. You can do more and you can lower costs by having a wider set of people that you are collaborating with. Scams which in a normal economy would be called entrepreneurial activity because this way you add capital to these resources to create new money. And then liquidation which is the trickiest part where you actually get cash out in the end. And so in some cases it's advertising basis. So there's a lot of stuff going on here. We try to break things down into categories. I think we have about a 120 different categories of different goods and services. But at the end of the day, there are two value creation strategies that exist. Alright? And there are probably a couple of contrary examples so anybody you would think of them almost everything fall either into advising—That is monitorized through advertising or it is monitorized through theft. And it is a spectrum. So like clearly advertising is good spam. People selling you Viagra and clearly they are after your credit card... same anti-viruses somewhere in between. You actively participate for you being defrauded and so forth. This inturn, all new capital comes from in turn the theme – This in turn funds everything I would call infrastructure. SO everything from botnets to banking, Trojans to underground virtual private networks to hypen police. There's a huge activity that is funded through that capital because that—The market for these services are the people who are doing the [indiscernible word].We're going to focus on the left-hand side in particular on the advertising and spam. [Break in the audio] So first I'm going to give you a picture of how a modern spam campaign works. This is a real life example. You've seen this one before, kind of--So this is an example from a couple years ago. Real life example. The grum botnet deposits this e-mail in our mailbox advertising online pharmaceuticals. It's not actually important that this is the e-mail. We call this the advertising phase. It can also be a search. It can be Twitter. It can be Facebook. The important part is that it gets a link out in front of a potential customer and tries to coerce them to click on it.

[Break in audio] So having clicked—Then a whole bunch of other things need to happen. You need to have that click take them to a site they can actually purchase from. We call this click support and there are a lot of moving pieces. So first, they need to register for this domain that they have asked you to click on. And in this it was a registrar in Russia who was mass registering the domains of questionable value. They need a name server to actually host this domain and in this case it was a name server in China. Then they need a website to host the content. In this case it was a compromised host in Brazil that did not actually host the content but who was a proxy that reached back to an affiliate network in Moscow run by a jail, child molester of all things. And now you can get to the site and purchase the goods but there's a third thing. They need to be able to take your money and deliver the goods. If you think they do that. So we call this the fulfillment phase. And so what's happened is this particular affiliate firm has have cut a deal with a Ajuba Jani bank in Baku to accept Visa payments on behalf of the customers of this activity. And they have cut a deal actually through a middle man but the picture gets too complex with a set of factories in India who then drop shipped goods back to the person. The important part of this picture is every single one of these lines has to work for them to make money on that spam message. If anyone of those lines were intercepted, you could not make money. And so the question that I'll ask you to be thinking about in the back of your head is what line should we cut? Where should we put our money to try to stop this problem? Right now where we spend most of our money is trying to keep from getting the spam to begin with. Alright? And I will tell you again it's the most expensive place to put your money because it's really easy to send a lot of spam and very easy for them to eventually get past the filters because they get accounts on the same site--And send mail to themselves until they modify until they can get through the filters. [Break in audio] So let's talk a little bit about the actors. So remember we have these three phases I talked about. The advertising phase is almost uniquely done by the spammer. The spammer is not selling goods. They are an independent contractor who is an advertiser. Alright? They are an affiliate of what's called an affiliate program which I'll go into more detail. They also handle sometimes the quick support phase of setting up the domains and so forth. The realization phase is almost entirely handled by an organization called an affiliate program and this is an evolution of this business structure. If you were to go back and look at how spam worked five to ten years ago—It would--The one person would do everything they would be swooped to nuts. We now have division of labor, ends up being much more effective. Sometimes to a third party. Alright? So let's talk briefly about affiliate programs. The way this work is they are kind of like a franchise business for advertisers. They hire advisers as independent contractors. They provide contents. The web pages and so forth. The back end engine for processing orders. They handle payment services. They have a relationship with payment processers and banks. They handle fulfillments that is getting whoever delivers the goods and services. And they provide customer service which actually you know they provide quite good customer service and that actually ends up being somewhat important. They are paid on a commission basis. So for pharmaceutical it is between say 35 to 50% of the net that comes in from a customer order. [Break in audio] So here is an

example of affiliate program. This is one called RX promotion it is run by Pavel Vrublesky who is now actually in jail in Moscow. The program has since been shut down. Among the things that they advertise –pharmaceutical programs. Programs for almost anything that you can imagine but form uh as your experience goes probably looking in your inbox is a pretty big one for spam. And so they will advertising a number of things. One of them--Whoops [Indiscernible phrases/ break in audio]. They'll advertise that they have different commission structures so they'll compete on the price of drugs. They have parties for their people. They have various kinds of incentives for selling the most or selling things in a particular categories and then they offer cash out through independence online currencies. They also let you run your own shop. They have a whole bunch of different templates these all kind of look like they are different—There will be whole websites around this theme but they are in fact all just geared toward different market segments. So, if you are selling Viagra to the elderly versus to the young. If you are selling scheduled two opiates, you know you will have different marketing content that goes along with that. Alright. So there's a bunch of different questions one might ask about this. One set of questions that garner—How—What does this business market looks like? How good is it? And how much money are we talking about? And the other is given this how should we intervene? We have done a lot of work. I'm only going to be able to talk about snippets of it. I am going to give you some flavor for the kinds of work that we have been doing this base to understand this problem from an economic context and from the stand point of actually understanding who the actors are and how they manage. So quick aside—Because Doug [indiscernible last name] gave a talk last time and we end up—We always get called out for some reason. And I think part of it is because we are frequently developing new methods and in particular new methods that involve you know direct engagement. We purchase from criminals and so forth. So, I wanted to be actually very clear first thing in light of Doug's talk that we pay more attention to this than almost anybody. In the list of oversight, we have a human subject review. We have our own lawyer who in fact is the lawyer who co-wrote Doug's report that we pay. We have then two independent lawyers who we check her opinion on by chance of research. We have general counsel who we are in very good basis with and we have sign off on the UC system wide office of research compliance. We are the most overseen research group that I think you have come through here in computer science in a long time. So you know I think perhaps we could do more on legal and ethics but we do as much as I think it could be humanly expected and still get work done. Alright so given that, let's get back to talking about spam. The first issue with all this is demand. Alright? So the stuff would only--Only happen because someone wants it. Someone has to actually click on this stuff and pay for it otherwise you can't make any money doing it. So there's latent demand out there. I've yet to meet a person personally who says, "oh yeah, I buy my drugs online from Russian pharmacies." And so we might ask ourselves where is this coming from? And so there's a case study actually appears in 2011 looking at this. And here we are particularly aided. So, we are very opportunistic group. We take data from anyone who will give us data and then try to figure out how much we can figure out about the world from that. And in this particular case, one of the

peculiar things about this pharmaceutical program which is one of the largest is called Evo Pharmacy and is used to be called Bulker Dot Bis is that to save on hosting costs, they save--They host the images on the site on compromised machines so they don't have to pay for the band width and so the very particular structure and we hunted down a bunch of these compromised machines and contacted their administrators that your machine has been infected. It's hosting pictures of Viagra pills. Would you be willing to share the logs with us? And we found a taker and this allowed us to basically look at every visit to one of these sites that anyone had made for I think about a period of five days. So when someone visits one of these pages, it pulls the main part of the page from an affiliate program but then pulls all the images from one of these five compromised servers that which we now have logged. When they then go and pick a product, there's a separate set of images on the product page so that allows us to see what product page they had gone to. When they then select a particular product, they are yet again some distinct images that appear on the selected product page. And there is yet another image that gets picked when they click on check outs. So, by building a basically parser that looks at these images and the data that we received was the source IP address was anonymized and replaced with a city. But that's all we pretty much have to work with. But we can tell here's what some person in this city looked at and then chose to put in their cart and said I want to check out. So, we don't know, maybe they never provided a credit card but they showed a fair amount of intent. So we're getting intent to the point 0of set check out. So we get 752,000 distinct IP's who are visiting. And they make about 3100 distinct additions to their shopping cart that they then want to check out from. And so what you find, everyone on the planet is visiting this site. But then if you restrict this to who is in fact trying to purchase something, it's much smaller and it is much more concentrated. So in fact, you end up finding 75% of all customers are in the U.S-- 91% are from western countries. This is an activity that is funded with western money. Alright? So it is a negligible amount of money that is not coming from the west. [Break in audio] So, we also know what is getting purchased. Alright? And so this, there is some—There is some interesting features. So, it's not too surprising [laughter in the background] that a whole lot is erectile dysfunction drugs. And then there are also--We've placed in a recreational category things that are abused. Alright? So things like opiates and stimulants and so forth. That is the lion share of what is out there. But there's a long tail, alright? There is about a third that are for things that are either acute or chronic conditions. That are like real conditions that people are buying stuff for. These include, I mean—We find AID's medication. We find cancer medication. We find diabetes medication, you know all over this mess. So then—Oh yeah, so--We've actually done this since for a number of pharmacies. And there are three clusters that all of us come out which is easy: opiate stimulants and then chronic Meds. We can then break this down by country of origin. Alright? And so let's just look at western countries, alright? Not from the United States until what you find is 92% fall in the recreational category and 8% are these kind of chronic Meds. If you look at US orders this is where all those other things come from. Alright? So, [break in audo] people in the United States are four times more likely to buy normal medications than in these other countries which

perhaps this is not the reason but actually subsidized the cost of drugs for their visits. And so I think one of the things that you—That you can find out with this kind of research are these structural aspects of what's going in the—On the victim side who are participating. Alright. So let's talk about business models. For the spammers, this is the same—This is direct mail. It's the same business model. So, as long as the advertising cost is less than the conversion rate times the marginal revenue, you make money. Same business model as LL Bean. Advertising costs we know--We can either figure out what it costs retailers it's about $60 per million sent out on the underground. It's probably much less so for the guys who do this on math because they just-- It's the cost to run a botnet. And so it's--You are what your skilled labor costs in the country in which you operate which is low so like a got. Marginal revenue, average sales is 100 to $200. Commission in neighborhoods is 40%. Conversion is tricky and so we did a study in 2008 doing an active measurement in this space where we found for one campaign they got one successful conversion for 3 million messages delivered. I think it's conservative because of some methodological problems we had in this study, but even at that rate, we figured out they could be pulling in about $3 million a year in revenue so it's--For one—For that campaign. Now the business model--The affiliate program is a little bit different. So first, the affiliate program model is absolutely brilliant because it transfers all of the risk both in advertising and in innovation to the advertisers. They are the ones who need to figure out how to deliver this advertising. You don't have to. If they don't do a good job, that's fine they don't make any money. In fact a lot of the affiliates don't make any money. In fact so we have data from inside one of these [indiscernible term], 90% of the revenue comes from 15% of the affiliates—So, its heavy tail like almost everything else. Cost structure, the commissions are the biggest single costs. Supply is about 15% of revenue. Another 15% goes for payment processing, which is incredibly expensive if you think about it. Normal discount rate on this should be like I don't know 200 basis points. Growth margins are not that big, alright? If you are venture capitalist you would not want to invest in something with growth margins between 10 and 20%. On the other hand these are largely operating out of countries in which there's no meaningful alternative with their skill set to engage in the western market and so this is actually--May be quite a bit better than what's available for them. Alright, so how much total revenue? I'm going to go through this really quick because I want to get to the end. Remember I said they have customer service? Alright, so we've placed quite a few orders now. We have placed over 500 orders from online sales. Not using federal money. Let me be clear, this is one of the many places that industrial partners come in handy. So one interesting thing if you get—Is you get these order numbers—Your order number blah blah blah blah blah. And we noticed when we would make multiple orders from different versions of the same program; we'd get these order numbers. So here is like 482065. Then it's 483939. Then at point--They are all going up. Alright? And so you can—We--You have this hypothesis—Word scientists, we call it hypothesis. It is a quick sequential update hypothesis that they just have a global variable. They increment every time someone makes a sale. [break in audio] If this were true, then you can make a purchase at time T--Figure out that the order number was say 4200 and then infer

[break in audio] having made a second purchase later, that order 224 through the power of the traction that--That in fact there had been 23 other orders that are not your own that were placed. And so in fact we did a whole bunch. You got basically two thirds of all the pharmaceutical programs and most of the capital software programs who had this characteristic. We did a huge amount of work validating this is true. So you know just instead of going through it all, it's true [laughter in the background]. So from this, you can actually get revenue estimates. Alright? Because you say alright let's look at various estimates of what the marginal revenue for sale is. We know how many sales are going on. And you get these numbers that range from a few hundred thousand dollars per month to a couple hundred million dollars per month. And in two of the cases we actually have ground truth data. And that is because a lot these-- These organizations have been escaped the Wiki leaks means. And so there is a huge amount of data that they have--They attack each other and leak each other's data. So in some cases we have like their financial statements and it's basically within a small constant factor of what the prediction is. So in the end, this is like a hundred million dollar business maybe two but not more than that. So, how big that is depends on if you-- Where you are coming from. Alright, so in our remaining time though I want to focus on what we should do. So, we've done work before looking at catches which is fascinating. It turns out that solving those little catchers or those little obfuscated puzzles, you need to solve online that in fact that has engaged a market of Third World labor. That's how those all get solved. We were able to establish what the market demographics were. Who those people were. How much they were making. What countries they work in and so forth and we've done stuff with tape down to the registrar, but I want to talk about our most recent work—Which is what we call a quick trajectory effort and the idea was that--Remember all those lines that went into monetizing this activity? Alright, so can we go and look at the full value change-- For every one of these messages and see where the bottlenecks. And by bottleneck, I mean, if we were to intervene in one of those lines, there's two kinds of things we were looking at which is how many resources at that tier would we need to eliminate to have the most impact on profitability. Are there thousands of registrars we need to take down? Or just two? And then the other question is what is the switching cost? So having taken those resources away from them by intervening, either in the technical or in the extra technical means--Would it be easier for them just to obtain substitutes? So, are there lots of alternatives and what's the cost to switch to an alternative? And so when we do this we are a hardcore [indiscernible term—impair-rsis]. So, we have a huge cluster that basically crawls off business. And then tries to purchase from it. And so I'll walk through that. We did this for three domains which dominate which aside from pornography, which we decided for institutional reasons not to tackle--Dominate the spam market. So we have a lot of different seasoned spam which we get from commercial partners. We also have a honey farm that through a project that we do on ONR that allows us to run all those botnets and look at all the spam they are sending. From all of these we extract the URLs that are embedded. And then we have a set of crawlers. One that comprehensively crawls all the name server infrastructure and this deals with all kinds of—Kind of low level issues like fast flux. Basically, get--Figure out

all of the servers that are supporting their ability to reach this site. And then we have a web crawler that actually goes to every single one of these pages, renders the page. We save all this. We have a 20 terabyte database that then holds every single thing that happens. We then take this data and cluster it. So, if the pages are structurally similar or if they look the same, then we'll cluster them together so it's all of the different pages that are—That have exactly the same brand and storefront get put in the same bin. And then we have a technique called content tagging. Where we take those and associate them directly with the organization with the business entity that is sponsoring them. And I would like to tell you that we did this through some very advanced techniques. And I can't-- What I will tell you is there is a huge amount of manual work in some cases by actually getting stuff from within these programs that they provide to their affiliates and think that these all are the options and in some cases by going through by hand. And other cases going to partners. We now actually have a machine learning approach to doing this but this was like literary a week of solid like 12 hours a day work. But at the end, and just to reiterate why this is the case—So here is a whole bunch of different structures you might see. These are all the same organizations. Alright? It's important for you when you are doing these kinds analysis to understand alright—At-- What organization are you dealing with? Not just what veneer they have put out. So at the end, we may not have covered every single affiliate program but we covered every one that matters. Alright? So the amount of spam e-mail that we were able to accurately characterize was very small. [break in audio] Alright. So then for each of these programs, we do selective purchasing. Alright? And that is we make purchases of goods from instances—Multiple instances in each of the programs. And so you might ask why would you do that purchasing you know-- Aside from being kind of salacious and sexy and so. What do you actually get out of it from a scientific stand point? And you actually try to get a huge amount. So one, is you get a lot of information about payments. So, we actually cut a special deal with a payment—With a credit card issuer. So we have our own credit cards that get issued to--We get a single credit card issued for each transaction we make and we get transactional data for the entirety of the purchase. And that allows us to tell what bank is-- The most important is a lot of information we get but the most important things of information we get is--What is a merchant bank that this organization is using in order to receive their payments. On the fulfillment side, we find out, first of all are you receiving anything? Which turns out I mean—I think—Are you in a fraud game or are you in a business? And then you—There's a bunch of issues about, you know-- Where to ship from and sometimes where the order is and so forth. Quick aside, this is—It turns out to be far harder to do than we ever thought. There are tons of operational issues in placing this kind of purchasing. You end up needing to escape their credit card fraud detentions. You need to have your IP address co-located with the address of your credit card. You can't have--You have to have e-mail addresses that are not from like public -- You can't have a hotmail e-mail address. It needs to be from a real place. You need distinct credit cards. There are cases where you need to have currencies that are not—That are basically off shore online currencies. In one case we had to get a relative to go into Russia to like get instances of this currency. Far

harder than we ever thought. And then as you might imagine when you go into your institution and explain that you want to make purchases from criminals. The very first question that I get is: Why can't they take a PO like everyone else? I swear to God--The first reaction that I got. And so—So we spent about a year doing trust buildings to get to the point--We are on a great basis at this point with everyone in our administration. But originally getting them to understand that we're going to take these cash equivalent currencies, give them to unknown parties and that they might be stolen and we couldn't account for that--That was a tough pill to swallow. Alright? And we then have a lot of issues. So then we have-- Students love this aspect of the job. There is Carolyn [indiscernible last name] who is the leader in the project and you can't—and that's-- [indiscernible first name] Chang is our director of our purchasing. And you start—And you send out and you do get stuff. It will be drop shift from pharmaceuticals as far as places like from India. For herbal supplements it's largely shipped from the United States because the regulatory regime such as that it is not as much concern about shipping from within the United States. And this actually came from a little post office right next to UMass Amherst. Replica stuff comes from China primarily and the counterfeit stuff--Counter software just you get online. So quick aside, this is not clearly a fraud game. So, we've made over 500 orders. And we also have done fake antivirus. And in all but one case we got a shipment that's not to it always made it through. So about a few percent that actually get picked up by customs. And I'll say that the FDA were very pleased because they were getting a few percent. There is basically no fraud loses. Alright? Our credit card—So we have a separate credit card for every order and we track what happens overtime. People are not defrauding our credit card. There's a significant reorder business. I can tell you more than 10 percent of the business is reorders. And for scheduled two drugs in particular it's much higher because they are seeking. So if they find something it works, they are going to stick with it. These people generally believe if you listen to them talk, they are selling products of reasonable quality. I cannot talk about that categorically because we have not bought all these different products. There is a lot of legal issues about what we are allowed to buy and not allowed to buy, but I will tell you for one drug that we bought—And then we run a hundred mass bet, here is the difference between the controlled and the sample. [break in audio] What's that? So, it's not clear that this is so different. Although, we did not test categorically for like--This is something for active ingredient we didn't look for additives and so forth—It's not our—It's not our specialty. Alright. So for three months we basically looked at all spam—All major spam campaigns on the internet. We crawled basically 98% of the domains of the URLs and made multiple purchases from all the programs. At this point we've now made probably 30, 40 purchases from every program. And then we said all right let's consider interventions at each of these levels. Let's go and snip at different places we can do it at. Let's go at the registrar level, at the hosting level, at the payment level or at the [indiscernible word]. First, let's look at the registrars. So, whoops—well now the problem is at some point [break in audio]. Somehow in this conversion from PowerPoint to WEBEX something happened. So, I am going to talk through this part. What you find is that in fact yes, there's one registrar out of Russia

that is like 40%. Alright? 40% of the domains are registered to them. And then there's another 25% registered to two domains in China. And then the tail is incredibly diverse. Alright? And so this is in some sense suggests—Hey, you know, there is a real opportunity to make impact because you can go after a small number of registrars. Three registrars and get 50% of those domains. The down side is at this point of view is that in fact the switching cost is incredibly low. So, we did another study looking at this, where we looked at active attempts to take down the registrar level. And what you find is people are able because-- The value of a domain is 50 cents or a dollar in bulk. And so if you take it down, that resource was not very expensive and we have almost about a thousand registrars and then you know—I don't know five extra re-sellers. There are so many different places they can go. So we just find if you shut it down here they go somewhere else. So while it's appealing from the stand point, small number of people can have an impact quickly. It's not a long-term impact [break in audio]. When you look at web hosting, it's even more diverse. You don't even get a bigger concentration and here the switching cost is even lower. As I said, the cost per thousand U.S. hosts is a hundred bucks. So, ten cents a host. If I shut down your name server, your web server it's not very hard for you to get another one. So the a ha moment if when you look at merchant banks. There are three banks that monetize basically all of the payments. The biggest one is the [indiscernable] bank in Ajuba Jan. And the one in Saint Kitts Nevis which did most of the replica and herbal stuff and then India and [indiscernable place]—Latvia and [indiscernable word] did [indiscernible word] and software. Not very many banks in general in this business. Maybe 30 all total—There are—Because they do--Not very many banks want to handle what is called high risk merchants. You want to find a U.S. institution that is willing--Willingly going to offer merchants services to someone who is going to sell an online pharmaceutical. The other thing that the switching account is high. You cannot—For unlike domains and hosting, you can't click on a button and get it. You need to meet someone. There is new diligence that goes into this. Visa actually needs to co-sign credentials so that you can actually get your connection into Visa Net. So you know light speeds for this system is like you create a new account in five days. Alright? So it's just a totally different order of magnitude two times. The other thing is you have to pay money to create this account. Let's say high risk accounts, you know give us 20 grand. More important though than the upfront capital which doesn't always have to be all that big is what is called a hold back forfeiture. They know that you are high risk so instead of paying you on net 30 they'll pay you net 90. Ninety days later you get the money and if something bad happens and your account gets shut down well the bank keeps that money. So, at any given point in time they have 90 days of capital in there that will be lost if something goes wrong. This is a really valuable resource and it's painful to replace. So there's a bunch of things you might—And so this is, kind of we--So we tracked this to this day. We check it-Actually every two weeks we do purchase from every major program [indiscernible phase] which banks are sponsoring which activity or not necessary knowing---I want to be absolutely clear about this. A number of things, do not know that this is going on or are being defrauded and so we have worked with a number of them. And in some cases there is a fairly complex

laundering operation going on. But there's a bunch of things you might do. You might go to the merchant bank and say hey you should stop—You should drop these customers. This is--They are violating U.S. law and regulations. There are some challenges, you have to do this bilaterally. It's going to work on human time scales. You could also do it on the issuing side. That is when you get your credit card as a U.S citizen like from Chase or Wells Fargo and what not. They can buck the transaction as well and in fact the nice part about doing it on the issuing side is that it can be—It's just like blacklisting of anything. You can do it arbitrarily cheap or do it unilaterally. And you know 50% of U.S. credit cards are issued by banks. So, there is a capacity—There is a challenge and the incentives here are aligned. It's not costing them anything. From their standpoint there's no fraud. Their customers aren't hurting. They got burned when they were forced to do this for gambling and so this is not necessarily their favorite bunch of activity. But the point is, there are policy things one could do here. I'm actually going to talk briefly about that in the last minute which is so we have had success in doing this. So we got some major press--Front page of the science section and then the New York Time ran an opt end basically saying and you should do that. The thing that they said would be a good idea for someone to do. We give a lot of briefings to both government agencies [indiscernible terms]. And oddly enough there is all kinds of communities who are interested in this stuff. We give support to the law enforcement community, industry, banks, investigations and [indiscernible word]. In the end, there are lots of people interested and not all of them feel that they are in a position to act. The community that was easiest in fact were branch holders. And so there's actually a major initiative now underway just launched yesterday between the International Anti-counterfeiting Coalition and I think some organizational help from the Executive Office of the Presidential and the Intellectual Property Section that are doing exactly this. Alright? Where there are brands that come in provide information and trying to shut down the merchants accounts associated with that. So, over the course of next six months we will have some idea about how effective this is. We're actually working with a number of them to do that work. So, to sum up--Our work basically we're big believers that security should be data driven. And that if you are going to do this, you should be get—You should be gathering data not simply about the purely technical characteristics but about the economic and social structure and there's a real achievable research agenda here. I only touched on one aspect of this. We've done stuff on the nature of reputation and underground social networks and Third World labor and so forth. There is a lot of work that can be done here and it's a place where everyone can have more meaningful impact than I think what is easy to have on the technical side. And thank you very much. [Applause]

Keith Marzullo: So I lied he did talk about his more interesting work. And you understand why I gave him such a short introduction because I didn't want to take anytime from that but we do have time for questions and we can also I think accept questions for people who are on web ex. And I also--Before I forget again is we will be going to lunch with Stefan afterwards. And anyone who wants to join us whose here is welcome to come along.

Q: [inaudible question]

Stefan: Sure, the question was about pure online forms of payment like Bitcoin and were they to succeed, wouldn't that undermine this strobe point because it's not going through the traditional payment network. And I think that if you were to take that to the logical conclusion that you are right. Bitcoin is a kind of invented payment mechanism that offers a lot of anonymity and is independent of any central broker. The—It will--I will be shocked if in my lifetime that there is a reasonable replacement. The reason again is that it's not individuals can't learn to use Bitcoin is that you have--You have—You are trying to reach the broad U.S consumer market. Alright? So, imagine that you are trying to get all your relatives to use Bitcoin as a matter of course. The only payment instrument that has a footprint of any reasonable size in the West are these big payment card networks. And so you know as of-- Regardless of what you think about this action I mean, were it true that you could replace it with something else. Wiki leaks would not be having problems because there credit card transactions would have been shut down. They would of found one of these alternatives. In fact, it's quite challenging to find these alternatives. It's not impossible but I think it's hard to do it scale, which is what the problem is here.

Q: You mentioned there were three registrars that are—That were [Inaudible] shut down. I'm curious about those these three registrars. I live in the world of registrars and they are divided into two board categories [Inaudible].

Stephan: All registrars have to have a contractual relationship or they can't operate.

Speaker response: So very briefly. They are divided into so-called generic registries and registrars and country codes registries and there are registrars that only deal with those. And those are not regulated [Inaudible]. But the three that you are talking about are ones you think that do fall under the [Inaudible].

Stephan: That's right. Let me think this--The question is really about what kind of registries we are dealing with. Are these people who are serving up things like .com and.net? Or are they serving up country codes, TLD's like .rue or.cm? And the dominant ones at the time, were .rue and .cm and--But I don't think that the reason is principally because those registries have less oversight than Verisign. Although they do, you are quite right. They are independent. They have more independence. They still need a contractual relationship with [indiscernible term] but they are more independent. But I think that the real issue is the cost structure. So we—From the fascinating thing is that it used to be that .cn was by far most popular domain for people to spam with and its because there was a campaign that the Chinese government did and it was run through [indiscernible term] where the cost was one R and B which is about 10 cents. And it was by far the cheapest domain in existence. And so what happens just recently, actually about a year and a half ago is they changed the requirements to require quite a bit of documentation if you wanted to register with .cn which in turn raised the cost for registrars who wanted to offer .cn in

China and it raised the price to about 67 R and B which is about 10 bucks. And so it became much less attractive because it is a more expensive resource and what happened is there is perfect tracking of the decline slope in the use of .cm and the rise of .rue which was the next cheapest domain that's available. So, I think like the cost is what draws people to particular registries. Now what brings them to particular registrars I think is people who are willing to bulk register and look the other way. But the fact you are registering 100,000 random character domains and maybe I should wonder what your business is. Another question? I see a question from Keith.

Q: You talked about the work you needed to get the university on board with your research plan. What are the obstacles available are there for people who wish to jump into this line of research?

Stephan: Keith was my chair at the time. So he is well aware of the problems that we have. So, I think there are both institutional challenges to overcome. And we've written papers on what some of those are. And I'm actually happy to talk with any research groups who want to be in this base. So, there is a fair amount of trust building that has to happen. Peace mail with your university. The other thing is you need, so for example, I have very little data of my own. Alright? I have some data from these bots that we run and then from things that we visit. But there's a huge amount of data we get from others. And those others may be independent investigators. They maybe industry, they maybe organizations in other countries. And a lot of them will, are not-- You don't just ask them for data. Back when we started this enterprise with C/SIDE, we approached all kinds of people and they laughed it up. And it took quite some time of kind of building up trust to the point where they were willing to share data with us which is pretty essential for doing this work. And so, I think that another big obstacle is getting them -- I wouldn't say getting them access to data because I don't think it's a directly solvable problem. But it's helping them learn how to build up trust in those communities. And some of it is getting to know those communities through the academic conferences. They don't come to the academic conferences to find out about your work. You have to go to them. So there are a lot of things you need to do to engage the people who actually have the data required to allow you to do the research.

Q: [inaudible question]

Stefan: Well, I would hesitate to generalize from that but just so for--Replica watches are a good example. They are right up from. They say it's a replica. So, continue your question.

Speaker from previous question: So I guess the question is, what are [inaudible question]

Stefan: Alright so the question is twofold. One is that if this is not a fraudulent activity then what--Then where is it on? And the other is what's the cost associated with spam? So, let me deal with the second one first and then I will get to the first one. So, the cost--There are a bunch of ways we can talk about cost. We have direct costs that we

pay simply because we spend a billion dollars a year on empty spam. Alright? So that cost has to be born somehow, right? It's just--That's the tax we pay for the spam problem and supporting the empty spam industry. There's indirect costs that come about because you actually have to look at it, when the stuff gets real filters you have to look and so forth. So people have tried to come up with estimates of what that is in labor costs and so forth. But then the thing I want to point about is that not all spam is used for goods, right? So, in previous work we've studies the storm botnet which taken you know over several hundred thousand hosts. And spread almost exclusively through using spam as a vector to attract unit sites that would compromise you. So, spam is a vector for malware for phishing, for all kinds other activities that do have time is still fairly important. Alright? It may not dominant the amount of spam out there but there is—You know I would not want to let spam go. There are definitely cases of people and this is why in fact we have laws in the books of people getting bad drugs that have killed them. Alright? So there is a public health issue here but there is no QA on the pharmaceutical stuff. The—I think those are the principle kind of cost issues that come to mind. On the--What's the harm part, I think, it really comes down to aside from the public health aspect, it comes down to what is your view on intellectual property crime? Alright? So I tell you if you are Microsoft or Adobe you don't like the fact that people download your software for free and that people don't pay you any money—That does not support your business model and I think that you would find lawyers who are Rolex, and Movado and Faizer who would feel fairly strongly the same way regardless. And so let's put aside the pharmacy, I think that the standpoint that the Rolex guys would make is look this is our brand, we built it. We have a culture in which we get value for brand. And so, I need to find a way to maintain that value. And that comes down to how you feel about that issue.

Keith Marzullo: Alright okay. Well, if there are no further questions Let's thank Stefan again.

Stefan: Thank you very much.

[Event concluded]