Transcript

Lecture 8: Security

Hour 1

(00:00:00)

DAVID MALAN: Welcome back to Computer Science E-1. My name is David Malan. This is Lecture 8, "Security," and it's the first in a two-part lecture series on security, which is largely motivated by the sheer interest that folks tend to have in this subject, particularly as, one, it's rather in vogue these days, right? It's hard to pick up a technical paper, or even CNN.com these days, and just once per week, at least, you see some article somehow related to the violation of privacy, the loss of data, the theft of some computers.

We'll look at a few such articles tonight. Before we dive into security, and by extension, privacy, I thought I'd do one last demonstration of multimedia, which we did not air last week.

Recall that we did show this example.

("Swedish horses" on screen)

This of course was an example of what file format? Shockwave Flash, or Flash files; a .swf. And the takeaway from this particular file format was what? What was interesting, beyond just being cute, about this file format?

So it's interactive. So that was sort of a key distinguisher from some of the other file formats we looked at, right? Just hitting "stop" or "play" isn't exactly interactive. Rather, last week…

("Swedish horses" harmonize)

…we had these guys, although I tried to harmonize them a bit better.

What else was interesting about this?

STUDENT: (inaudible response)

DAVID MALAN: Yeah, that was the neat thing. If we blew this thing up, just by expanding the window, we seemed to have a lossless blow-up. That is to say, it looks just as good at this resolution as it did at the smaller resolution. Why was that, fundamentally?

STUDENT: (inaudible response)

DAVID MALAN: Because of vectors. What was does that mean?

STUDENT: (inaudible response)

DAVID MALAN: Sure.

STUDENT: (inaudible response)

DAVID MALAN: Good. So the essence of a vector-based file format is that it uses mathematics, it uses shapes, like polygons, but defined with lots of *x*'s and *y*'s , so to speak, instead of, as you said, just dots; pixels. And it was the so-called bitmapped file formats that only used pixels. And we only have finitely many pixels. Like we saw for that mountain range, as soon as you blow it up, the best the computer can do is sort of just double the size of all of those pixels. It can't give you information where there actually is none.

So a fun introduction to tonight I thought would be to do one more Shockwave demonstration. This was a little animation that was put together by a Harvard undergraduate a year or two ago. I used to play this with a bit of trepidation. But I think, as you'll see in a moment, I'm no longer as worried about word leaking out that we promote this here, in Computer Science E-1.

("Shaniqua" video plays)

So, anyhow, we'll link to that on the course's Website, if you'd like to pass it along, or show some friends. We'll just leave it at that. We won't try to draw any lessons from that one.

Anyhow, let me draw your attention to three articles that debuted on CNN.com in just the past couple of weeks.

The first one that I'll draw your attention to—these are all linked on our Lectures page—is perhaps the most recent one here, a snippet of which I shall read you.

"Starbucks loses four laptops with employee data. Nearly 60,000 employees in the United States affected by loss of personal information." The article goes on to elaborate. That's dated November 4, 2006.

From a couple weeks ago, we have another article on CNN.com: "Apple: Some iPods infected with virus. Less than one percent of video players made after September 12 infected with virus, affecting Windows users." And the article goes on to elaborate on that topic.

Another one, from a few weeks ago: "Crooks hijack online brokerage accounts. Spyware used to steal account details; then liquidate, manipulate stocks, the SEC says." And that article, too, goes on to elaborate.

So this is just in the past few weeks. I literally spent two minutes on CNN.com before lecture just searching on various security-related topics, and it was quite easy to find such topics as those, some of which you might have heard about yourself.

What tonight's lecture and next week's lecture are all about is, one, trying to put the fear into you, in some sense, so that you do understand not only just intuitively but technologically what threats actually exist, as they relate to computers and the Internet, but also what threats don't exist, as they

relate to computers and the Internet, because there's a lot of hype out there. There's a lot of misleading information. Probably several, if not all of you, have at some point received some email from some helpful but perhaps naïve coworker or friend, who forwards you an email saying, "Forward this to everyone you know because of this new virus that's going around, and you have to do *x*, *y*, and *z* in order to protect yourself.

Well, any time you receive an email like that, especially if it purports to come from "the security team at Apple," or "the security team at Microsoft," well, there are no such teams that send mass mails to the entire Internet user base about viruses, and worms, and so forth. Odds are, if you go and pull up Google, Google a few of the seemingly key words in that email, you'll find a link to something like urban legends.com, or various other Websites that document hoaxes. And those, too, are sort of a threat if you, for instance, do something you're told under the guise of being proactively secure and protective, when in reality, you're just being tricked into opening some, say, backdoor to your computer.

An email that was popular a year or two ago contained a zip file, which is a compressed file format, inside of which was an infected file. But the email was written in such a way that, in addition to the attachment, it was always signed "the Harvard.edu team," or "the gmail.com team." In other words, this email was being forged and sent out to random people via various mechanisms. But it had this sort of lure of officialness, in that, at the bottom of every email was sort of a customized signature, but it was customized just based on what your email address was.

Moreover, this infected file was password-protected—the zip file was password-protected. And in the body of the email were instructions as to how to access the contents; that is, the email itself contained the supposed password to this file. And that, too, sort of gave this phony email the lure of authenticity: "Wow, so these folks are actually being so proactive, the security folks, as to give me the solution to my problem, but they're also trying to keep the solution safe by password-protecting it and telling me the password."

Well, what happened when users double-clicked that attachment, typed in the secure password, opened the attachment, well, it's like letting the age-old Trojan horse right into your computer, not realizing that what's inside of that attachment is all of the bad guys that are subsequently going to take over your computer.

So we'll cover topics like those tonight. But what I also want to do is solicit as best as possible as many questions and concerns that you might have, whether they are grounded in technical reality, or it's just something you've heard. Let's also try to dispel whatever myths might be out there tonight.

With that said, what are some of the threats that exist on the Internet, as they relate to computers and the Internet today?

Well, let's take a simple one. You've all used the Web before, and you've all probably filled out a form before, if only to search for something on Google. But how many of you have bought something over the Internet?

Okay, so almost all of you. So you fill out a form like this. This is a screenshot of Buy.com, where one might go to buy computer equipment, or various other things. What are some of the concerns that you might have upon visiting a Website? Or already, before even taking E-1, what are the precautions that you take when visiting a Website like this to buy something?

STUDENT: That it's a mainstream Website.

DAVID MALAN: So it's a mainstream Website. Hopefully something you've heard of is a good bet, if only because at least if you're buying from this Website, there have been others before you who have done so.

Other things that come to mind; thoughts, concerns you might have?

STUDENT: (inaudible response)

DAVID MALAN: Sure. The last time I checked, even Amazon today, even though they don't encourage you to call them to complete your order, I believe even Amazon today does have a delicately hidden option where, if you're really paranoid over typing your credit card information in over the computer and over the Internet, well, you can call them up and relay it verbally.

(00:10:05)

Well, let's actually tackle that particular feature for just a moment. How many of you are concerned about using your credit card on the Internet? Okay, so most of you this time—not all of you but most of you. Why, any of you who are concerned?

Someone else—why this worry?

STUDENT: (inaudible response)

DAVID MALAN: Okay, so the obvious, perhaps: Someone could steal your credit card number. How might that happen? What's your worry?

STUDENT: (inaudible response)

DAVID MALAN: Okay, so somehow someone could find the information. In what ways might someone sniff this information? Well, let's consider a simple scenario, right? Just like people can wiretap telephones, legally or illegally, suppose that someone was sort of sniffing your connection between your computer right here and Buy.com? Well, couldn't someone who's just monitoring the Internet connection, or the CAT-5 cable, or the wireless connection between the two of you, could that person not just watch the zeros and ones go across the wire; figure out that those zeros and ones represent your credit card number, and therefore steal your information?

Well, hopefully not, right? Hopefully the Web gives you at least some measure of security. What is it that those savvy among you look for in a Website, when actually doing something that's sensitive, like inputting your credit card number?

STUDENT: (inaudible response)

DAVID MALAN: Okay, good. So you might look for a little box that tells you it's a secure Website. The most reliable place to look for such a box is where? Sort of leading, right? This yellow padlock that I'm sort of obviously standing beneath is unfortunately a rather subtle clue, right?

You have this huge screen—maybe 1040 x 768, and they decide to indicate that this is something you should trust by using this maybe 16 x 16 pixel icon in the bottom right corner of the screen?

Firefox, I believe, does something slightly more obvious, such that in the address bar up top, where it says buy.com, it will change colors to, I think, yellow these days, to indicate even more obviously that the site is secure. The funny thing about the boxes that you offer as a suggestion is that there are a lot of Websites—in order to draw your attention to the fact that they're secure, they don't even tell you to look here. They just put a GIF or a JPEG in the Web page itself. And guess what that GIF or JPEG says? Well, "This Website is secure." And, you know what? They sometimes put a big yellow padlock that's this big—maybe 100 pixels x 100 pixels. But what's the irony here?

STUDENT: (inaudible response)

DAVID MALAN: I could put that on E-1's Website and tell you all day long it's secure, when really it's not, right? So you have to be aware, too, at what visual cues you're looking for.

What is another way of telling that a Website is secure, beyond trusting this icon, which—in contrast to one that's in the Web page—this is in Internet Explorer's window.

STUDENT: (inaudible response)

DAVID MALAN: Yes. So we talked about this, I think, briefly in one of our Internet lectures. If the URL starts with "https://", what that means is that the Website is using a protocol—a sort of language called SSL, Secure Sockets Layer. And long story short, this is simply a means of encrypting information between my laptop and Buy.com's server.

So in reality—though in theory someone could be monitoring the traffic between me and Buy.com, if they had physical access to, say, the wires that somehow connect me to Buy.com, or maybe physical access to the routers through which my data is traveling—well, the beautiful thing about SSL is that it's end-to-end encryption. And by encryption, this is just a big way of saying the data is somehow scrambled. And it's scrambled so much these days that you can't figure out what the secret code is just with paper, pencil. Even a very fast computer—a Pentium 4 with a 3 GHz CPU—even that, in theory, should not be able to figure out how to break this so-called encryption.

But what that means is that you have end-to-end encryption, whereby, yeah, someone could still sniff those zeros and ones that are going between you and Buy.com, but they look like gibberish. It's as though you in grade school might have come up with some secret code to use to write letters to your best friend during class. You can pass it through the audience, other students in the classroom. Even if the teacher gets the message and confiscates it, ideally he or she can't understand the

message because you're using some kind of gibberish that you and your best friend might have constructed. Or, you know, more simply, you write it in a foreign language that the teacher doesn't understand.

And so in spirit, that's what happens when you have this little padlock, is that the data is entirely being scrambled between you and the other Website. So threats in between you and the Website really aren't of concern.

Now there is an esoteric attack, whereby you can have what's called a "man-in-the-middle attack," whereby you could create an environment whereby your browser is not connecting to Buy.com, but it's connecting to some bad guy. And that bad guy, in turn, is connected to Buy.com, and is sort of relaying the information back and forth, but in the middle is decrypting and reencrypting the information. But that is a very technical attack that certainly no one in this room, even the teaching fellows and I, should worry about, certainly, when just conducting business over the Internet. That would be a much more targeted attack.

But with that said, you go and enter your information to this Website. You go ahead and click "Submit." Does there remain a danger to your credit card?

Well, we seem to have—hopefully I've reassured you that you don't really have to worry about what's between you, point A, and Buy.com, point B. So where else might the threats be?

STUDENT: (inaudible response)

DAVID MALAN: On your hard drive. What do you mean by that?

STUDENT: (inaudible response)

DAVID MALAN: Good. So the danger, too, is that these days computers—PCs, Macs—are so flexible in what they can do. You can do almost anything computationally with a computer today, that that means you often leave traces of your behavior behind because they're not really designed to be secure systems, per se.

The most obvious incarnation of this is that, can you figure out where someone's been if they sit down at your computer and use the Internet? You can usually just click the little dropdown menu and go to the History, for instance. And there's even other evidence of this. We might spend more time on this next week, and I'll try to show you exactly how much information you can unearth about someone's Internet behavior.

But if you're typing in something like your name and your credit card number into a Website, theoretically there is a trace of that left on your hard drive, or left somewhere in RAM. Now, again, the means of accessing that information sometimes are pretty complicated, so you probably don't have to worry about your sibling, or your significant other, or really anyone who shouldn't be using your computer, getting your credit card number off the computer after you use it.

But what is more dangerous—and we'll come back to it later tonight—is a little something called "spyware," software that literally is designed, often maliciously, to spy on your behavior.

Suppose that you have previously visited some sketchy Website, or honestly, you have kids who just tend to visit game-related Websites—and with games often come free downloads. Unfortunately, with free downloads often come malicious software that's designed to… maybe just to pop up advertisements on your screen.

Worse might be software that's designed to monitor the Websites you're visiting so that the ads you do see are more targeted to places you've been. Worse yet might be a piece of software—spyware that logs every one of your keystrokes, and then every night at 4 a.m., uploads that information to some server, so that that guy who wrote that malicious software says, "Oh, you know what? It looks like David logged into his gmail account. He logged in with this username, and he logged in with this password." Even though the password on the screen might show up as just dots, well, that's just a browser thing. That's just aesthetic. The information's being stored somewhere. And if you type the letter "A, B, C," well, if spyware is running on your computer, it can simply write to a little text file, hidden somewhere on your C drive the letters A, B, C. And then it can just email or send that file off to the malicious software author at some point in time.

So where's the lesson here? Well, if you trust your own computer, you don't really have to worry about these threats. If you practice safe computing, which is a topic we'll spend a good amount of time on tonight, you probably don't really have to worry. After all, if you can't trust your own computer, what recourse do you have other than not to ever use your credit card on the Internet.

But—counterpoint—I would never, for instance, use an Internet café or a kiosk at the university to log into, say, my bank account, or something that… information that's particularly sensitive. Because if I might leave traces on my own computer of logging into some Website, imagine how many traces are lying around on the computers in the labs here, or in, say, an Internet café. You certainly don't have control over what software is on there.

In fact, Harvard now has, what, 30,000 FAS accounts? I would bet you that you could within minutes, and just a slight bit of cleverness, figure out the usernames and passwords of many users on campus, just by dropping a piece of spyware on the computer labs here. And even though the computers are designed to erase programs after a while, well, you notice—it's sort of a nice thing. If you ever try to check your email on these kiosks, or in the lab, and wow, you don't have to log in because someone already logged in for you.

Well, if someone already logged in, what does that mean? They've had access to the machine. What might they have installed? Anything, right?! And if they've installed software, and it's now running, and you sort of just have this nice open terminal that you can now do anything you want on, who knows what that computer itself might be doing, and where your information might be shipped off to.

(00:20:05)

But back to this credit card scenario. So if we're worried now that—not so much about the transfer between and A and B, but about the vulnerability of data at point A, especially if I'm not on my own computer—where else might we be concerned?

STUDENT: (inaudible response)

DAVID MALAN: All right, so we did A to B; we did A. Where's the other threat?

STUDENT: (inaudible response)

DAVID MALAN: All right, so B. So on the other end there is information being stored. Perhaps that information is being stored ephemerally, such that, as soon as the transaction's complete, Buy.com forgets your credit card number, and subsequently just requires that any time you want to buy something, you input it again.

But how many of you have used Websites multiple times that remember your credit card information, right? So how obviously is the server remembering that information. It's storing it on its hard disk somewhere, in some database of sorts.

And so some of these articles that you'll see on CNN.com about the loss of data, the loss of backup tapes, the theft of laptops—well, any data that's on these computers, if someone steals these computers, or perhaps, worse yet, a remote hacker hacks into the servers, and doesn't physically steal the machines, but copies digitally all the data, any information you provided to that Website could theoretically be lifted as well.

So is that to say that you should not buy anything over the Internet?

STUDENT: (inaudible response)

DAVID MALAN: Well, what's the takeaway? Well, there are a few push-backs on this. One, I would argue that typing in your credit card number, and expiration date, and name, and so forth, into a Web page that does have a modicum of security, and actually uses strong cryptography to scramble the information; ship that data not even to a poorly paid human, but rather to a computer system that immediately tucks it away into a database is arguably more secure than a transaction you might make going to the local store, or to buying something over the telephone, where you, by definition, have a human involved.

Even when I was growing up, I remember in high school, a kid got expelled from our high school because—and this is one of those stupid criminal stories—because he was at the Foot Locker, I think, an employee. And for whatever reason, he thought it would be a reasonably deceptive thing to do to just jot down people's credit card numbers, as they bought things, or maybe keep carbon copies of the receipts. And then order stuff himself by phone or via catalogue—there really was no Internet back then. And then, and this is always the catch. How did they catch him?

STUDENT: (inaudible response)

DAVID MALAN: He was mailing it to his home! So, anyhow, I think he was probably arraigned; certainly expelled. But this is only to say that when you have humans involved, and many fewer transactions, there's less likelihood, I would argue, that your data is going to be compromised.

I mean, back in the day, remember, like, at restaurants, you would get not only a receipt printed by a machine, but you'd also get literally the carbon paper. And those, too, were threats to your credit card information. Because what was on that piece of carbon paper that usually people would tear off and just leave on the table? Well, what's imprinted on the carbon paper?

STUDENT: The number.

DAVID MALAN: Right, the credit card number; your name, probably; the expiration date; perhaps some more information. This too was sort of a weird memory from my childhood. I have many memories of dinners with my grandfather and our family out at nice restaurants, where, all of a sudden there'd be a burst of flames on the table because my grandfather's exit strategy from these nice restaurants that gave these carbon copy slips, to dispose of them wasn't to tear it up; wasn't to put it in his pocket—was to burn it in the candle in the middle of the table. But it's secure. The data's not being stored in a form, if you've already destroyed the data.

So again, question on the table: Should you be buying things over the Internet? What's another protection in place, beyond the cryptography, and beyond the fact that there are relatively few, if any, humans involved?

STUDENT: (inaudible response)

DAVID MALAN: So you can certainly have defenses that we'll revisit tonight and next week—software that protects you against such spyware. So you could minimize that threat.

But also, I mean, if any of you have ever had a credit card stolen, how much money have you ever been out, right? This isn't perhaps the recommended way to deal with theft. But, frankly, if my credit card information were ever lifted or stolen from my pocket, or somehow stolen from a database, call American Express, explain the situation, and I would bet that it's not going to cost me anything. In fact, my American Express number is 3723-98… that's all you get.

But the point is that even credit card information, these days, especially given that electronic systems are so fast and so constantly online, it's very easy to shut these things off. And so I would say, we're not talking about your home address, necessarily, unless it's your billing address. We're not talking about your Social Security Number, which, in and of itself, has never been secure, but people tend to think more along those lines. You're also covered in just a real-world logistical sense by the credit card companies.

And, to be honest, if you take one lesson away from tonight, I would say genuinely that there is no reason, no sufficiently compelling reason not to use your credit card on the Internet provided you know there's an encrypted Website there, it's a reputable Website, and it's not, for instance, a debit card. I might be a little more concerned about a debit card, only because you might have to fight with your bank more to get those funds back into your account. But if it's a credit card, where the

charges are only going on your statement, you're in typically a better bargaining position if the bank doesn't have your money yet. But that would be the only rule of thumb I follow. I would have no concerns these days over buying things over the Internet. Heck, even with the first, what eight digits of my credit card number. You can try to figure out the other seven or so. Take you a while.

Okay, questions on credit cards; forms, as they're called, on Web pages?

What about these things? Everyone always seems to be scared about cookies. What's a cookie?

Who's scared of cookies? Okay, Eric's scared of cookies. Who else is scared? Or who else thinks, if Eric's scared…? Remember Eric was the guy who knew everything in "Jeopardy." If Eric's scared, how many of us should be scared? Okay, two. All right, so most of you are not as scared of cookies as Eric is.

Well, what are they, in the first place? Well, this is just a screenshot of a folder on my computer. This is my C drive; Documents and Settings/malan/cookies. If you use Internet Explorer, this folder—using your username, not mine—will accumulate these little text files, over time. And each of these files is usually named with your username at some Website. And inside of each of these files is usually some kind of information. It's usually just a random number, a big number.

But it can also be personal information, like your username, or your first name, your last name. It entirely depends on the Website. Because what a cookie is, quite simply, is a file that a Website you have visited installs on your computer, typically in order to remember something about you. And installs actually makes it sound scarier than it is. It saves a file on your hard drive. The design of cookies, though, is such that only that same Website should be able to read the contents of that file subsequently. Well, why might a Website want to remember who you are? Or, relatedly, why might you want a Website to remember who you are?

STUDENT: (inaudible response)

DAVID MALAN: Excellent. Just to save time, right? When you visit Amazon.com, it's a nice feature that I don't have to type my login name, my email address every time I want to log in to Amazon, because it just remembers my email address.

I do have to type my password, though there are some Websites where you can also say, "Save my password." Well, that's a convenient feature. Why do we need Websites to remember us using cookies? Why can't they just look at, you know, my IP address and say, "Oh, I've seen this guy before? This is Malan at post.harvard.edu? What's the danger in just using IPs to remember?

STUDENT: (inaudible response)

DAVID MALAN: So the IPs can change. If you have dial-up, it probably changes every time you connect to the Internet. Even if you have Comcast, or Verizon DSL, or a cable modem, even though these guys tend not to change your IP address all that often, you are by no means guaranteed to have the same IP address day by day, or even hour by hour.

And so, moreover, what do we also know about IP addresses, as it might relate to this problem? Why is not sufficient for another reason not to just remember someone's IP?

STUDENT: (inaudible response)

DAVID MALAN: Good. So two computers can appear to have the same IP address. Recall that we talked about home routers, and how they share one public IP address among many computers in your home? Well, imagine that this is a scenario that a hotel uses, for instance, so that they can really skimp and share one IP address among multiple rooms in the hotel. Consider this in your own home, or your dorm, or just your apartment with some friends. Well, if all of you to the outside world seem to have the same IP address, you probably don't want your roommate to pull up Amazon.com and be half logged in as, or fully logged in as you. That is to say, if you can't necessarily trust that someone's a returning visitor, based only on their IP address, how can you do it?

Well, you store, like, a little token, a little breadcrumb on their computer, so to speak, that proactively tells you who they are. And even though a Website could store in this cookie your email address and your password, maybe even, for some Website, they tend not to. They tend instead just to store a big number, a big random number, but a unique number. So that essentially, when you visit Amazon the next time, your browser—IE—is designed to send to Amazon, with the request for the home page, whatever the contents of Amazon.com's cookie is. And it might be a big number: 123456789; maybe even bigger.

But Amazon then uses that number, looks it up in their database, and says, "Oh, you know who I gave 123456789 to last week? David Malan, with this email address. Let me present him with a Web page that's customized to his criteria, as he informed me last time."

(00:30:20)

So where's the danger in cookies? They sound like a wonderful thing.

STUDENT: (inaudible response)

DAVID MALAN: Well, it turns out that they can be used to track where you're going on the Internet. And this is big business in the advertising business.

When you visit a Website like CNN.com, you've probably seen ads, right? Banner ads, flashy ads of some sort. And you've probably seen those on other Websites, too. If you look, as we will in the future, at the HTML for those Websites, you'll see that a lot of those ads don't come from CNN.com. They come from, say, DoubleClick.com, or Ads.com. That is to say, in a Web page, you can include content that's not only stored on CNN.com, but you can reference content that's stored elsewhere. So that my browser, Internet Explorer, not only gets the news from CNN, it gets all of the related ads from some other server, Ads.com.

Well, suppose that it's Ads.com that is storing a cookie on your computer. Well, if Ads.com has partnerships with CNN.com, MSNBC.com, Amazon.com, and a whole bunch of other Websites, if

Ads.com is sort of the middleman who's always present, just because they have really good business relationships with all of these sites, Ads.com can realize, via this simple mechanism, that, "Oh, this guy. I don't know who he is, necessarily, but I do see that he's visiting CNN.com. And I see that he's visiting MSNBC.com, and Amazon.com." Why? Because the browser is always sending the same cookie to Ads.com, even though the Web page being pulled up is a different Website all together.

Internet Explorer calls these things "third-party cookies." And one of the only features I disable when I use Internet Explorer, at least, is if you go to Internet Explorer's Internet Options, go to Privacy, Advanced, you'll see this. You can override "default cookie handling." A first-party cookie means a cookie from CNN.com itself. A third-party cookie means a cookie from, say, Ads.com when you visit CNN.com. And so what I do always is I just click "Block third-party cookies" largely because it doesn't really impact me and there's no reason that Website needs to know this information. Is it a severe threat? No, not really. Maybe I'm getting more ads about whatever topics I tend to be interested in on the Web. Maybe that's not a bad thing. But personally I feel, eh, it's not their business. There's no reason they need to know that information. And with a click of a button, I can block that information from them.

However, there have been bugs in browsers. And this is when people used to be scared more so about cookies, a few years ago, when Internet Explorer might have a mistake in it, such that it would give to CNN.com not just CNN's cookie, but someone else's cookie. And this became a concern, because now not just third parties who are supposed to be embedded in the Web page were correlating information, but any old Website could figure out who you are, not specifically, but who you are by these relationships among the Websites that you might be visiting.

Well, let's make this a little more real, beyond something like cookies. Well, what about just where you go? Right, we mentioned in a previous lecture that the U.S. government at least has been calling on ISPs to start retaining data on clients for, you know, at least three months or so to facilitate federal investigations into computer-related crimes.

Well, what does that mean? Well, that means the feds have essentially been asking Verizon, and Comcast, and other such entities, "Could you please keep around a log of every Website that your customers visit; of perhaps every instant message your customers visit?" although that would very quickly take up too much space. But odds are, they're asking for what Websites did the person visit, what are the headers of the emails that they sent? So even if the ISPs aren't keeping around all information, because you can only imagine if, in just an hour, Comcast kept copies of every byte that flowed across its network, you know, that probably is not cost-feasible.

But it is a little more feasible to maintain copies of "Who did David email in the past month? Was he emailing, you know, someone at AlQaeda.com?" All right, this is essentially, in goofy spirit, what the feds are asking for, that kind of data. "What Websites was David visiting?"

Well, it turns out that, even if the feds do or don't get what they want, someone does have this information. Who does know, by definition, what Websites you're visiting, beyond your own browser, obviously?

STUDENT: (inaudible response)

DAVID MALAN: What? Very noncommittal answers tonight.

STUDENT: The service provider.

DAVID MALAN: The service provider does. But let's assume now—and this is generally true—they don't generally keep this information around now because, one, they don't need it; two, it's expensive; and three, it's just an administrative pain to do so. So who does know where you're visiting?

STUDENT: Ads.com.

DAVID MALAN: Okay, so Ads.com, perhaps for the cookie-related reasons. Well, what if you go to CNN.com? CNN, by definition knows that you are visiting them, right?

How many of you have Google toolbar installed? So a few of you. Google toolbar is this little thing in the browser that lets you search the Web without having to go to Google.com. You can just type it into this search box up here. Well, if you don't pay close attention when installing this software, you might unknowingly say to Google, you know what? Report back to Google where I'm visiting. And the benefit to you being that they can try to give you better results if they know something about your Web browser behavior.

If you have a gmail account and you have Google toolbar, well, this little icon up here will become green when you're logged in. What that effectively means is that Google has the ability, then, and is sort of warning you with this little light, that they can be and probably are monitoring the Websites you're going to, not to be snoopy, but rather to help give you better results, based on the types of places you're clearly going.

When you type in, you know, a search query into Google, this too is becoming a big concern. AOL goofed recently, or they realized they goofed, when they disclosed all of the data related to searches that some of their clients had been using. And there was a story—I don't remember the specifics—but a woman who had no idea that published, I think, on some Website, or maybe in some paper, were details as to, you know, what kinds of Websites she had been visiting.

And on first glance, it sounded sort of weird, or sketchy, because they were medical terms. It was certainly personal, if nothing else. Well, that is so say, any time you use a search engine, that search engine knows certainly what you're searching for.

Well, just how much data is being accumulated? Well, here's an example of an analysis we ran on E-1's own logs for E-1's Website, and this is just for the month of October, from last year. So I essentially went one year back, ran a program on all of the logs that we maintain on the course's Website, which doesn't really keep information about you personally, except, like, the grades area, and so forth, which we do keep a close eye on.

But notice, that just over the month of October, this is the number of requests for pages or files on the Website. What's sort of curious about this graph? What's sort of obvious about this graph?

STUDENT: (inaudible response)

DAVID MALAN: So it hugely spikes around October 27, 2005. Why?

STUDENT: An exam.

DAVID MALAN: Exam 1, right? It is clear to us that last year's students, at least, were studying for the exam, because we could just infer from the logs that there was clearly an interest in Website content. Apparently, though, folks were not so engaged in the beginning part of the month. But this was apparent.

Well, what if we look at the data a little more closely. I'll pull up this same report, but with more information in it. Notice that among the things we can determine are the following. So here's that same weekly report, shown slightly differently, but you similarly see the peak. How much information could we get?

Well, we can look at this data, which—this is just sort of curious. We can get a sense of the most popular hours of the day that the Website is used. So it seems that the folks last semester were not really using the Website between midnight and, say, 6 a.m. But as soon as it becomes 8 a.m., 10 a.m.—folks are logging in from work, and so forth—activity seemed to pick up. So that's sort of a neat confirmation, if nothing else, of intuition.

Well, how worrisome does the data get? Well, we know, for instance, that 43 percent of the accesses of the Website came from organizations ending in ".net"; where 23 percent, or roughly 24 percent came from folks in the ".com." Okay, so fairly… somewhat interesting, but not terribly violating of anyone's privacy.

What more can we do? Well, now it gets interesting. So it appears that there's at least one student who was making 10 percent of the requests. You know, might we infer that this was the student most behind that semester, in that this IP address, or rather this fully qualified domain name—someone who has a Comcast account—was accessing a lot of information: 3,000 requests. Now, this doesn't mean we even had 3,000 downloads for him. But embedded in our Web pages, recall, are little graphics sometimes, and text, and so it just adds up. If every time you pull up one page, you're really pulling up ten small files, ten images, well, how that's how these numbers get inflated. But still, as a percentage, that's pretty significant.

Well, what else can we find out? We can look at this referring URL report. And this doesn't look very revealing right now. But it essentially tells us where people came from when they visited our Website.

So using this information can you sort of infer who on the Internet has linked to your own Website. Unfortunately, it's a little disappointing in that we seem to be the most popular link to ourselves, because have all of these cross-links.

But we can look at something that's a little juicier, "Search Query Report." We can keep track of the search terms that people used in order to find us.

(00:40:28)

So last year, last October, the most popular search on the Web that led people to us was actually the name of one of the course's teaching fellows. So twenty-six people searched for "Roman Rubinstein." Google, or whatever other search engine led that user to our Website, because last year, obviously, Roman's name was on our Website. And you can sort of see the other kinds of things people were searching for, and how common these were just within a month—not huge numbers, but certainly revealing and helpful to figure out where people are coming from.

This one, on first glance is a little weird. But you might recall this? So what this suggests, to be honest, is that, um… Well, why is this in our logs? How did someone search on "twinkies project.com" and find their way to us?

STUDENT: (inaudible response)

DAVID MALAN: Ah, wasn't the final project.

STUDENT: (inaudible response)

DAVID MALAN: It was one of the problem sets, which an astute student would know from an earlier problem set. It's one of the URLs where we ask about the significance of Whitespace. Well, it looks like some user intentionally or accidentally typed that into Google, or something. Because we had a PDF last year that mentioned twinkies.com, Google had found that over time. And there are relatively few Websites, apparently, on the Internet that say, "www.twinkies project.com." So the user found his or his way back to our own Website, which probably didn't really answer the question, but was sort of funny to see showing up in the logs there.

You can also see much more information, among which is what browsers people are using; what operating systems people are using—all of this information is disclosed when you visit a Website.

So, with that said, questions on these potential, or supposed threats to one's privacy? Yeah?

STUDENT: (inaudible response)

DAVID MALAN: Absolutely. If you wanted to actually find out who this struggling student was, you would have to, for instance, subpoena Comcast and get the information from them.

More cleverly, how else might you get this information? Let's start thinking like a security expert.

STUDENT: (inaudible response)

DAVID MALAN: Okay, bribe someone that works there. You can do it less expensively and more cleverly, as the staff, at least. What could we do? Hm?

STUDENT: (inaudible response)

DAVID MALAN: Email him and do what? How do we know who he is? Chicken and the egg.

STUDENT: (inaudible response)

DAVID MALAN: Ah! Fool him into telling you who he is. How might we do that?

STUDENT: (inaudible response)

DAVID MALAN: A survey! Tell us what your IP add… You know, we could, right? It'd be sort of clever of us to just incorporate into a problem set, "Hey, for ten points, tell us what your IP address is." People would probably give up themselves for ten points.

Well, we could also… how about this? Embedded in any email you send is the address of the computer from which you've sent that email. If you're using something like Outlook, or Eudora, or some email program that's on your computer, and isn't a Web-based program, like gmail or Hotmail, embedded in every email you send is your personal IP address. And in fact, if it's a PC, the name of your computer, which—maybe the name of your computer at home is "desktop," or mine you've seen is "laptop." So embedded in any email that I send you, a student, from my laptop, from home, using Outlook, will contain somewhere in the headers, in a place you don't normally see. But if you go to the right menu option, you'll see it, is your IP address.

So what I was thinking of, and these other tricks work as well: We email the class, and we ask, "Please respond to this email, confirming that you got it." Right? Now we've got the IP address of everyone, at least the IP address from which they've sent those emails. So that might be a clever way as well.

There's another trick you could use. This is used by spammers sometimes. Well, how many of you have ever received an email that's not just text, but looks like itself a Web page, with graphics and HTML, as it's called? Well, probably all of you, at this point. If you've ever received an email with pictures, and colors, and more, animation, even, it's probably because that email itself is HTML. It's like you've emailed a Web page.

Well, if you've been emailed a Web page, Web pages tend, by definition, to link back to some main server. So what this implies is that if I sent you an HTML-based email, with the E-1 logo, for instance, and you opened it with something like Hotmail, or gmail, or Eudora, or Outlook on your computer, to view that email, by definition, it would grab all of those images—the E-1 logo—from out server, if we constructed the email properly.

What this means is that we now know that you have read our email. Moreover, we know from where you've read it. And that, too, can be a useful trick. Even law enforcement might do this, right? If you don't want to go through the time, and the expense, and the trouble of trying of trying to

subpoena the information, and you know who your suspect is, you can solicit them in a way electronically, such that they, thinking it's completely innocuous to open an email, or respond to an instant message, they are giving up information about themselves. And this is because—you've seen in our Internet lectures how the Internet works, which basically is about those virtual envelopes, which tend not only to have destination addresses, but also those source addresses.

Well, what are some measures that you, a user, might take to defend against this information—or protect this information? Or rather, as a teaser, let me do this. This is one of our to-be-continued slides. We have Sherlock Holmes here, sort of looking for data to recover. What we'll likely do next week is talk briefly about this article, and the related topics. This was an article written a couple of years ago by an MIT PhD student—two of them. It's fascinating. This is an excerpt from an IEEE… which is an engineering magazine, related to specifically security and privacy. And we'll spend more time on this next week.

But you've probably heard that data from your computer can be recovered with the right tools, or the right expertise, or with enough money. This article sort of puts that into perspective. So if you've ever read articles about how people's credit card numbers were divulged, or people's ATM records, or their health records from a hospital, what these guys did in this research project was they bought, I think, 300 or so hard drives off of eBay. They then analyzed each of those hard drives, using their own computers, for data.

In some cases they found that the hard drives contained just zeros—completely uninteresting. In many of the cases, in worrisomely many cases, they found people's medical records, Social Security Numbers, sketchy files, financial data, all of which had been just either sold unintentionally by the people, or naively sold. Some of the cases, though, were actually traced back, as I recall, to say Best Buy, or companies like that, where, when these people traded in their computers, at least in one instance recently, they were told, "We will destroy your data before reselling this hardware." Uh-uh. It's a lot cheaper not to wipe information than it is to just resell it. This is terribly common. And one takeaway from next week, hopefully, will be, if you ever get rid of a computer and get rid of it from your possession—sell it, throw it away—if nothing else, you should wipe the hard drive with software we'll talk about next week, or at least you should physically destroy the hard drive.

And, as we've seen in this class, you all are pretty good about destroying our hardware, so we'll just give you the appropriate screwdriver to do that.

So, with that said, we'll come back to that next week. How can you actually protect against these violations potentially of your privacy? Well, how many of you have one or more passwords?

All right, and how many of you, for ten points on the next problem set, would tell me what it is?

All right, we won't put you on the spot—Okay, Eric!

So, passwords are a funny thing, because they're omnipresent. They're incredibly popular. They make good sense intuitively, right? If you want to protect access to something, just protect it with something that presumably only you know.

This is an excerpt from the script for the movie *Spaceballs*. How many of you have seen this film? Okay, how many of you then remember the following line:

(dialogue from *Spaceballs*):

Dark Helmet: So the combination is one, two, three, four, five? That's the stupidest combination I've ever heard in my life! The kind of thing an idiot would have on his luggage!

President Skroob: Well? Did it work? Where's the king?

Dark Helmet: It worked, sir. We have the combination.

President Skroob: Great. Now we can take every last breath of fresh air from planet Druidia. What's the combination?

Dark Helmet: 1 2 3 4 5.

President Skroob: 1 2 3 4 5? That's amazing! I've got the same combination on my luggage!

DAVID MALAN: Okay, think about it now. How many of you have an ATM code, have a pass… not just that, but have an ATM code, have a password that's your birth date, or maybe your husband's birth date, or your kid's birth date, or some amalgam of multiple ones. How many of you maybe use a keyword like your hometown, and spell it out on the keypad, or something like, you know, heck, even just use your Social Security Number for your password, using the last four digits of your Social Security Number? Pass phrases: You might remember the *Seinfeld* episode, where George Costanza's ATM code was what?

Oh, come now! Bosco! Yes? No? Bosco? No. All right, tough crowd. No pop culture here. No Jon Stewart. Can't use *Seinfeld*, so keep trying.

So this is to say maybe you're all more clever than we might expect. Even I have some pretty stupid passwords, but I sort of take that risk, thinking, frankly I don't really ever care if someone guesses what my password is to yankeecandles.com, where I bought a candle once. Right? So it's sort of calculated risks sometimes if you just choose an easy word to remember, or 1-2-3-4. But some people try to be more clever.

What's a good way to come up with a password?

STUDENT: (inaudible response)

(00:50:33)

DAVID MALAN: Good. So use characters other than just alphabetical characters. Use capitalization in a sort of random way. Use numbers or letters. What else might you do?

STUDENT: (inaudible response)

DAVID MALAN: So no dictionary words—more to the point. And that's an excellent point. Even though there are a lot of words in the *Oxford English Dictionary*, you know, there's a lot of gigahertz in computers these days. And a computer can guess words from a dictionary and compare them against your potential password much faster than you, the human, could. Which is to say, if you are using a password on your computer, it is probably not very difficult for someone with the right software or expertise to figure out what it is.

In fact, a friend of mine demo'd for me once… His brother was moving into a local apartment complex and, you know, he didn't even have TV, electricity, or anything set up with the local utilities, but he did have his laptop, right? These are computer science friends of mine.

This was the first day. I was helping them move in briefly, and the brother wanted to get on the Internet. Unfortunately, as you've probably seen at home, if you have nearby neighbors, a lot of people, even if they have wireless Internet access, what are they doing these days?

STUDENT: (inaudible response)

DAVID MALAN: So they're turning on security, so that you see a little padlock next to, not the Web page, but to the access point, or the wireless router—the AirPort in Applespeak—which requires that you have some password.

Well, one of the technologies being used for most of these routers, daresay, these days, is something called "WEP," wired equivalent privacy. WEP is broken. And WEP is the security mechanism that comes with, to this day, most routers. It has been supplanted by and should not be used these days in favor of something called WPA. WPA is much more secure. There's even something more secure, which is WPA2. However, if you pull up—in my own neighborhood, and in my apartment complex area, there's probably twelve or so access points that my laptop could reach. None of them use WPA or WPA2. They all are using WEP. And you can tell that if you have the right software. It'll tell you what's being used.

And the relation to the story I was telling was that there exists software, just like there exists software to crack people's passwords, there exists software to crack people's WEP pass phrases, or secret numbers, such that my friend was able to pull up this really neat Macintosh program, and within minutes—literally minutes of just sitting there, running this software while he went to the kitchen and got something to drink—by the time he came back, minutes later he was able to access that person's access point, because the computer just very rapidly figured out what the pass phrase was for WEP. And this is not uncommon. WEP is broken. And WPA is much more secure so far as everyone knows currently, at least in its known implementations on most popular platforms, operating systems, and so forth. But even WPA is not secure if your password is 1-2-3-4. It's really not that hard for a computer to guess that kind of number.

So, in short, if you are using wireless access at home, and you are worried about someone sniffing your data—not so much your SSL-protected data, right? This is irrelevant to visiting places like Buy.com. The worse someone could do is know that you're visiting Buy.com. But you have that end-to-end SSL encryption, which means they're just going to see gibberish going across the wire.

But if you're sending emails; if you're visiting sketchy Websites; if you are sending instant messages, all of that stuff is not sent usually via SSL, which means it's being sent in the clear. And using software known as a packet sniffer, you can literally watch what someone is doing on the Internet, especially in this age of wireless.

You go to a local Starbucks, which has access points which don't use encryption, which means all the Web pages and instant messages being sent from the local café are going through the airwaves. And with the right software, you can sniff all of that information.

And so it's sort of ironic. Even though, in recent years, people moved away from using devices called hubs, which are sort of a precursor to what we call switches… Recall that switches were those central nodes that were pretty smart. If data was coming from A to B, only B got the data.

Well, in the past, a hub would, upon receipt of data from A, would send it to everyone connected. And that was a bad thing. And that's the way it was in my undergraduate days, whereby you could pretty much sniff the traffic of everyone locally.

There was really no good technological solution to that. There was a very good administrative solution to that. What do you do in a university to prevent people from actually sniffing each other's data?

Right? Stiff penalties: Throw them out if they're caught doing this. So for the most part, it was not done.

The irony is, if done right, it's not detectable. So, go figure.

Nowadays, even though we've gotten toward a more secure wired solution, as soon as you move to now wireless, you've sort of opened the doors again to these potential attacks.

So, again, if you're worried about sending information in the clear, like emails or instant messages—one, just don't; or, two, you have more technological solutions, some of which we'll revisit next week. A buzzword that will come up will be a VPN, a virtual private network, for instance.

So, passwords: Use letters; use different capitalization; use special punctuation symbols.

What about this? What about being clever and saying, "You know what? Instead of using the words, um… instead of having a password of "1-2-3-4" (writes on chalkboard), why don't I use an "l-2…" uh, what does a three look like? An E. So "…E." And a four, if you flip it over is an "h." So I'm sort of being silly here, but this is a common heuristic. People will often substitute letters that look like numbers, or numbers that look like letters, because it's still sort of a mnemonic for them.

Even though it's not the sequence "1-2-3-4," at least if you speak… You know, you can do, uh… Let's see if I get this right. (writes on chalkboard). I don't even think I got this whole thing right, but Leetspeak ("l33tspeak" on board). Right? This is sort of cryptically written words, where "3" tends to

denote an "E," because it kind of looks like an "E," if you flip it around. Well, l33tspeak… And I didn't even write it out. You can write it more cryptically than that.

The thing is, if *you* know how to speak Leetspeak, or do these sort of trivial substitutions, so does a guy who is trying to crack your password. And all that guy is going to do, in addition to trying all dictionary words, is going to try all dictionary words, substituting a "1" for an "L" or an "L" for a "1," and vice-versa. So even these tricks—not so reliable. Useful, because you can remember them better; not so much more secure.

So what's the catch here, then? What's the most secure type of password?

STUDENT: (inaudible response)

DAVID MALAN: Okay, so that's…

STUDENT: (inaudible response)

DAVID MALAN: It's an excellent point, right? You have these conflicting interests, right?

You probably want a good password for at least the most sensitive information. But it's too secure if you don't remember it subsequently. All right, so a little quiz. How many of you, perhaps at the office, have a little Post-it Note on your monitor? Yes? No? How about the drawer; next to you, under the chair? Okay, at least, even if all of you are sufficiently security conscious not to have this, stroll through the office tomorrow, and see just how many folks do have Post-it Notes on their monitor, or on the top of the drawer.

This is a solution that many people employ, and it's sort of an IT person's nightmare, right? You're trying to promote safe computing, but the pushback is that it's too hard for them. It's too difficult to remember. And, as you say, I mean, even I have probably logged into hundreds of Websites over the past few years, and I've used hundreds of usernames and passwords. It's just not realistic for a human to remember all of these things.

So what do people end up doing?

STUDENT: Use the same password.

DAVID MALAN: Use the same password. What's an obvious problem then?

STUDENT: (inaudible response)

DAVID MALAN: Right. If one of those Websites—Amazon's database is stolen, or part thereof, and someone is specifically targeting you, or just takes a guess: "You know, if this person's sort of Web-savvy enough to buy at Amazon, I'm going to guess they've bought something on eBay, or they've bought something on some other popular Website. You could write a computer program that tries those same usernames and passwords on other Websites.

Questions? Comments?

Well, packet sniffing we've covered already. This was just to depict the sort of scenario that was worrisome before. We'll skip over this slide. This was just a reminder of how a TCP header is laid out. And the lesson here was simply that all that esoteric information that we waved our hands at before, at the end of the day it's going through in the clear on the Internet, which is to say anything that is being sent from your computer can be sniffed by someone in the local area, either on a wired network, if poorly implemented, or particularly on a wireless network.

Yeah?

STUDENT: Can you show you us what the interface of a packet sniffing program would look like?

DAVID MALAN: Can I show you what the interface of a packet-sniffing program would look like? Uh, maybe. Let me, during this five-minute break, test something out and see if that would be feasible—a little how-to of sorts.

Okay, let's take a five-minute break.

Hour 2

DAVID MALAN: We are back.

(01:00:00)

EUGENIA KIM: Oh, so I'm pitching tonight's and Saturday's section as well as workshop. So the section has been billed cryptically as "TF's Choice." And the TF's choice turned out to be "Content Management Systems." It's one way of creating and publishing Web pages and keeping them organized, so I highly recommend that you come to the section, if you can. And then we have a very exciting workshop, headed by Rei Diaz. It's the "Gaming" Workshop, and he will take you through a really comprehensive, like, history, and background of gaming, and plus he has an exciting demonstration planned.

(00:60:41)

DAVID MALAN: This workshop on Saturday will be Rei's dream-come-true, since he 's all about the games. And to call it "work," and to have you all there, he would love to see you. So, all right. Thank you.

Any questions—administratively or otherwise? No?

All right, so, hacking. What is it?

You're all worried about it.

STUDENT: When an outside person gets into a system that they're not welcome to . . . (rest inaudible).

DAVID MALAN: Yeah, that's pretty good. Yeah, so to hack a computer is to sort of compromise it; to break into it, using some form of technological or sociological means, even.

The term "hacker" actually used to have a positive connotation. A hacker at, say, MIT, years ago, used to just be a kid who was really good with computers, and could do neat things.

Even at MIT, what they do annually, like dressing up the dome as R2-D2, or decorating the water fountain with the fire hydrant—those are MIT hacks.

But in more recent years, the term has acquired quite the pejorative connotation, such that if you're a hacker, you are ostensibly a bad guy. So what does it mean, then, to hack into a computer? It really depends. Hacking a computer might mean writing some special software that logs someone's keystrokes. It might mean accessing their computer physically and trying to try every possible password at the prompt.

In fact, how many of you, when you log into your Windows computers, have a password to protect others from logging in? It's not that hard to circumvent that. If someone has physical access to your computer, you can put as long a Windows password on it as you want. It's not going to keep that person out if they have the right savvy, the right time, and the right expertise.

How many of you have a BIOS password? A BIOS password is one of those ones that comes up—and Dan focused on setting BIOS settings, actually, in one of his Videos of the Week in Volume 1.

A BIOS password is a password that your computer—not Windows, but your computer, the BIOS itself offers you. So that when you first turn on the computer, before you even seen the Windows logo, you see the password prompt.

How many of you use one of these things? Yes? No? Well, some companies might use this, at least to password-protect, like, those BIOS settings.

If you ever seen one of these BIOS passwords, turns out there's got to be a way to get around these things, if someone forgets the password to a computer, right?

So to remove one of these so-called BIOS passwords, it suffices to take the case off, find the little jumper, as it's called, if you attended one of our hardware sections. Connect two little pins on the Motherboard. Gone—no more password. It's literally as simple as that.

How many of you have a Dell, or have used a Dell computer that, when you first turn it on, it says, "Warning: Cover has been removed"?

Wow! All right, so, suffice it to say there exists Dell computers—I don't think they do this anymore—such that when you turn them on, if the cover had ever been removed, it would inform you as much. Sort of a nice idea on first thought, but thereafter, sort of stupid. Because, clearly, you

need to open the case at some point. And so what you have is thousands of computers out there that say, "Warning: Case has been removed," because someone put in a new hard drive, or put RAM in. And the idea was that it would suggest that your machine had been broken into, but only insofar as you, yourself, had never lifted the case off of the thing.

So you don't even see that anymore. And if you haven't seen it already, in Volume 1, I believe, "PC BIOS Settings" is one of Dan's.

And if you haven't noticed already, at the ends of some of these Videos of the Week, there are bloopers of sorts—DVD-style outtakes. I have to say, at the risk of embarrassment, Dan, I think, makes the best outtakes. In that one in particular… Am I allowed to relay this? Everyone on the Internet knows it. So…

(students laugh)

There's a wonderful blooper at the end, where this video, as I understood it, was ostensibly about how to set a BIOS password on the computer. And Dan's reaction is priceless when he boots up the computer and says, "There's no BIOS password," after having spent the eight-minute video teaching you how to set a BIOS password. So that's in the outtakes. And there's more juicy stuff like those in them.

So let's make this a bit more concrete. Hacking is sort of a general term. What is one way that one might try to get you to divulge information, which isn't necessarily hacking, but it's certainly in the same spirit of these malicious activities.

Well, phishing: This is something that many of you are familiar with. How many of you have ever received an email from, say, Citibank, asking you to confirm some of your account information, the irony being you don't have a Citibank account, right?

How many of you received such an email from PayPal, or Bank of America, or Bank of the West? Now, this was a clever one, actually. This was beautiful. One of… We talked briefly, I think, in one of our Internet lectures about domain names looking somewhat similar to reputable domain names. Well, this particular phishing attack, as it's called—and a phishing attack is whereby you sort of try to hook someone, as you would with bait, to try to entice them to divulge information to you through social engineering: presenting them with a Website or an email that looks legitimate; therefore, it looks like it's safe to type your information into. What you're really doing is typing your information into some guy's server, you know, somewhere in Europe. Or he's logging your username and password and, faster than you can go the real PayPal.com, he goes to PayPal and wires himself some money. Right? If this is done internationally, you know, whether you're using checking accounts, credit card accounts, it's not necessarily as easy to protect or go after finances as they've lost from actual debit or checking accounts.

But this phishing attack was neat, because the URL that people were being invited to go to was (writes on chalkboard) "bankofthevvest.com." In fact, if you went to this Website, it looked quite like… Okay, no "I" in "of." There we go.

Okay, so it looked like this, right? Some random bear is their logo. So this is the Bank of the West, a West-Coast-based bank. Looks pretty credible when you visited this Website. Turns out, what most people who were hooked by this scam were logging into was not that—was this (points to "bankofthevvest.com" writing on chalkboard). Now, I'll grant you my handwriting is atrocious. But I did do one thing intentionally in this scribbled writing.

STUDENT: (inaudible response)

DAVID MALAN: Yeah, not even. It wasn't "Bank of the West," it was "Bank of the VVest.com," which, frankly, in, like, a ten-pixel font at the top of your screen, when you visit www.bankofthevvest, as opposed to bankofthewest, you're not going to notice the difference, most likely. Even the most astute of computer scientists probably would not notice a subtlety like that.

Wherein lies the problem? Well, I mean, part of that is just English alphabet, or the printing of a font. It's hard to imagine a robust way of preventing even against an attack like that. And just to contextualize this, essentially people received emails that said, "Go to this address." Looked like a legitimate address. They pulled up bankofthevvest.com. They didn't notice that it was two Vs instead of a W. They tried to log into their Website. They're actually logging in, again, to that guy's server in France, who's logging their account information; logging and transferring the funds elsewhere. They've been duped.

It's a clever attack. And it's hard, again, I think, to imagine ways of protecting against each and every one of these things.

Well, what are the tricks that you can at least use to lower the probability that you're going to be duped? Well, what's perhaps the most reliable way to ensure that you are not being scammed.

STUDENT: (inaudible response)

DAVID MALAN: So you could call your bank, certainly, and try to vet it there. I would wager that the typical person you get on the phone probably isn't going to know what you're talking about. So while reasonable, you probably wouldn't get through to the right people. So probably not a useful call to make.

STUDENT: (inaudible response)

DAVID MALAN: Absolutely. So a very good defense and a good habit to get into is just, when you're accessing online account information that you're all worried about—especially banks—don't follow links. Just go open a blank Internet Explorer window, and you type in "Bank of the W-E-S-T." Odds are, you are not going to type "V-V" accidentally. But you do sometimes make typos. And so it is in fact important to be careful as to what you're typing. Because if you accidentally reverse transpose two letters, well, we saw on our Internet lecture that it's very easy to buy a domain name.

And what people will do, scammers will often do, is just try to buy domain names that are very similar to real domain names, just banking on the fact that, hey, if there's thousands of people every day accessing this Website, odds are 0.1 percent of them are going to make a typo. That typo is

going to lead them to my Website, which I made look just like Bank of the West's Website. But in fact, it's just my server in France, for instance. I keep saying "France" because there was an instance where a guy in France was doing this, a few months ago.

So what do you have to do? You have to be careful as to what you type.

What's a useful trick? Well, one way, too, if you're not quite sure what your bank's Website is—like Bank of America. It is bankofamerica.com, but maybe you're not sure: Is it BOA.com, because they use that acronym a lot?

Well, what's a good trick—not wholly guaranteed to work, but a good trick for checking what the real Website is for an address, for a Website, rather than looking up some printed material?

(01:10:10)

Well, a wonderful approach is to leverage Google itself, right? We talked briefly about... We've used Google, certainly. And if I type in "Bank of America," typically the best match will come up; the most credible match; the one to which most other people in the world are linking to.

Now, this doesn't always work. You can't necessarily trust that the first link, or the first page of results from Google are the legitimate ones. But it's another check. It's another sanity check.

And even I've done this with certain banks that I've used over the years: I'm not quite sure. What was the URL? I don't want to guess. I'd rather at least get some positive reinforcement if—probabilistic reinforcement that where I'm going is the right address.

STUDENT: (inaudible response)

DAVID MALAN: Oh, this one here? So, what Google does under "Sponsored Links"—and you might have come across this when reading up briefly on the click fraud—this Bank of America's paid-for link. They capitalize it differently. But we know from our Internet lectures, that's okay. That's even more reinforcement. Because odds are, some scammer is not going to pay the relatively high costs that Bank of America is probably incurring to get their link to be number one there. But then again, you never know.

And just as . . . you've probably seen this before. If you type in "miserable failure." Trying not to be too political here, but this is a common example. So... that's sort of funny. This is an example of what's called "Google bombing," whereby some folks via some means, a while back, sort of collaborated via blogs, and emails, and so forth, and said, "Hey, guys, let's all create a hyperlink on our Web page that goes to whitehouse.com/george, or whatever. But the text of the hyperlink is "miserable failure."

What Google tends to do, because it's a dynamic system, it sort of infers from Web pages what's the most popular and what pages are about miserable failures. If you have thousands of people on the Web with way too much free time, creating links on their home pages to George Bush's Website,

but calling those links "miserable failure," Google sort of infers a relationship between "miserable failure" and "George W. Bush."

If you go to Wikipedia, you can search for "Google bombs," or "Google bombing," and you'll see perhaps a couple of other examples of this. It's just a hack that works with Google. And Google's policy is pretty much hands off. They don't want to even manually tweak things like this. People accuse them of being political, but this is sort of a side effect of having the sort of dynamic system they do.

So, with that said, what else can you do? Well, here are four emails that I've kept, over the past few months. This is in my "show folder" for my E-1 folder here, and they're just four spams, really, that at least are sort of interesting for discussion.

Take a look at this first one. So I got this email here from E*TRADE FINANCIAL. Notice that using the latest version of most mail programs is generally a good thing. Because Microsoft and others have gotten smarter about how to help you help yourself when it comes to security, such that I mentioned earlier that HTML-based emails can be dangerous, right? They can reveal where you are because they load the E-1 logo. Then we know you've read our mail.

What Outlook 2003 will do, if you configure it as such—and this is the default behavior—is it won't show you HTML-based emails, unless you proactively say, "You know what? Go ahead and display as HTML." And then it will sometimes even go so far as to warn you, saying, "Hey, are you really sure?"

In fact, notice what it did? It showed me almost everything, but not the pictures. So, case in point, it's not downloading the pictures as one additional measure, because as soon as I download those pictures, where are they coming from? It's coming from this phisher's Website, if it's even still up. So let's try that.

All right, my email address is out there. I'll show you how many spams I've gotten in the past few weeks. So it's still working.

Now, where they're pulling that from is unclear. They could even be pulling that link from E*TRADE, just to borrow their imagery. But the point is that this is what the email looks like.

Now, you tell me, what are some suspicious signs in this email? What should be the most obvious?

STUDENT: (inaudible response)

DAVID MALAN: Sorry?

STUDENT: (inaudible response)

DAVID MALAN: Yeah, it's a little ugly, to say the least. It's not the most beautiful work of art. The biggest warning flag should hopefully be I don't have an E*TRADE account. All right, that works.

So you might wonder, well, this seems pretty stupid, this scam. If they're sending millions of emails to most people who don't even have E*TRADE accounts, what's the point? Well, again, it's scale. And the marginal cost of sending another email? It's practically zero. It pretty much is zero. Which means even if, of the million people you email, only 1 percent of them have E*TRADE accounts, well, that's still, what? Ten thousand people. That's still a lot of potential victims at relatively no cost.

Well, what else is a little worrisome, or should be?

STUDENT: Poor grammar?

DAVID MALAN: Poor grammar! This is sort of the funniest thing on the Internet, when it comes to the scams. I have yet to meet a criminal on the Internet that reaches my inbox who can spell English correctly. You would think—and this is such a common thing, honestly. Almost every solicitation that I've seen like this has some typographic, or just aesthetic bug. And you would think that, if you're going through the trouble of sending out a million emails, you might somehow spell-check the thing, or copy the prose from some real E*TRADE email. Honestly, that is one of the biggest things to look for is a typo.

I mean, even stupid things. Like look at this (points at slide on screen). There's a space before the period. And you might have to be pretty anal to notice things like this. But if you use a computer long enough, and you notice... Maybe even stupider is "eznetconnect." I mean, I don't even know what that's getting at.

So there's another telltale sign. What else is there in this email that's suspicious?

Yeah?

STUDENT: (inaudible response)

DAVID MALAN: So E*TRADE is really enthusiastic potentially. But that too. We'll chalk it up to aesthetic weirdness. Who's it sent to?

STUDENT: (inaudible response)

DAVID MALAN: So it's sent to "undisclosed recipients". That's a little weird. Now, there's an efficiency argument there. It can be more efficient for E*TRADE to just bcc everyone. However, most companies tend not to do this for the very reason that you picked up on: It looks a little suspicious. And that's effectively what this guy did: bcc'd me and probably thousands or hundreds of other people.

Who's it sent to in the greeting? "Customer". Hey, if this from E*TRADE, and I have an account with them, they probably know my name. That too is a good cue.

Now, this too, is sort of a tar pit potentially, though, because a lot of companies, reputable ones, will tell you these days in the email, this is legitimate because we know who you are. And then they'll put your name in as "Dear David Malan."

Now, that's something, right? Because it means your email address wasn't just harvested from some random Website. But frankly, when you've seen my email address on a Website, what else have you seen on that Website?

STUDENT: (inaudible response)

DAVID MALAN: My name. So even this, not surefire protection. It's not that hard to get not only lists of people's email addresses, but also their names.

So even there the industry is sort of promoting that as a way that, "Hey, only E*TRADE knows your name." Eh, that's not so true.

Anything else in this email? Well, let's scroll down. What about the link?

STUDENT: (inaudible response)

DAVID MALAN: So it's faulty, or at least the text thereof is faulty. This is sort of neat. A neat thing about Outlook, at least, and some other email programs, is if you just—don't click—but hover over the link, it will show you, if briefly, where the thing is actually going. And we'll talk about this in our HTML lecture in a couple of weeks.

But for now, know that when you make a Web page, or you send an HTML-based email, the way via which you make a link in the email is literally you type this (writes on chalkboard):
<a href=
for hyperlink reference. And then I might say something like …
<a href="http://fastmortgage.com">

then I put the name of the link. So that's why you have this dichotomy between what you're seeing and where you're going to end up.

And I could say, as they say here,
etrade!

So if this looks cryptic now, wait till you see two weeks from now what you'll be capable of doing. But that gets across the basic idea. When you have a hyperlink—be it in a Web page or an email—there's two parts: where you're going and where you're told you're going. And those do not have to be the same, as in this case, they aren't.

In fact, if we pull this up… So it looks like this guy has been shut down somehow, or other. It goes to fastmortgage.com. There's some weird spaces there. If we go to the real Website...

So that's sort of interesting. So this actually seems to go to at least a Website that exists. Just how reputable they are is perhaps unclear. So buyer beware.

Well, there were three other emails in here. Let's see if there's anything interesting.

Oh, these are always fun. So one... Again we have bad formatting, bad English, bad punctuation. And, honestly, what's really sad. You might be laughing at how silly this is. This essentially is, what, telling me he's got a lot of money. Yeah, this is coming to me as a surprise. He was totally convinced to write me in reference to the USD$11.5 million.

You know what's sadder than the English here, is the fact that are people who do get duped by these things. Not even just in email, right? You'll see on, what are those, like *Dateline*, over the years.

(00:80:00)
There's always a story occasionally of someone who's been duped out of their life savings by some scammer, which, to a typical person, would seem so obvious. But not necessarily to all folks out there. So clearly things like these work, even though hopefully to the astute eye it would not catch you.

What's with this, uh, let's see. Yeah, it's just funny when the English—they don't even try. So hopefully you would not at least fall for something like that. Is there anything else here?

Oh, so here's a fellow from Tunisia. Medical treatments. She was married to Mr. Robert. So that's... Again, I don't understand this. Like, it's one thing to get the English wrong, but the commas in the wrong place? I find this fascinating sociologically that this is so common, over the years.

Here's another one: "I have now defined the new Newton's law..." (chuckles) "I have also explained the five cosmological blunders of the last 85 years. Go to josephnfrance." Dare I? We might have to edit this video. (clicks on link.)

Oh, it's legit. Again it's in France. So there's some sketchy stuff going on in France it seems.

So we'll leave that to you. If you're interested... (reads lengthy URL).

Anyhow, where are these phishing attacks coming from? Well, they're essentially spam, right? At least that's how they're masqueraded. Just to put into context here. If you think you get a lot of spam, I dug up my spam folder here. The first email in my spam folder that I have not yet deleted is from December 20. So three or so weeks ago. And the last one is from probably this morning, as of 6:55 pm. In this folder I have received in three weeks 2,221 spams, fortunately all of which have been automatically flagged as spam. Those were not manual spams that I put there, that went into my inbox.

But there's a lot. And again, this gives you a sense, not necessarily as to how many mailing lists I am on, but just how cheap it must be for these things to be sent if, just in three weeks, you're receiving this much.

Yeah?

STUDENT: (inaudible response)

DAVID MALAN: It's a good question: Why are subject lines sometimes misspelled? Why is there gibberish in some of these emails? Let's see if we can take an example.

So, yeah, so here's an example. It's simple. So here's a spam: "Re+move y*ur email", where there's a plus here and a star there. It's just weird, if nothing else.

But, yeah, the short answer is that this sort of gibberish is often inserted into the emails so that they'll slip past the most naïve of spam filters, which might just look for keywords like "Remove your email" spelled correctly.

And simple tricks like this clearly didn't work for my email system, since this automatically got put into my spam folder. But you'll see even crazier things, where it's simply random words, or random letters. And there'll be more random words, or more random letters in the email than there is actual content. But that too, because this is a very difficult problem in computer science. It's sort of an artificial—it's an intelligence test: Is this email being sent from a human, who's sent coherent sentences, or is it gibberish? It's tough to tell if it's just gibberish, and if it's more gibberish than it is English, because if you can understand that this is gibberish and not English, what you then have is a sort of artificial intelligence, which, as you've probably seen, doesn't really exist. Computers are not so good at recognizing and repeating convincing English. So it sort of goes hand in hand with this.

Is there anything else juicy? Oh, these are always interesting. See if I regret pulling this up. A lot of the spams going around these days have random text, or even sort of legitimate text, maybe excerpted from some book. But the spam comes from the attachment, the attachment being a GIF or a JPEG, which is a nice relation with our last week's lecture.

Fingers crossed. Okay, it's just an investor's email.

So what you have here is an HTML-based email, where the spam effectively is in the form of a GIF. And what Outlook is doing is showing you that GIF as well as the email text. But this is much, much harder for a computer to detect because this requires essentially optical character recognition, which, for an arbitrary graphic like that, is difficult. That is to say, we notice this as spam; computer: not as easy to detect as much.

STUDENT: (inaudible response)

DAVID MALAN: Correct. Most likely. If we click... Actually, if we click this, nothing is happening. So nothing happens when I click on this GIF. So this is one of those... But this is one of these investment things, where they're trying to push up the price of some penny stock. So what they're hoping is you'll not click a link necessarily, but call up your broker and request a million shares of "SBNS", therefore driving up their stock price. It's clever. Send out of millions of emails. You got enough doofuses on the Internet who then call their brokers; drives up the stock price temporarily. You sell from some foreign country. Make a quick profit. How well this works I don't know.

Some spams... And we'll keep an eye out for these. Because these are another curious one. I'll get spams sometimes that are relatively short; no links, no attachments, just gibberish. Why?! I can only

imagine that these are often tests, or accidentally sent emails. But I get enough of them that it's this funny thing.

There's clearly things on the Internet I don't understand. All these misspellings for ten years of Internet usage—spams that make no sense and you get them quite frequently.

So what's another threat? Well, you've all probably heard of the notion of a virus or a worm.

What is a virus? Yeah?

STUDENT: (inaudible response)

DAVID MALAN: Destroys your computer. Pretty much. I'll qualify that. *Can* destroy your computer, pretty much. So a virus is just a piece of software that someone again, with probably too much free time, or with malicious, or with malevolent financial intent, wrote some piece of software in hopes of doing something bad with it. A virus will typically infect your computer if you put into your computer a CD that's infected; if you put a floppy in that's infected; if you put in... if you open an email that has an attachment that's infected; or, as per that article, if you buy one of those few video iPods that was infected and connect it to your PC.

A virus, however, generally requires human interaction. Simply receiving an infected email is not sufficient, for all intents and purposes, to infect your computer. You can only get infected by opening the attachment.

And dangerous attachments include anything that's executable. You should never—here's another takeaway for tonight: Never, ever—even if it's sent from someone you know—open a .exe on a PC that was emailed to you.

Fortunately you don't see these that often coming through, because most system administrators just block emails that contain executable attachments, which is sometimes annoying if you know what you're doing and you want to send an .exe.

Similarly potentially dangerous are zip files, because they, in turn, can contain files that are infected. For a while you could even get infected by, as I recall, JPEGs. There was a bug, I think, in Internet Explorer's rendering engine, where JPEGs—that is to say, if you pulled up a Web page that had an infected JPEG, because of a bug, I think in Internet Explorer it was, you would get infected by loading that image. But in a sense, that was your fault for having gone to the Web page.

Similarly have there been other bugs in Internet Explorer, and other browsers, such that if you visit a Website, and that Website sends you a GIF, or a graphic, or some kind of content that takes advantage of a bug in the browser, similarly might you get infected with a virus. And unfortunately, I can't really tell you what viruses do, because there are so darn many of them.

In fact, if we pull up Google here, and choose, let's say, "virus database", let's see what this gives us. "Symantec". "Threat Explorer". Let's see if this is good.

So just in the past few weeks, these are all viruses or worms that have been identified.

If we instead go to... "A to Z Threats and Risks". There are so many different types of malware—malevolent software—these days, that they don't even call them viruses, just, anymore. They call them "threats and risks." These are the viruses, and worms, and other such things that start with the letter A that the company called Symantec has identified and can protect against. These are those that start with the letter... Suspense building.

Okay, so suffice it to say there's... There we go: D. There are a lot that start with D, as well. Suffice it to say there's a lot of malware. There's a lot of people with free time. There's a lot of people with malicious intent that have written these things.

What can they do? Anything. Virus is just a piece of software. And if you can write a piece of software to erase the contents of your hard drive, you can write a virus to erase the contents of your hard drive.

If you want to write a program that sends email spams to everyone in your Outlook address book, you can write a virus to do precisely that. All they are, are programs. And the problem is, once you open an infected file, by definition, viruses tend to infect other files. So it's often not sufficient just to delete the infected file. The thing sort of moves around your computer and tries to hide, in a sense.

These antivirus programs, what they try to do is detect all files that have been infected, and they try to remove the piece of infection. It's like a biological virus that attaches itself to other cells. In this context it attaches itself to other files and tries to hide there. But it uses a host file to infect. It doesn't just exist on its own.

A worm, by contrast, is similar in spirit. It can do anything, but worms are dangerous in that they can... Worms are dangerous in that they can travel on their own. And this is actually where most of my research has focused over the past few years. Worms are more scary than viruses because they don't need foolish humans clicking on attachments in emails.

(00:90:12)

If you are infected with a worm, it can jump from your computer to any other vulnerable computer on the Internet, simply if you have an Internet connection.

And for a while there was one worm going around which created the sort of Catch-22 situation. There was a bug in Windows XP, and other versions, such that if there were computers on the Internet infected with this worm, and they were attacking your computer—which essentially just meant sending unsolicited virtual envelopes of data that themselves contained the worm itself, well, if your computer was on and connected to the Internet, and not behind a so-called firewall, your computer would immediately start shutting down within thirty seconds. In other words, you turn on the computer, you connect to the Internet—Comcast, whatever—thirty seconds, your computer will shut down.

Now that Catch-22 is that for the most of us is not sufficient to go to windowsupdate.microsoft.com, search your computer for available updates, install said updates, and reboot. That doesn't happen in thirty seconds.

So I even went over to a family friend's of mine to try to help resolve this for her, and it literally was this Catch-22. To get rid of the infection, I had to download the fix from Microsoft. To download the fix from Microsoft, I needed to connect to the Internet. To connect to the Internet, I was subsequently shut down after thirty seconds. And you can imagine the nightmare this causes for companies like Microsoft if you can't even fix people's problems because the threat itself renders them unable to fix it. These are very expensive, very scary problems. There are recourses, some of which we'll come back to next week. But, in short, they can do anything.

Turns out most viruses and worms, daresay, don't do anything that interesting. They are often buggy. They'll break things. But in theory they could delete all of your files, email all of your files to someone else, just send generic spams, or they could even damage hardware in select circumstances.

Yeah?

STUDENT: Then how do they fix it?

DAVID MALAN: How do they fix it? You have very smart researchers at companies like Symantec, and McAfee, and other such companies that essentially... They let one of their machines get infected, or they get the file from someone else, and they analyze it. They figure out effectively how to reverse the effects of it.

In my situation, is that what...?

STUDENT: (inaudible response)

DAVID MALAN: So I put the machine temporarily behind a firewall so that the worm, even though it was trying to get at her computer, couldn't actually get into her computer, and that let me get to Windows update. Alternatively, I could have sat down at another computer, downloaded the patch elsewhere, and then installed it with, like, a floppy, or a USB stick, or a CD. So there are ways around it—not ways that this sixty-five-year-old friend of mine could have done on her own, and that's the scary thing. That's the worrisome thing.

Yeah?

STUDENT: (inaudible response)

DAVID MALAN: StarForce.

STUDENT: (inaudible response)

DAVID MALAN: Yes, so there are... There is some software that is thought to be good, but in reality is very similar to something like a virus or worm, in that it does something without your

knowledge. Sony got reamed in the press a few months ago, because on some of their music CDs that you would buy from the store… All right, everyone's all up in arms these days about DRM, digital rights management, making sure you can't rip their CDs, copy CDs, upload them to the Internet. So what Sony thought would be appropriate would be, when you put this CD from, you know, Tower Records, into your computer, unbeknownst to you, it installs some software in the background. That software essentially makes it impossible for Windows to let you duplicate that CD.

Unfortunately one, they told you nowhere in the wrapping that this was being done; two, it installed itself stealthily, such that you could not even detect that this software was installed. You go to "Add or Remove Programs"—not there. You go to the "Task Manager"—not there. It essentially was designed to be stealthy, which suggests some suspicious intent by Sony, and frankly some stupidity if you actually think people who want to steal music won't figure out how to circumvent this.

The worst part, though, was that the software they wrote was buggy. And it was buggy in a way that could theoretically have let an outside hacker take over the computers of anyone who had played this music CD in their computer. And this was very bad PR for Sony, who's already struggling as a company, of late.

And it sort of spoke to this conflicting interest of protecting people's intellectual property, and just implementing stupid technological solutions, particularly without users' notice.

So, in fact, very much related in spirit to that is what's called "spyware," and we talked about this already tonight. Spyware, like viruses, like worms, is software that, as the name suggests, does bad stuff; it tries to spy on your computer.

The earliest incarnations of this stuff was perhaps to just pop up ads, ads relevant to where you're going on the Web. These days you see spyware being used to harvest your email address, your password. It can upload things to other Websites.

If any of you have kids and you let your kids play games on the Internet, playing games like, you know, Frogger, and free games that they might find, these kinds of sites are riddled with spyware. And spyware usually has to be installed by a user. But browsers tend not to be designed very well, such that...

How many of you have ever received a prompt that says, " Yes or no, do you want to do this?," and you realize that if you hit "No," you're not going to be able to proceed. Well, what do you hit? "Yes," especially if you don't understand the message.

Most of us, if you want to get at something, it's yes, yes, I just want to get to the content, and you don't even appreciate what the Websites are doing. And that is one common infection mechanism, where you yourself have said, "Yes, give me this infection," when really all you really wanted is "Give me Frogger, or give me the game I want to get to."

If you download file-sharing programs, those, too, are often riddled with, like, Kazaa, in particular, comes with a bunch of stuff installed with it, which is a little risky for you.

We have the fortune tonight of having a brave volunteer, in the form of Dawne, here, who brought in this nice sexy tower for us that, on first glance, looks like it's already quite buttressed with various protections, some of which keep popping up here.

I'm going to try to ignore that for now, and draw your attention to, in our final minutes tonight, to—I'll spin it this way—the only protective software that I use. And again, take that for what it's worth.

But I will say that, for the most part, it is not necessary to pay for software to protect yourself against viruses. It's not necessary to shell out for Norton AntiVirus, or McAfee Antivirus. Frankly, there's enough free stuff and enough good enough stuff out there that if you inject with that your own common sense—E-1 savvy, daresay—and practice safe computing, you won't get infected. I think in years, I've only been infected by one thing, which was some piece of some fairly innocuous spyware. I don't know how it got there. But one piece is far fewer than some of the situations you'll see. Again, that elderly friend of mine, she has grandchildren who come over and use her computer all the time. And often what I'm going over there to do is clean off the kind of stuff that she has littering her computer.

So Dawne has volunteered her computer to be a quick disinfectant case for us. We'll try not to step upon anything too suspicious or sketchy. The program I'm going to show you is Spybot. This, from personal experience, is the only program I've used to eliminate spyware. There's a bunch of options out there. I would never download something that's suggested *to you* by some Website, only because there's a lot of antispyware software masquerading as antispyware software, that in fact is spyware. So take it only from the recommendation of a friend or a reputable company.

The nice thing about Spybot is that it's free. And sort of Julia Child style, I downloaded it earlier and installed it, and it put this link on the desktop.

What I'm going to do is just run this program. We have a nice handout that will be linked on the course's Website as to how to use Spybot that will hold your hand more than this quick demonstration. But I want to show you two things. And we'll ignore any of Dawne's protections that pop up.

The two things that I personally do with Spybot, is if I use... I use Internet Explorer. The Immunize button is a wonderful thing in that it will immunize against up to 13,222 possible attacks on Internet Explorer, some of which are related, so the number's slightly inflated. But even if it's a tenth of that, that's still a lot of stuff to worry about.

What the Immunize button does, essentially, is protects Dawne's computer against Websites that might try to infect her just by visiting them, for instance. And that's a pretty quick process. Now that's done.

But what Spybot is also useful for is disinfecting a computer. If I click up here, "Search & Destroy", and "Check for problems", what that's going to do, and we'll see, perhaps in the progress bar in a moment, that it is searching now her hard drive for the 52,708 forms of spyware that the authors of

Spybot have detected over time. You will find that there exists products from, like, McAfee and Symantec that do this same thing, and they're not necessarily all equivalent. There's probably good overlap in a lot of them, such that some catch something that the others don't. This is free. Those other programs might come with your computer, and that's great. But, again, I don't recommend paying for this stuff, necessarily.

There's relatively little software you need to pay for these days. You can also, for antivirus protection, we recommend, or I personally recommend AVG, which is the only thing I've used the past few years because it's free, and it updates itself every night. It works pretty well, certainly well enough; arguably no much better than the paid-for software, even if, again, it's missing some things. But again, you can go to this Website here and download the free version.

And they, too—I just noticed today—seem to have some kind of antispyware software as well. I've not tried that.

What you want to be careful with these products is sometimes they're a little overzealous, such that you'll get these weird prompts that I keep clicking "Cancel" to that Dawne's computer is prompting me with. And this is the tradeoff with this software.

If you want really good protection, you're going to get false positives sometimes; that is, false alarms. The computer says, "Whoa, this looks sketchy. Do you want to proceed?" And then what you find yourself doing is exercising your own judgment, which, one, is probably annoying, in that you're constantly clicking prompts; and, two, you probably installed that software because you don't know better than the software. So there, too, it's not necessarily the best of situations.

What's nice about Spybot, at least as we recommend it to E-1 students, is it doesn't really ever really prompt you, and it doesn't... Maybe it errs on the side of letting something through, but it's not confusing, I think, once it's actually installed. It just blocks things silently, which is nice.

AVG similarly just runs in the bottom of one's computer. I won't install that. But I will show you one last piece of software, which is the only other tool I've ever used to remove spyware from a computer. This is the tool that I personally have had to use in the most tricky of situations, whereby the spyware just isn't coming off the computer.

And I realize that we can't really do justice to the process of disinfecting a computer in a lecture. We actually do have a section on this, "Disinfecting a Computer." Tonight's meant to just introduce you to... See, even that's sort of interesting. So we'll just see... Even I... Oh, let's just ignore that.

So what our goal here tonight is just to mention these software names; let you play around with them, if you download them off the course's Website.

HijackThis is nice because it really gives you... really gives... Where is it already running? It really gives you finer control over what you might want to remove from the computer. Why am I not seeing Spybot? HijackThis?

All right. We'll cancel this. And we'll run it again.

What HijackThis does is, if you click its "Scan" button—and again, I realize this is a cursory tour—it shows you pretty much line by line, all of the stuff that is loaded into your computer's memory at startup. A lot of this stuff comes from what's called the Windows registry, which essentially is a huge poorly designed preferences file, among whose features is that entries in this so-called registry can be loaded at startup time.

Some of you probably know that... Is it okay if I go here, Dawne, in your...? Okay, so Dawne does not presumably have any sketchy software loading up on startup.

If I go to her Startup folder, it's kind of far down. But this is where most Windows users, if they know it at all, know that stuff in your Programs' Startup folder is loaded at Startup.

But there is a whole bunch of stuff that's loaded at Startup that's not in that folder. You can see such things in an interface like this, HijackThis. And I won't spend time dwelling on this, because it sort of is, by nature, scary. And I wouldn't suggest yourself running a program like HijackThis, and solving your problems by checking every one of these boxes, and saying "Fix checked," which means remove all this stuff. You really want to exercise some discretion.

But the discretion, for those of you who are willing to be a bit adventurous, if you're experiencing some suspicious problem, is to look among these lines for the names of programs that you just don't recognize; things that are suspiciously named in weird locations.

Now the gotcha here, of course, is that—and this is the irony—a lot of spyware is not cleverly hidden. It's installed in some folder, and given some weird name, for whatever reason. So you can infer from that weirdness that it is malevolent software, and you can probably just delete it by checking a box and clicking "Fix checked". But clearly a smarter spyware author might call his spyware "MSWord.exe," put in a folder called "Microsoft Office" in your C drive, so as to just, in a social sense, confuse you, or make you think, "Oh, well, Microsoft Word, that's okay." But in reality, it's actually spyware running there.

For whatever reason, you tend to see this that often. Spyware authors—not so clever, not so concealing, probably because they don't have to be, or they don't realize they should be.

But if you're particularly savvy and a bit adventuresome, and if you're ever having a persistent problem with spyware that things like Spybot cannot solve, this is a wonderful tool to use. But you want to run it in what's called "safe mode." And this, too, I'll defer to our "Dissecting a Computer" Section.

But realize this Catch-22 as well: Often you can... It's actually not so much a Catch-22. I'm overusing that tonight. So it's a predicament. Often spyware, if it's running on your computer, and you try to remove it with tools like Spybot, or other tools, you won't be able to, because it's running. Safe mode on Windows is a mode that you can put your Windows computer into such that nothing, except Windows itself, is running. And therefore, you can run programs like Spybot to actually disinfect it.

So the teaser we'll end on tonight is what exactly is on Dawne's computer. It looks like there were—just read it off—eighteen problems found, perhaps of varying worrisomeness. There's some kind of error there, which we'll ignore for tonight's purposes. Some of these might just be cookies of some sort. Spybot's a little overzealous when it comes to protecting you against cookies, even if they're not that bad. But every one of these lines here is something that's being loaded into Dawne's computer, usually at Startup. Some of these seem to be cookies, which is fine. But these settings, these are more worrisome, since you don't want things that are clearly making some settings happen if you don't understand what they are. So a nice trick with Spybot is, if you don't know what it is, you're probably fine in just removing it, and I would promote doing that, as much.

So many thanks to Dawne for bringing in her computer. We will see you next week in our continuation of what's scary out there and how to protect yourself.

(01:45:55)

(end)