Computer Science E-1: Understanding Computers and the Internet
Harvard Extension School
Fall 2006

Transcript

Lecture 9: Security, Continued

15 November 2006

Welcome back to Computer Science E-1. So nice to see you all again. This is Lecture 9: Security, Continued. We got an e-mail just the other day from someone who had tuned into the podcast and thanked us, actually, for talking to John Stewart segment on why the internet is like a series of tubes and pipes or rather not like a series of tubes and pipes. And this podcast viewer lamented the fact at the time that no one in the audience here live, really knew what we were talking about. So a fun way I thought to start off today's lecture would be to show this 4 minute clip from the Daily Show showing John Hodgeman. This was a running gag on the Daily Show for a while, and I tried to pick up one of the most interesting/humorous clips. You can visit this on the web yourself at comedycentral.com and without further ado. Here we go.

CLIP

All right. Well, welcome back to Computer Science E-1. Let's see if we can put more of a technical spin around some of these topics and others. Last week we had our first introduction to computer security. Last week we focused more on threats and bad things. Tonight we will focus more on good things you can do in defense of those bad things. But first a bit of review.

What were some of the threats we discussed last week.

STUDENT: Viruses, worms…

Viruses, worms; all right. What is a virus?

STUDENT: Inaudible

Something that gets into your computer and spreads. Ok, how does it get into your computer?

STUDENT: Inaudible

Good, so you might accept this bad thing from an e-mail by clicking on an attachment. What kinds of attachments are particularly dangerous these days? So, anything that is executable. And for the most part we are talking about the PC world here, simply because there are… most of the viruses and worms exist for the PCs these days for various reasons; not the least of them is Windows popularity. Dot exe is an executable; dot scr is also an executable that just refers to a screen saver. So one of the lessons we tried to preach last week is that if you ever get an e-mail, even if it's from a friend saying this is the best new animation of singing horses I have ever seen… resist that temptation to open it because cute as it may be, there have been many cases of cute things that carry inside some type of infection.

When a virus is on your computer, what can it do to your computer?

STUDENT: Inaudible

Erase files, what else?

STUDENT: Inaudible

Slow it down, certainly, just by consuming CPU cycles. One of the most common activities for worms or viruses today is to send out e-mails. They come with a SMPT server of their own. SMTP stands for simple mail transfer protocol. This is just the protocol via which e-mail is sent. So, some infections come with e-mail software built in. That e-mail's sole purpose is to send out spasm. The beautiful thing about that is that the spam now appears not to be coming from some big server that you could easily blacklist that you could say is a definite source of spam, but from your mom's computer, or your computer. So this poses a much more challenging problem for anti-virus companies and anti-spam companies because you have spam not just coming form open central source, but from all over the internet.

STUDENT: When you get a spam from… [Inaudible.]

That's a good question. So, if you get an e-mail form some say rei_diaz@harvard.edu does that suggest that Rei's computer is infected? These days, no. These days what is common for spam and also for spam generated from worms and viruses is to have the headers of those e-mails forged. That means that even I can send you a spam-like message, but pretend it is from Rei. Rather if you wanted to get a sense of whose computer is actually infected, a better way of doing that is to look at the e-mail's header. Not just the: to, from date, subject line, but go to the options menu, the view menu in Microsoft Outlook for instance. Go to options and it will tell you a whole bunch of information.

Actually a question was asked a few weeks ago on "Not dumb questions" so, if you are curious to see an example of this pull up notdumbquestions.com and you will see an example of an e-mail header that I pasted in and tried to give the question asker a sense of what information is in there. Among the information in there is for instance, the sender's IP address. Sometimes when I have gotten an infected message I will occasionally glance to see if I recognize the IP address, because if it is from Harvard.edu, I might actually be able to figure out that it is from an office-mate or someone on campus. FAS computer services will sometimes do this… they realize that some kid's computer on campus is infected and it is doing some kind of malicious behavior, they will do some type of technical trick such that they can figure out whose computer is infected, and then call the kid up. Unfortunately some of the junk you will get is from someone over whom you have no control.

For instance, E-1's website was being hit by something suspicious at one point, but we traced it to some IP address abroad. At that point it was out of our hands. And unless you get an administrator who cares about your problem, there's not much you can do other than just ignore the packets that are coming in.

STUDENT: Inaudible.

Being hit… someone downloading a lot of content cyclically from the website for seemingly no good reason. By hit, you might generally mean it's like what's called a denial of service attack. This was a threat last week that we did not discuss, but a denial of service attack essentially what exactly what it means. A computer or multiple computers just request so many darn times per second, or per minute of a website, that nobody else can get through. This has happened in the past. A couple years ago, I think it was amazon.com or someone similarly big was essentially unavailable to a lot of people because there was a distributed denial of service attack. DDoS going on; it is just the result of a bunch of malicious computers, or infected computers just requesting the heck out of that server. By requesting the heck out of it, I just mean requesting a home page repeatedly. That could block a lot of people from getting in. So, it's like a crowded throng of people at a store's door… only so many people can get in at a time.

What about a worm? There is a fundamental distinction between a virus and a worm. How is a worm different?

STUDENT: Inaudible.

A worm doesn't have a stationary source… that's true, though not the best way of distinguishing them.

STUDENT: Inaudible.

Excellent. If not necessarily by the web. Just through the internet. "Through a series of tubes and pipes" do worms travel from computer to computer. This is only to say that a worm is self propagating, whereas a virus, by definition needs some stupidity or naivety on the part of the user to execute it or to get it to deliver its payload (it's badness, whatever it happens to be.) A worm, once on your computer, can jump to another computer simply on its own. Once it's on your computer it's running. And I would go so far as to say 5-6 maybe more times out of 10 today, worms get into your computer simply because of bugs in operating systems. Windows has been a huge offender of this because windows has a whole bunch of servers running on your computer, sometimes unbeknownst to you. These are servers or services as they tend to be called that do useful things. Sometimes they are more esoteric than you need to care about, but they are nonetheless services running on your computer, such that outsiders connecting to your computer (not in the sense of controlling it, but in the sense of passing information to your computer) these services have sometimes been buggy. hey have been buggy in such a way that if a worm or adversary sends the right kind of packet or the right kind of data, what they can effectively do is send an EXE through that hole and that executable can then start running on your computer. Ever since windows service-pack 2, a lot of those holes have begun to get plugged and the fact that some of you even have a firewall running on your computer. How many of you have Windows firewall? Just by turning something like that on is huge these days. Simply having your computer behind a home router in your house is a huge way of preventing or at least discouraging yourself from getting infected.

13 minutes 43 seconds

That is a perfect segue to one of the first topics on tonight's agenda which is the notion of a firewall. You have probably all heard of this. Certainly in the context of the real world, in buildings or retail centers, what is a firewall meant to do?

STUDENT: Inaudible

Block bad things out. More specifically?

STUDENT: Inaudible

In the real world now, for a moment, not the virtual world. So, in a strip mall, there are generally in well designed buildings firewalls in between the units. Why are they there?

STUDENT: Inaudible

They are meant to be walls against fire. The idea being if some restaurant in a strip mall, for instance, all of a sudden has a grease fire in the kitchen, even if that restaurant goes up in flames, the wall depicted here as just a big brick wall, which might do the trick in some cases, discourages the fire from spreading to the nearby units. That analogy is applied similarly in the virtual world. A firewall is meant as a virtual wall that is put in place to discourage bad stuff from getting into your computer, or more generally, for getting into your LAN. Think back to that "warriors of the net" video. We like showing that because it is a fun video and it sort of takes the fear factor away from TCPIP, but if you recall… part of the video had a few holes in a big wall. There was this weird thing that was sort of throwing data into the holes. Some of the holes were 25 or 80. Do you remember this pictorial? The idea of that segment in the film was that wall represented a firewall. It only had a fixed number of holes in it though which data could travel. The only types of data in the movie were allowed through, were port 80. Any packet coming in with TCP port 80 stamped on the virtual envelope, was let through. But if some data came through from port 3389 or port 443, it would have just hit the wall. Well, what was port 80? What was that a good thing that it was allowed through?

STUDENT: Inaudible

Close. HTTP, and we'll just distinguish those 2 things properly in the next week. That's fine, but port 80, recall is the port that is used by HTTP. And just so you have a bit of familiarity with a couple of others, what is 25 used by?

STUDENT: Inaudible

Good. SMPT, AKA outgoing e-mail. Any others come to mind?

STUDENT: Inaudible

We will do 2 others because you might be using them for your final projects. If I said port 21, you might say FTP, and if I said 22 you might say (This is the challenge of the acronyms) SSH, which you may have used or might use for your final project. If these aren't familiar, don't worry about them. We will get to them. Don't worry about memorizing these things. Just generally remembering these 2 numbers tends to be helpful because when you configure your mail client (be it outlook or Eudora) you will be asked what is your SMPT server. You might have to ask Comcast or Verison for the name or the fully qualified domain name of your SMPT server. And then in that same window there will be a box in which you can type the port. 25 is the default, but because there exist things like firewalls, it is not uncommon for companies or for individuals to run these same services on different ports.

The curious thing here is that most companies (for instance the ones you work for do tend to let traffic on port 80 through.) because that just means web behavior and even though they might not let you visit sketchy websites from work by filtering actual content, they generally do let all web traffic through. They might block AOL instant messenger. The might block SSH, but it turns out (and we will talk about this tonight) there are some neat tricks you can do. Intuitively, if you are behind a corporate network that will only let port 80 through, the suggestion is that you can only use the web… how might you as an engineer leverage that fact and nonetheless get around these corporate restrictions… and access for instance an e-mail server on port 25?

STUDENT: Inaudible

Good, so the simplest solution is just change the port that your service is running on. Even though we have said that the standard ports that is useful for convention in the world, but there is nothing stopping you from overriding that to your own server. You can also set up what is called a proxy (we will also come back to that tonight) which essentially means that you can send all of your traffic to one port and let that machine or that computer figure out what ports to actually send it to.

And as time permits, I will relate to you the experience I have had helping a friend working for his company in Dubai for a few months. It is the United Arab Emirates for a few months. He corporately (and I think in the entire country) was behind outside the country outside of China that prohibited delivery of certain kinds of traffic and certain websites. Well, being a Harvard graduate, he was particularly eager to exercise his freedom of speech, albeit in the Middle East and access whatever content he wanted. Nothing sketchy is what I requested of him. But being here at Harvard, I helped him set up what was called exactly that, a proxy server. Whereby he simply accessed this proxy server and by going through this proxy server he could access anything he wanted. Moreover we set up the proxy server in such a way that it was encrypted so that no one in that country or even this country could even see what he was doing. That was useful too so you couldn't even put a filter on the content that was being sent back and forth. So, I mention this one as sort of a technical aside and to say that no matter what governments or companies set up in the interest of restricting people from doing certain things, almost always there are ways around this. Sometimes the only way to truly prevent such deception is to use (as the example I used a couple of weeks ago) Harvard's AD board.

If you can't stop kids technologically from sniffing each other's traffic, well, if you catch them, just expel them; raise the penalty high enough so that you don't need a technological solution. But tonight, we are talking about technological solutions to these kinds of problems. So, with that said…

What is a firewall all about? Well, consider your typical home network. You might have your PC sitting at home and you might have your laptop here. Your PC and your laptop are going to be connected to some central point in your house, assuming that you have a cable modem or DSL modem which most of you tend to have. We will say that this is your modem (it doesn't matter the technology). It is connected somehow or another to the internet which we will represent with a capital I in a cloud. The computer is presumably wired to this router and this actually (I left something out didn't I?) This is a modem. Let's create the router here. So, this is your router, but per our conversation in lecture 5, routers these days as you would buy them for your home come with so much more functionality. What else might these routers do? Or what else might you call them aptly?

STUDENT: Inaudible

Ok, switch usually; and by switch we mean a box that has usually 4 Ethernet ports that just splits the connection so you can connect multiple wires to it. What else are these things?

STUDENT: Inaudible

Not hub, switch and hub are usually mutually exclusive. The switch is the better of the 2 because it's smart. It doesn't broadcast your incoming data to everyone, it only sends it to the guy to whom the data is intended. And this, as we said last week, discourages packet sniffing. I have omitted (as bad as my drawings tend to be) one feature of these home routers, as most of you now have them.

STUDENT: Inaudible

Yes, wireless. It has little bunny ear antenna on top which makes these things an access point as well. I also said in lecture 5 that these router switch/access points are also fire walls. Well, what does that mean? Well, when you pull up the internet on either one of these computers, and we will assume that this laptop is connected wirelessly to this router/switch/access point. Clearly all requests for internet data go through this point and go into this internet cloud. And all responses come back to the router/switch/access point here and get delivered to the appropriate machine. So, in this picture where might be the perfect place intuitively to insert some kind of protective filter?

STUDENT: Inaudible

It's always a leading question. In this router, and I will just start calling it a router for simplicity, but realize it does all of these things. If you have this central point through which all of your data is going out and coming back in, why not insert some intelligence there that does whatever kind of filtration you want? Even though I have drawn and described this as a home network, think of this left side as representing maybe 20 computers or 100 computers that are in your university or corporation. Similarly would your corporation have some kind of switch and/or router or wireless access points in the building. They would probably cost more than the $20 ones you would buy for your home, but in spirit they work exactly the same. And corporations generally do on these things is actually impose some kind of filter. Well, technologically speaking what would you want to do in this devise to prevent all the children in your home for instance or everyone in your home from using AOL instant messenger? Put on your engineering hats. What would you have to do technologically here to prevent instant messenger traffic from going back and forth?

STUDENT: Inaudible

Lock it using the firewall. Let's dig a little deeper. What do you mean by blocking it using the firewall?

STUDENT: Inaudible

So, tell the firewall not to accept this program. And how do we identify AOL as a program?

STUDENT: Inaudible

Perfect, by its port number. Does anyone know the port number for AOL instant messenger? No, so this is a good exercise because this is not an uncommon thing for you to want to do in your home network either to block such traffic or more common in home networks, is to allow certain traffic doing something called port forwarding which we will come back to in a moment. Suppose you have a question these days. Say AOL instant messenger port and let's see if within 5 seconds we can see, yep, the port for AOL instant messenger is 5190. It is much higher than these, but think of it this way… internet has been around for a while and some of these standards have been around for a while… AOL instant messenger less so… and intuitively it makes sense that these newer services tend to have higher numbered default ports. Yes?

STUDENT: Inaudible

No. Are there an indefinite number of ports? No, I believe 16 bit values are used for TCP's port numbers, and we can confirm this by looking back at that fairly scary slide in our second internet lecture that depicted the layout of a TCP datagram, and I think, if I'm getting this right, it's only 16 bits. And if you only have 16 bits for a value, roughly how many different values can you represent?

STUDENT: Inaudible

You were required to remember this for your exam. Roughly… no not 4 billion, that was 32 bits. I only asked you to remember 3 numbers in this class. You got one of them, sort of… what is 16 bits? Good, 65 thousand…65,536. So let's do a bit of review.

We even did this in Jeopardy. So, 2 to the 8th is what? 256, that's easy. What is 2 to the 16th? 65 thousand and then it fades from there. 2 to the 24th we said was roughly… a million. And then 2 the 32nd? It is roughly 4 billion. Now you say it with confidence, good. So, unnecessary information, but we have somehow asked you to remember and it doesn't seem to sink in, so we will keep trying.

That is to say there are about 65 thousand or so ports that are allowed. Frankly that's more than enough. The only ones that are officially reserved for very standard services are the ones below 1024, so pretty much the numbers beyond that the world just generally agrees on some standard, but there's no central authority there that requires the enforcement of those numbers.

STUDENT: Inaudible

So, why does Joe Shmoe's website not have a specific web port number? So, if you want your own website to be accessible by a lay person you want to run it on some standard port because by default internet explorer for instance, when you pull up http:// it assumes you want port 80. If Joe Shmoe were to run his website on any other port, you (the visitor) would have to know that port number and specify it. We haven't done this in here before, but what you need to do if you want to specify a port number in something like internet explorer (and I am pulling up WordPad to show you in big font) this should be used to visit web pages. http://cnn.com:80/ would be equivalent in most browsers to visiting just that. You will occasionally on the web see web servers being run on nonstandard ports… 8080 is one that people tend to use. You could use 1234 if you just wanted to have a bit of obscurity to your website. If you had some kind of administrative interface that even though it is password protected and everything else you don't want it to be a target of potential hacker attacks, you might just for a slight bit of additional security run it on a nonstandard port.

If by contrast, you are running your site via HTTPS, which we said last week means what? It is secure. It is using that protocol called SSL, well, SSL's standard port which the world has agreed on is 443, so why all these ports? Why may they actually be useful to you? So, for the typical user with these home routers, you are not going to be packet filtering most likely, and you are going to be a really technically savvy parent if you are prohibiting your kids from running AOL instant messenger during certain hours of the day. With that said, these $20 routers these days, often come with parental controls that allow you to do exactly that. They have tried to dumb them down by allowing you (with a menu) to choose what services to block, so that you (the parent) don't need to know any of these numbers. Underneath the hood, all it's doing is…if I say don't allow port 51190 through between 5pm and 10pm, what that's doing is configuring the software in the router to say if you get information coming in or going out that is using TCP port 51190just drop it. So, just tear up the packet and don't let it come through. That is in fact what would be happening in a firewall.

STUDENT: Inaudible

Absolutely, so if your kid is smart enough, couldn't the kid just change it somehow/ well, the short answer is yes, but the more technical answer is you need the help of someone on the outside so to speak. You can't change the port AOL is using, but what you could do for instance, is call up… you need a proxy server actually. So, if you had little Timmy form down the street who has his own computer on the internet and his parents are not so savvy, they are not restricting any of his traffic… well, Timmy could be running a server on port 80, as though he is running a web server, but it doesn't actually need to be a web server. It can be a proxy server. And if your child sets up their computer the right way, essentially saying any internet traffic I send should not go directly to the internet, but rather should go to Timmy's computer on port 80. Timmy's computer will be set up to "port forward" so to speak, any traffic that comes in out on the normal port. It's going to go to its destination. It's then going to come back to Timmy's computer. Timmy's computer is then going to them be responsible for forwarding it back. In effect this is exactly what I did with my friend who was in Dubai. I was "Timmy in that situation. (I should have chosen a better name.) But all of his traffic was going through my server, then going out on the internet. Then it was going back on my server and back to his computer. Of course, being that I was therefore the man in the middle, clearly the potential for me (little Timmy) to do what?

30 minutes 53 seconds

STUDENT: Inaudible

Mess with his information, log all of the websites he is visiting, read all of the e-mail he was sending; but we sort of did this on a trust basis. But you have to realize that with technological solutions come potential costs. The first cost is that you need someone on the outside to be doing this. But you have never seen a smarter bunch of kids (if you find the geeks in your local high school) who are trying to use the internet from their library and they want to circumvent all of the restrictions, it's not hard to get around the school systems. Especially, truth be told, the IT folks in public schools these days are the librarians. They might have installed the software that they were given but, if you know this level of detail and you are 12-years-old (16-years-old) it's going to be hard to outdo these kids if they are determined to download the latest widgets and flash-based animation games. So, it's doable.

So, with that said, firewalls block the data that you might want to keep out of your network, or to block from going out of your network (as in the case of AOL instant messenger.) Proxy servers, in spirit, do exactly what ewe described in "little Timmy", but why else might these numbers be useful to you?

How many of you have tried to use something like AOL instant messenger, but not just for IMing, but for sending files via direct connection, or used Skype, or Google talk, or any of those voice over IP programs whereby you want to establish either a direct connection for files or for voice or for video? A few of you have tried this. How many of you have tried this with some friends and failed? It times out and just doesn't work. Why might that be? This is incredibly common the more common these kinds of firewalls or routers become. Imagine if you hear (like we were doing with Victor) [Inaudible]from the typical PC user podcast. We had set up a video conference (even though I wasn't using my camera) he was sending video to us. That worked very easily for us that day because I was not behind a firewall. I was on Harvard's network, which for the most part lets almost everything through. Victor was, similarly, on a network that lets almost everything through. More common for you and for me when I am at home, is for us to be behind a firewall. It is common for your friends to be behind a firewall. The problem with things like sending files via instant messenger, or using audio or video is that you are starting to transfer a lot of data. AOL is already hugely popular for instant messaging (however their ISP is tanking) but, if it's that popular and they already have to send all these instant messages, what is the problem if now people wasn't to send files and videos, and voice through AOL instant messenger?

STUDENT: Inaudible

You need much bigger "pipes" for AOL instant messenger. Does AOL instant messenger really need to pay for that bandwidth for the hardware to support this? Even E-1 has been paying to broadcast its podcast because we needed, for a while, to pay someone to give us the bandwidth we needed (up to 1-2 terabytes per month.) Oh, I promised to do this… We are no longer paying for our bandwidth since the kind folks at switchpod.com have kindly offered to host E-1's podcast gratis for us for the coming year. Many thanks to switchpod.com. We encourage you all to visit swithcpod.com as well as to drink Coke Cola. So, with that said, it's is just silly for AOL instant messenger to be receiving and retransmitting all of your video content and all of your video content, if they can avoid it. Your instant messages meanwhile have always gone through some central server. So, as an aside, if you are ever doing sketchy things over instant messenger, or talking about building bombs, or generally things that are frowned upon, don't do it over AOL instant messenger. Theoretically they could be logging absolutely everything.

Back to the story at hand, if you want to engage in voiceover IP, or exchange video conferencing with someone and you are both behind firewalls, what does that imply about both of your IP addresses?

STUDENT: Inaudible

So, they are basically the same, but they are technically private, or as we have said, fake IP addresses. Because when you are behind one of these routers, your router does have a real IP address assigned form your ISP via your modem. The same goes true from your friend of the other side, who has a similar setup. Your individual computers (the laptop and desktop) have phony IPs. 192.168. Somehting is the typical form. The problem is when you want to exchange video or audio content, you have "the chicken and the egg" problem. Your computer (if it's not going to go through AOL) needs to make a direct connection to your friend's computer, but you don't know the real IP address of your friend's computer because it is hidden behind that firewall. Similarly, your friend does not know yours because you too have a phony IP address (or private IP address as they are more appropriately called) could be anything behind this router. What some instant messaging programs like Skype and Google talk are very good at and AOL instant messenger and MSN messenger are less good at (at least as of recent years) is figuring out a way around that. Long story short, when this does actually work, there exists special tricks the world has adopted.

STUN is the name of a protocol that is a language that programs can speak and the trick is essentially, in spirit, works like this: Both computers sort of guess where the other one is, and they start transmitting simultaneously. This will often work through firewalls such that the traffic is allowed through when typically, it would not be. If you are curious for more technical details, just pull up STUN on Wikipedia. There's a really nice article there for instance.

STUDENT: Inaudible

So, sometimes you have friends for whom you can start a direct connection, but they can't you? That's likely a function of who's behind the firewall and exactly what the firewall is doing or what the ISP is allowing through. So, it is tough to answer more specifically than that. The relevance to us is the following: Suppose you are not just trying to do these direct connections, but you want to connect to your computer from another computer of your own. For instance, what I do all the time now is I use this thing that comes with Windows called remote desktop connection. This just lets me literally connect to and control my PC in my apartment. So, for instance, earlier before we started rolling film, you all were treated to a good half hour of the Daily Show (not the John Hodgeman segment, but those of you who came early saw a good half hour of the Daily Show) and that was being streamed via the equivalent of remote desktop to this lecture hall from my apartment. Well, what remote desktop that comes with windows lets you do is not only stream content, but literally control things. If I have never said it before, the reason I have such bland desktops like this up here (the black background with the word laptop in gray at top right) is that I get confused myself. I so often connect to my different computers; I sort of need to be told what computer I am connected to. If you use remote desktop, you literally see your desktop, you literally see your desktop and your start menu, and you're my computer as though you were actually sitting in front of it. The only fundamental difference is that you will often feel that it is slower than actually sitting in front of it, obviously because you have a network in-between you and it. However, this requires that (if I want to connect to my apartment right now) I need to know my apartment's IP address, which is step one. But I also need to do a little trick. If I want to connect to my desktop (which is the one I normally want to connect to when I'm out here) I might know the IP address of this guy, but what's going to happen when I say I want to connect via remote desktop to IP address 1234? That address is going to arrive at my router and what's he going to do with it?

STUDENT: Inaudible

Well, the router will already have received an IP address presumably when I booted my desktop up, but (for instance I have multiple computers in my apartment, in fact I have 6 now I think) how does the router know to which computer to route this incoming request, if the only means by which I can identify my apartment is by it's publicly accessible IP?

STUDENT: Inaudible

How would you do it anyway? We're going to put you on the spot. All the tools you need to acknowledge have been discussed in some form tonight. The goal again, is (when I'm sitting here at this computer and I initiate remote desktop connection to my home apartment by way of it's IP address) how can my router know to whom it should send this incoming request? What might I want to do?

STUDENT: Inaudible

Ok

STUDENT: Inaudible

Ok, not bad. So each computer will be called A, B, C, or say that each computer has its own IP address. All I have to do in advance is to configure my router to do is what I called "port forwarding" earlier. That name kind of says it all. What you can do is hard code into your home router (and you can do this yourself if you want to start doing this with remote desktop) you just have to tell your router always send requests on port 33898 (I believe it is) to computer A. That is to say, always send it to the computer with IP address 192.168.1.10, for instance. And if you hard code that into your router, and you assume that your desktop's computer's address is hard coded as well (which you can easily do, or you can tell your router to always give it the same IP address.) Henceforth, anytime I try to connect to my apartment form any computer on the internet, any requests on port 3389 reach the router: the router says "oh, I'm supposed to forward this specifically to this computer" and then I will be able to connect to that computer. What I was actually showing you earlier when we played a good half hour of the Daily Show, was a little toy of mine which we had actually used before. How was I streaming Comcast's cable modem to this lecture hall?

STUDENT: Inaudible

That way, sort of… What was the toy called? Like $100, $200… worth the investment? Slingbox? Yes? No? Tough crowd? We can review tapes and I can prove to you that I have said these things before. Slingbox. So, remember I did this analogy, whereas Tivo is the sort of time-shifting devise that lets you record something and watch it later. Slingbox is a devise that lets you record something on your Tivo, but watch it, not only later, but elsewhere. So, essentially, I have a little computer called a slingbox in my apartment that is connected to my Comcast cable box, and it's also connected to my home network. I have a little devise at home that looks like a trapezoid. It too is connected to my router. It too has a private IP address, and slingbox operates on port 5001, I think. so, I've simply configured my router in addition to saying remote desktop should go here (that is 3389 should be mapped here) well, I have also said that port 5001 should be mapped here. What that means is I can also connect to my slingbox via the internet by just connecting to my publicly accessible IP address.

So, one more question here, and we have clearly moved away from security. I started talking about things in my apartment, but that's ok. So, how do you know what your own IP address is? How do I know what my apartment's IP address is? So, you can always go on your computer (to the run menu) and if you do this as I will do here, and type IP config on a PC (it's a little small) I'm currently connected to E-1's router here. Unfortunately, I'm just simulating my apartment here (just with not as many cool toys or big TVs) this thing has given me a private IP address. It's this router that is connected to Harvard's network. So, I just found out that my private IP address--- but that's not sufficient. I need to know the IP of this thing.

Well, this is a neat trick. It is not a function of the internet, it's just clever website. When in doubt these days, type what is my IP address? This would be weird if a website could tell you what your IP address is, because that suggests that the website knows the IP address of everyone on the internet. But what happens when you visit any website have we said?

STUDENT: Inaudible

You are sending information to a server; that virtual envelope that you send to the web server inside of that virtual envelope is not only "Give me your web page" but also your return address (that is your IP address. One of the most useful addresses on the internet is this one. What is my IP address? And it will tell you what your IP address is. Notice, this is different from what I just saw with IP config because what does my IP address look like to the outside world? It looks like the IP address of this thing (or equivalently this thing.) only inside my apartment does that 192168 address actually get used. The outside world knows nothing about it. So, what I could now do, if I really wanted to, if I knew (for the sake of discussion) my IP address was let's say 25.43.67.9. Suppose I knew in advance (because earlier today I had sat down in my apartment and gone to whatismyipaddress.com; it had told me; I had written it down; and now that I am at Harvard, if I wanted to connect to my computers.) I can just type that in and hit enter… via port forwarding will my request go to the right computer. There's only one problem with this. What do you know about IP addresses as they are assigned by ISPs?

STUDENT: Inaudible

They can change. They are not often changing these days. With dial-up they will change all the time. If you have dial-up you don't want to try controlling your PC via the internet anyway. You can't, certainly, watch TV via streaming video. But if your IP address might change occasionally and you are traveling or generally, you are relying on this ability to connect… that's problematic. One of the neat things you can do to solve this (and this is one thing that I use) there is a website called dyndns (short for dynamic DNS) dot com. If you go to this site, one of their services you will see (and there are others that do this) if you go to this link here dynamic DNS- a free DNS service for those with dynamic IP addresses (which is almost everyone) You can essentially sign up for a free account and it will give you a user name (like David Malian) and it will give you a password. All you have to do is install free software on any of the computers in your network (and only one of them.) so, you can download a windows program here. You configure that windows program with the username and password that you got for free. What that little program does every day, every few hours, it just contacts dyndns.org, just saying hello and it says hello from David Malian. But because it is making an internet request of dyndns's server of what is included implicitly in that request?

STUDENT: Inaudible

The IP address. So, what this informs dyndns of, effectively, what David Malian's current IP address is. And what allows me to do is… what my username does is that… it doesn't just allow me to run that software; it also allows me to visit say… davidmailain.dyndns.org. I can then type that into remote desktop, or I can type that into my slingbox software and effectively (because of the way the internet works) this will get converted to an IP address with the help of dyndns's DNS server. (And a DNS server [we didn't spend much time on it]) simply converts IP addresses to fully-qualified domain names, back and forth. So, with this layer of indirection does dyndns now tell my computer what my current IP address is, and so in this way do I never even need to know what my IP address is. The little special free software constantly updates it. 99% of the time, you will be able to connect, assuming that thing updates itself often enough.

It's a wonderful resource, if you have never used remote desktop on windows, and you can do this on Mac OS and other operating systems as well. On a PC it's as simple as going to your system control panel; going to remote; and clicking allow users to remotely connect to this computer: if you have set up say, port forwarding in your router if you are doing this at home.

Now we can tie this back into tonight's lecture. As soon as you check that box, and click OK, what are you now running the risk of?

STUDENT: Inaudible

Worms, potentially if the remote desktop software that Microsoft wrote has a bug in it that allows someone with the right type of packet to essentially infect your computer, absolutely. You have just made yourself vulnerable simply by allowing connections to be made, to begin with. What else might be problematic?

STUDENT: Inaudible

You are going to have to stop sitting her like this. What else might be problematic?

STUDENT: Inaudible

So, other people can find you or more worrisomely, other people can potentially connect to your computer. How many of you have a password of 1234? Well, if your username is davidmalain, and your password is just 1234, it's not going to take a hacker very long to just guess your password. So again, there is a tradeoff here: the convenience of accessibility of your computers and your files. What is the tradeoff? The security threats that you potentially render yourself vulnerable to. There are ways to tweak remote desktop so that (for instance) after 3 bad passwords it will just lock you out for 5 minutes. That can at least discourage the hacker form just pounding on your machine with a whole bunch of randomly generated passwords. Question?

STUDENT: Inaudible

Let's go ahead and take a 5 minute break.

49 minutes 18 seconds

All right, so we're back. So another defense, another way of keeping people out, but still letting things in that you want to come in. How many of you are familiar with the term "VPN"? So, what's a VPN? What's it good for?

Virtual private network. Good.

A lot of offices use it, and businesses, hospitals certainly. So, what is typically useful is to keep everything out of your network. That is of course the most secure LAN to maintain, but it is certainly useful to let some things into your network. Traveling sales people might want to gain access to your local file servers. Maybe they need to get access to the internet (maybe just a website that only people in the company are supposed to be able to access.) Well, you could certainly try to just password protect everything and so forth, but more secure is if you could somehow let people in your company in other locations (be it in their hotels while traveling, or be it at other offices that you might maintain elsewhere in the country or elsewhere in the city.) Wouldn't it be nice if you could link multiple networks or individual travelers with your own primary network in such a way that you have all the protections in place, but you allow them a secure tunnel (so to speak) into your network?

Well, that's exactly what a VPN is. A virtual private network creates the illusion between a network and a user, that the user if actually on that network. It creates a tunnel, in a sense that all the traffic from the traveling sales person's laptop is encrypted, sent to the company (to a firewall effectively, or more specifically to the VPN's server at the company) where it is decrypted, and then it is forwarded along to the company's fileserver or data base or internet site, or whatever medium they are trying to access. How do you use VPN software? Well, Harvard actually has a VPN. Harvard has a VPN software that I don't have installed on this computer. But if I did, imagine the screen that came up and it said enter your username and password; hit enter; and it really should be as seamless as that. Harvard has a VPN so that people off campus can connect to on-campus resources. If you have ever tried to download some of Harvard's software, it is what is called "Keyed" with a program called key-server or key-access. What this means is that you can only use some of Harvard's software that they license for your use if you are on campus. The more mobile people have become, and with distance education and all, that doesn't fly with a lot of your students can't physically access the campus to begin with. So, what Harvard offers is a VPN: vpn.fas.harvard.edu, you can download the VPN software for free from FAS's website. What this allows a distance student to do is to connect to Harvard's campus; be given a Harvard IP address (140.247.something) and then appear to the outside world as being in Harvard.edu even if they might be in some Marriott half way across the world. The downside potentially, is that if you are connected to a VPN, usually all of your internet traffic while you are connected to that VPN, goes thorough your company's server. If you are sitting in your hotel room and you are connected to your company's VPN (using your company's VPN software) odds are those instant messages you are sending to friends back home- the e-mails you are sending- the web pages you are pulling up- even though you are in the Marriott, it's possible (and it's common) that all of that data is similarly being routed through your company back out through their internet connection, back through the company of the way back, and then back to the Marriott. So, beware when you are connected to VPNs because they will sort of usurp your internet connection and start controlling it for as long as you are connected to the VPN.

But it is useful if you want to provide people with secure access to otherwise protected resources. That's all it is. It's encryption between points A and B. That too is a good segue to this topic: that of cryptography.

Cryptography is just the art or the science or the method of encrypting or scrambling data. Scrambled above you is a phrase. What is this phrase? In other words can you decrypt this phrase for us?

STUDENT: Inaudible

So, there's a bit of a clue there. Below this encrypted phrase is what appears to be radio orphan Annies. SS something or other… what is this? Unless you happen to hack away, perhaps on paper, I will seed you with some hints. This is an example of a Cesar cipher, the legend being that Cesar, back in the day, used a Cesar cipher to encrypt communications between his army personnel and so forth. I find this hard to believe though because of all the ciphers, or encryption mechanisms. The Cesar cipher is perhaps one of the dumbest ones or easiest to break. Granted, if the world had never seen cryptography before, it perhaps would not matter how easy the thing is to break if no one has even thought of how to break it in the first place. But what the Cesar cipher simply does, is it takes (let's say for the sake of our discussion) the English alphabet which has 26 letters, and it rotates them some number of places. It uses a key. That number is just from 1-26. So, if your secret key is 13, that means that you simply rotate the letters in the alphabet by 13 places. So, that is to say if I tell you that this has been encrypted with a Cesar cipher using a key of 13, can you tell me now what this thing decrypts to? In other words, just rotate the letters in the other direction 13 places. Obviously if you rotate past Z, just flip around to A; hence the circle. That's what the circle is getting at.

STUDENT: Inaudible

Anyone?

STUDENT: Inaudible

You will be hugely disappointed when you find out what it is. Just as disappointed as little Ralphie was in a Christmas Story perhaps.

STUDENT: Inaudible

Be sure to drink your Ovaltine, is in fact the answer. So, a Christmas story which will soon be on stations like TNT 24 hours a day 7 days a week, has little Ralphie saving up like, cereal box tops for months because he wants to send them in and get a Captain Midnight secret decoder ring, with which he can finally decode the message that has been delivered to him on each and every one of his cereal boxes. Unfortunately, once he is lucky enough to collect enough box tops, or whatever, and get his secret decoder ring by mail; he sits down and very anxiously and in a very tense moment in the movie, decrypts the secret message, and it's just an ad from Ovaltine. It just says be sure to drink your Ovaltine. I have just spoiled part of the movie for you, my apologies. But for tonight's purposes, this is an example of a cipher.

A cipher is just like an algorithm, or a mechanism for scrambling information in a way that is reversible. It is very easy to create a secure message if you just change things to random letters. You need to be able to reverse that process. A cipher is kind of like a mathematical formula. In this case it is just plus 13 or minus 13. In this case it is very easy, which is why I find it hard to believe the legend that says this is what Cesar was using to encrypt and decrypt messages with his generals. But so goes the story.

A Cesar cipher is just an example of a general idea which is "how do you scramble information." Fortunately the world has gone much more advanced these days. We will put the challenge to you. You want to come up with something less inane than a cipher like this, because you are more concerned about your data than what Cesar was. How long would it have taken someone to figure out what this message was? What would you do if you were to be a cryptanalysist? (which is someone who decodes messages.) How do you bruit force figure this thing out?

STUDENT: Inaudible

Try all the different combinations, right? Add one to every letter. Does it look like English? If not, try adding 2 to every letter. Does it work? Nope. So try adding 3. if you know it's a Cesar cipher, it's not going to take you that long. If you know how to program, you can have a computer do it nearly instantaneously. But, if you didn't know that it is a Cesar cipher, you might stare at it for a while as some of you might have done. You might have just done substations. Maybe the most common letter should be swapped out for A, E, I, O, U or whatever the most popular letters are in wheel of fortune. The idea is sort of the same. You can do what is called a frequency analysis and just kind of infer that if the letter B seems to be pretty common, maybe B represents the most common letter E in the English alphabet. (I think it is E.) Or in this case it was an actual cipher. It wasn't just a substitution of one letter for another.

Web browsers and web servers do use encryption we said. What is the protocol that web servers and browsers use to encrypt and decrypt information?

STUDENT: Inaudible

Yes. SSL. Secure Sockets Layer. This is just a mechanism for cryptography fortunately is much, much harder to break. In fact, a lot of the actual ciphers in use today by computers don't use 26 letter keys. Let's translate this into computer speak. If you need to represent 26 different letters A-Z, how many bits do you need? Let's try to put this into perspective. So, what if you had 3 bits with 3 bits how many different values can you represent? This is the easy one. 2 to the 3 . 2x2x2

STUDENT: Inaudible

Yes. 8. If we have 3 bits, we can only represent 8 letters. That's not enough. What if we have 4 bits? See, this is why this stuff is relevant. It's not just esoteric knowledge.

STUDENT: Inaudible

59 minutes 58 seconds

16, ok, that's not enough… What if we have 5 bits? Then we can represent 32 letters. So, it looks like for Cesar, all we need are keys that are 5 bits. We need 5 bits, even though it is more bits than we need, because we will have some wasted values. This effectively means that the Cesar cipher as we have seen it used, takes 5 bit keys. Contrast this with something like RSA. (This technically is the algorithm that something like your web browser and server use.) It is the algorithm that SSL itself uses. Common today are the use of keys that are 128 bits. Other methods of encryption will use 256 bits, even 1,024 or 2, 048 bits. The faster computers get, the easier they can handle these kinds of values.

Now, put this into perspective. A 5 bit key means that there are 32 possible keys. That doesn't take that long to guess, for human or a computer. Well, let's round up… so, let's suppose we use 8 bit keys. Now, how much harder is the problem? How many keys might you have to guess if you are using 8 bit keys? 256. Fast forward to 32 bits; if I use 32 bit keys, how many possible guesses might I have to make to guess someone's password, or secret key? So, 2 to the 32 is roughly…

STUDENT: Inaudible

4 billion. So now if 32 bit keys give you 4 billion possible values, can you even fathom what 2 to the 1024 gives you? The take away being you can do this if you want, so take 4 billion times 4 billion... that gives you… what? 2 to the 64. So it again. It takes a while to count up to 2 to the 1024. I would wager that none of us know the word; I certainly don't for what the value 2 to the 1024is. Most of you don't have calculators that can count that high. That's a lot of digits. This is only to say that when computers use cryptography, it ends to be pretty strong. And unless there is a bug in the implementation of the algorithm, or like the browser or the server, this is to be mathematically secure.

One of the things the problem set actually hints at (and this is in its extra credit) is that prime numbers these days have a lot to do with the security of a lot of cryptographic algorithms. Clearly not Cesar ciphers, but something like RSA actually, which is commonly used by web browsers and servers as part of SSL to protect your data. It relies on it being very difficult mathematically to factor large numbers. And by factor I just mean to take a big number and figure out what 2 numbers you have to multiply to get that number. And for computers, if you are talking about values that are 1024 bits long, it will take more seconds than there are atoms in the universe to determine what those 2 numbers are. (I kind of made that up, but it's in the right spirit.) It's a lot. It is more than is humanly possible these days. However; one of the scary things is (even though it is a theoretical scariness) that if someone wakes up tomorrow and announces that I figured out how to factor large numbers quickly using some special fancy algorithm, or maybe using quantum computers (if you have ever heard of these things) that does pose a problem for the cryptographers of the world because so many things like your ATM machines, your web browsers and servers, your bank security, a lot of corporate security rests on algorithms like RSA which in turn makes this mathematical assumption. Now, weather that is a real threat is sort of up for debate. When we ask you in the extra credit for problem set 6 to tell us what the 2 prime numbers are that compose this big number, you will get a sense of just how long it takes to factor a number with just 10 or 20 or so digits, let alone hundreds of digits. So, more on that in the problem set. The take-away for now is that one of the defenses against some of the bad things we have discussed… adversaries, hackers trying to get at your data is clearly cryptography. It, fortunately, is implemented by people who are very typically good in the stuff, and you (as the user) get to use it. You don't even have to know how it's working. You do have to trust it.

Well, what about these things? What are the other defenses you might install to protect against bad things? Well, what does a virus scanner do? We talked last week, when Dawn brought in her computer, about something like ABG, MacAfee virus scan, Norton anti-virus; these are all the same thing fundamentally. How does a virus scanner work? How does it detect viruses and worms and get rid of them?

STUDENT: Inaudible

Excellent. So, a virus scanner will typically search your hard drive or scan your e-mails or scan web pages you are visiting for known threats. Typically, a virus or worm can be uniquely identified by some sequence of bits. That is to say, if you are infected with a worm or virus called XYZ, what that means is that you are infected with some piece of software that has some pattern in it. And it might be a few kilobytes long, but most viruses have some unique signature. They start with the same pattern of 0s and 1s. Simply by looking on your hard drive for exactly this pattern, though probably something longer than I have written, you can determine that is the XYZ virus. Because the smart people at MacAfee or Norton, or ABG have discovered that that is the sequence of bits that is always on someone's machine that is infected with XYZ. As soon as it detects that string of bits, what the software can simply do is erase them, or delete the file, or tell you that the file is dangerous. Don't open it.

What is the downside of this approach? Or that is to say are virus scanners the cure for all types of mal-ware like viruses and worms? Clearly not or otherwise it wouldn't be a problem.

STUDENT: Inaudible

Good. So, yes the software you bought off the shelf yesterday might protect you against all the viruses and worms that existed yesterday, but suppose someone wakes up this morning and doesn't figure out how to break RSA, but instead writes the brand new worm or virus that ends up on the CNN news all day long because it was so effective in attacking people's computers. Humans, typically, need to take time to figure out what worms and viruses look like. This is why, if you have anti-virus software installed on your computer, what is it usually doing once a day?

STUDENT: Inaudible

Yes, it's down\loading new virus definitions or updates or whatever the company calls them. But that is just updating new signatures for new worms and viruses that have been discovered; or finally, vaccines that have finally been developed for them in the form of this anti-virus software. Well, that implicitly already gives the bad guy a good 24 hour window after your computer, assuming that it is even updating itself every 24 hours. Moreover, what if you have what is called a metamorphic or polymorphic worm? That's not a term you hear tossed around in the media, but this refers to a worm whose shape changes every time it infects a computer. That is to say that the worm partially encrypts itself, and when you encrypt something you have seen it scrambles things up. Well, if you scramble things up, we don't know what it is going to look like in advance, and that's an even more dangerous threat. If you just can't protect against it, at least with this type of software.

So, what do most programs do? Well, the thing I don't' like, actually, about things like Norton and MacAfee… one of which was installed on Dawn's computer, you remember the pop-ups we kept getting? It was warning me about this and that. A lot of times these programs are behavior-based. What they do is not look for known patterns, but they raise red flags when something sketchy seems to be going on; when something is trying to write to your C drive, maybe: maybe it's trying to make a network connection and even though you don't have any browsers or anything open. The problem with this kind of software is that if it has to infer from behavior that something bad might be going on…what might be the case?

STUDENT: Inaudible

You just downloaded Skype for the first time, and you double-clicked it. It has never seen Skype before, but you want to load Skype. So, what you get is this risk of false positives, as they have been called. In it, you get alerts that say this is bad, when really, it is not necessarily bad. So, again, it's a trade-off. Do you want your software prompting you (punting the decisions to you) odds are that if you are like me, you just say "I don't understand that message; and just click ok; and just move on." The problem is infection time--- so again, it's a trade-off. It is an unsolved problem that continues to plague people and probably will for some time.

Let's consider the flip-side. This is how you might protect your computer. What about the people who write software to protect your computer? Or the people that write software that runs on your computer? Well, piracy is certainly an issue in the domain of security. Piracy refers to what in the domain of computers?

STUDENT: Inaudible

Sharing software, distributing software, copying software for which you don't have licenses for, or for which you don't pay for. Have you ever opened one of those CD envelopes and on the back of the envelope is a very small print on a sticker, and as soon as you tear that sticker, you are agreeing not to steal the software to give it to a friend. But, I bet if we did a bit of anonymous survey, we would come up with a non-zero number of people here who are running software that they didn't pay for and yet somehow it is somehow on their computer. Or perhaps you have a friend who is running software that you paid for. The thing is that it is very easy to copy software. It is just bits, so what do companies usually do to discourage people from copying software? You get things like this. How many of you have installed software that requires that you type in a sequence of letters or numbers to actually install it. That is certainly one way of discouraging it. Because not only does someone have to download thee software, they also have to know this semi-secret number.

Well, you can take the fifth if you want to… How many of you have shared that secret number with a friend or said do with this you will? Well, CDs don't necessarily protect the software in any real way. So, what else can the companies do? Well, product activation as it is called. Microsoft is doing this more with windows. TO register your software, not only do you have to type in the secret code, you also need to make a network connection to Microsoft: Microsoft essentially then records some information about your computer that is not officially identifying, but does identify your computer uniquely. What they then do is if someone else tries to register that same copy of windows or office on some other computer, and the information form that computer doesn't match your won, a red flag goes up. Either the software won't install, or you will be informed that we are finding this suspicious, you can only install this software 3 more times (for instance.) This is a common approach that companies are taking. The software installation expires after a while. If you ever experience this, you might have the following problem: if you are like me, you might format your computers or you may add a piece of hardware to your computer, and you can certainly create a situation legally in which you have just changed your computer, such that it looks different enough to Microsoft, that they think you are installing the software on another computer. This too, is a fundamental problem. Oftentimes, it suffices to call the number that is on the screen, and their screening process is such that they assume that if someone is taking the time to call Microsoft's 800 number to explain that this is a legit piece of software, can you please tell me how to install it; odds are that that's not a malicious person doing this and revealing their phone number and so forth. So, you can usually work around this. The point is that this is a nuisance.

So, what about what is in the media all the time? Music, similarly, is pirited these days. It is illegally copied, movies even more so. What are companies like Apple doing when you download music and you have paid for it, to prevent you from giving it to someone else?

STUDENT: Inaudible

Have you ever tried buying a song from say, I-Tunes for 99 cents and then giving it to someone else?

STUDENT: Inaudible

What happens?

STUDENT: Inaudible

Good, so what Apple does, and what other companies are starting to use is the big buzz word "DRM" (digital rights management.) That just means nuisance for most people who just use multiple computers, and multiple devices for whom it is very reasonable to want to play you brand new MP3 on your I-Pod or you PC desktop, or your laptop, or you maybe take it with you elsewhere. DRM is sort of an interesting approach that the music and video industry has now taken. It is very similar in spirit to all the hoops the software industry has tried to make people jump through (weather or not this persists remains to be seen.) But as you know, what you have to do with stuff that you download from I-Tunes is you have to log in to Apple's server in order to listen to that music. Or you can copy it to your I-Pod but only to your I-Pod. For the most part it has proven to be very difficult to circumvent Apple's protections, although inevitably all such ciphers or protections do tend to get broken. In fact, I believe it was Sony, installed similar DNR software on their CDs so that they didn't want people copying their music or data CDs. (I forget which it was. Within day of these disks being released to the world, someone figured out how to copy these copy-protected CDs. How did they do it?

STUDENT: Inaudible

I thought it was black, but same idea. So, we will go with either. So, they took (it was as simple as this) the way they had protected the data was to put the protections, essentially, on the outside of the CD. We know about the (what do you call it?) lans and groves, sort of like a phonograph record, they used the outside part of the disk to store 0s and 1s that collectively constituted some kind of copy protection. It turns out that with a black or blue, or green Sharpie marker, you just cover that line of the disk over, put it into your computer and Bam, it is copy-able. It's amazing when you think about how many man hours went into developing that cryptography and it was defeated by a sharpie marker.

There are many stories of similarly silly protections being defeated. How many of you had a kryptonite lock for your bike that needed to be replaced? How many of you have see the video on the web that took a 20 cent Bic pen, popped the end off so you just had a plastic cylinder; wedged the thing in there and turned really hard; Bam, you have just opened a $80 lock for free. It cost them a fortune I'm sure. I was really disappointed because I had one of the locks that was supposedly vulnerable. I hurt the hell out of my hand trying to get the pen into the lock, and I could not break my own lock. I was very disappointed. They did replace it for free. It's the same idea.

So, that is to say that even though you might put your faith in some of these defenses, with enough time, money, spare time, and people will almost always figure out how to circumvent these things. So far as the industry is concerned, a lot of these measures are intended simply to raise the bar, and increase the cost of actually cracking into things.

Well, lastly, what can you do to protect your own data in the event that you decide to give your computer to someone? Or if you decide to take it to Best Buy and you don't want them looking at whatever financial files or sketchy files you have on your computer? I use the term sketchy and sketchy files in this class a lot, but so be it. So, why is this a concern? Well, if you hand your computer over to someone on E-Bay and sell it or you bring it to a service technician, well, they clearly have access to your computer. You might think "well, I have a password and username." Well, we said last week that with the right software and savvy, it's not so hard to get rid of Malian's password on your computer. In short, if you have physical access to a machine, you have access to the data. It doesn't take a huge amount of savvy to figure out how to get at it. So, what can you do before hand, if you are selling a computer? Well, just highlight everything on your C drive; drag it to the recycle bin, and then what?

STUDENT: Inaudible

Delete it. Does that solve your problem? Can you then sell it safely on E-Bay?

STUDENT: Inaudible

All right. So, no. Let's take it one step further before we answer the why behind that. Well, what if we not only erase the files like that, we "format the hard drive" by using a boot CD or a floppy disk to boot into a little blinking DOS prompt and type format; enter. Would that do it?

STUDENT: Inaudible

Formatting is sort of a thing of the past, as far as typical users go, but the answer is no. So, let's consider the deletion one. If you go and drag every darn file on your computer into the recycle bin, and remember to empty the recycle bin, why are you arguing that it is not sufficient to protect?

STUDENT: Inaudible

It's still on your hard drive. How is that? I thought I deleted it.

STUDENT: Inaudible

Good, so you are basically saying that the space formally occupied by the files you have "deleted" are simply available for use by new files. They are not actually removed. If you think back to our hardware lecture, remember that our hard drive has one or more of these platters inside of it. On each of these platters is magnetic particles of some sort. A file that takes up a megabyte or a kilobyte, might take up this much of the hard disk. Here are just 0s and 1s, but they physically be some magnetic particles. This is, for instance, my resume. Resume.doc and it happens to refer to those bits, and you might have other files similarly referenced elsewhere. Well, when a computer saves a file, the means by which it remembers what file is where is by way of a table, essentially. That table, quite simply, or in simple form, has 2 columns. One is the name, the second column is the location. You might have in this column resume.doc, and over here you might have the address 1234. What that just refers to for instance is the 1,234th bit on the disk, or the equivalent thereof. So, when you go and drag something to the recycle bin, you probably know that is not sufficient, because it is just in the recycle bin. When you empty the recycle bin, what actually happens?

STUDENT: Inaudible

It erases that. What does that mean about your resume? All of the 0s and 1s are still there, and they are only going to be overwritten by different patterns of 0s and 1s IF your operating system and hard drive decide to reuse that space. Most operating systems do not proactively overwrite those bits with, say, random bits or maybe all 0s. Mac OS does have an option that is called secure empty trash, which does erase this and that, but even most Mac users, I would hypothesize, don't know that that exists. Or they don't know that they should use it in the interests of their own privacy.

So, what does this mean? This means, if you are downloading files and saving files, and erasing them via the recycle bin, you are leaving remnants all over the place. One of the things we use to do at the district attorney's office when a new hard drive would come in for forensic analysis is just run software that knows that there are going to be a lot of deleted files on there. It just scours, no this table, because this tables has had entries removed, but it just looks at all the 0s and 1s on the platters. It looks for patterns, similar in spirit to what a virus scanner would do. Because beyond what files being identified in the windows world, with file extensions, like with doc or jpg, or gif; each of those file formats similarly, is defined by a unique pattern of 0s and 1s at the start of it, and sometimes at the end of it. Which is to say, that just by looking for the right pattern of 0s and 1s, you can find all of the JPGs that have been downloaded by that computer, even if they were in your internet cache, and subsequently erased by internet explorer. You can find all of the word documents that were on the computer, even if they have been proactively emptied from the recycle bin. In the worst case, maybe some of the bits get overwritten, the operating system might decide I need this amount of space, but it doesn't need those bits yet, so you might get 50% of the file back, or you will get half of a jpg showing up on the screen. Depending on your goals, that may be sufficient.

The short of it is that deleting files is nontrivial. Moreover, even if you proactively securely empty your trash on, say, Mac OS, or on other operating systems that have special software, you can still have traces of files elsewhere. Can anyone think, based on conversations we had in past lectures, maybe lectures 1 and 2, back in our hardware lectures… where else on your computer might there be remnants of files that you have loaded into memory, RAM, besides where they are physically stored on your hard drive?

STUDENT: Inaudible

Excellent. Virtual memory… so what was virtual memory back in lecture 1 or 2?

STUDENT: Inaudible

Elaborate, what do you mean? It's not quite like RAM, so what is it in respect to RAM? What is virtual memory?

STUDENT: Inaudible

It is an extension…

STUDENT: Inaudible

So, it's this use of hard disk space as though it were RAM. You use virtual memory so that you can generally run more programs at once. Remember that pipeline we had that ran from the hard drive to RAM to L1, to L2 cache, to the CPU? Well, we said that sometimes if you can't fit everything into RAM, which is the fastest place to be, the computer will punt some of your programs back to disk. This is why sometimes your computer will sometimes feel slow, especially when you are toggling between programs: because they are being loaded from disk and into RAM, and then back and forth. Well, a lot of programs including programs that, so-called, scrub your hard drive, fail to scrub all locations that data might have been. So, when we have talked about virtual memory, what we are really talking about is the fact that part of your hard drive is reserved for virtual memory (by Mac OS or by Windows.) Unbeknownst to you, out of your control, some files might end up elsewhere on disk, albeit temporarily, but scrubbing programs don't necessarily know how to find those bits. They certainly don't seek them out proactively. So, though you might pay for software like Window washer, I think, is a very popular one, evidence eliminator is another popular one, those programs almost always have faults on them.

The articles we passed around last week, by Simpson Garfinkle and other people at MIT, (if you haven't yet had a chance to read that, do, [because]) one of the things it eludes to is the fact that even people that have sold stuff on E-Bay, or returned it to Best Buy, have proactively scrubbed their data. The programs that you have paid $50 for or $99 for to "sanitize" your computer, almost every commercial product has bugs in it. A very scary paper was written a few months ago by a fellow at Carnegie-Mellon, I believe, and this paper essentially analyzed a whole bunch of these popular programs, 8 or 10 of them, and every one of these programs failed to do what it was advertised on its shrink wrapped box. This is to say that even people who put their trust in products you bought with the intention of doing something securely, you still have to trust the humans who are behind that software.

So, what can you, the user, do if you want to sell your computer on E-Bay, or just generally you are uncomfortable with giving your computer to someone else, or just junking it? You can scrub or wipe your computer, or wipe your computer, but you can wipe the entire thing. All of those commercial products purport to do is erase just the sensitive parts of your computer; your internet cache, your recycle bin, your temporary folders, and it is much harder to get that right because Windows is so complicated, Mac OS is so complicated, that it's hard to get every piece of evidence that might be on your computer. By evidence, I just mean sensitive data that you might have, certainly. So what you can do is wipe or scrub your whole hard drive. The one program that we recommend for this based on its history, and based on its being free, is actually linked on the course's website. It is pretty easy to use, in that it doesn't have many options, but it doesn't have a nice, pretty interface. Rather, you have to follow a fairly arcane and old-school interface. But, if you go to the security section of the website, you will see that the 3 programs that we mentioned last week are there: ABG, hijack this, and spybot. The 4th security program I would say that I personally recommend is called "Derrick's boot and nuke." You can download either a floppy disk image, or a CD. And again, it's not the easiest thing to do, but you can install this onto a CD or floppy: boot up your PC or even Macintosh these days, or Linux computer: and wipe the whole hard drive. By that I mean that you can change all of the 0s and 1s to say all 0s, or to random 0s and 1s. Even with these programs, you will get different layers of security. You will get options like do you want one pass of 0s and 1s? Do you want 7 passes, like the department of defense officially recommends? Do you want something even more secure than that? Well, to this day, there has been no published research saying doing more than just 1 passes of 0s and 1s is any more effective than doing 1 pass of doing 0s and 1s. There is a lot of myth out there, or stories, or lore, that folks like the NSA can recover your data even if you have scrubbed it 7 times over. This has yet to be demonstrated publicly or empirically. So, most any software that scrubs the data at least once, certainly 7 times is sufficient protection against someone with an electron microscope or other such expensive tools.

STUDENT: Inaudible

1 hour 27 minutes 38 seconds

System recovery? What about?

STUDENT: Inaudible

Good question. So if you have backed up all of your important data, your documents, excel spread sheets and so forth, then ran your system restore CDs or the special partition you might have. Does that overwrite all of the data? Odds are, no. Because the bits used by your hard drive is partly a function of the operating system, partly a function of the hard drive itself, and it's not guaranteed that the software is going to overwrite the same 0s and 1s. Moreover, when you restore your computer to its original state, you have less information there than you did just before you restored it. This is to say, that if your operating system only need theses bits, you probably had a lot of data that was here, here, here, and here; in windows and these other programs don't go on your whole hard drive and erase everything. They only use what space the want. In fact, on e of the slowest parts of installing a new operating system is the formatting step. You will see a progress bar that maybe takes 10 minutes, 20 minutes, an hour to "format" your hard drive. And if you format your hard drive, you will often receive a message that says "warning! This will erase all of your data." It is a lie; it will not erase all of your data. It will not erase most of your data. This is to say, that process formatting is usually doing if it's slow, is checking the integrity of all of the bits, or the sectors of your hard drive just making sure that they are still working. Because, what a computer can do is if part of your hard drive is starting to fail, it can map around it, so you can use the rest of the hard drive. That's what's slow. It is a age old lie that when you are informed by Windows this will erase information on your disk. It is not true. It will simply make it slightly harder to get it back.

Other questions?

STUDENT: Inaudible

STUDENT: Inaudible

When you format your hard drive all the computer is doing is setting up partition tales usually. Though, that is technically a different process. The computer will typically partition it, which essentially means to set up the patterns of 0s and 1s that make it look like a C drive or D drive to an operating system. And too, it often checks the integrity of all of the locations on the disk. It tries to fix or write around any that are broken. The formatting does not officially erase data, it just happens to overwrite during that process, some of the data. To actually wipe data you need something (free, for instance, like this Derrick's boot and nuke.)

Other questions?

STUDENT: Inaudible

Yes, so how are CD keys and cereal numbers and so forth generated? Usually it is some type of algorithm. It is very similar in spirit to the stuff we briefly discussed. With that said, you can often… part of the world of piracy, is the world of wares. And here too, is sort of lip-speak type stuff, where you might see it written with weird capitalization. Wares just mean legally distributed software. There are plenty of wares floating around. You can download things, no only like EXEs these days, but if you have ever seen them, ISOs. ISO is just a file format for CDs. This is to say that you can download entire images of CDs or DVDs and with them movies these days. And this is what a lot of the traffic being used by programs like bit torrent, if you have heard the names are downloading. ISOs, MP3s, movie files, and so forth, and you can read crazy statistics these days that a huge percent of internet traffic these days is the result of people using things like file sharing programs, and downloading again and again lots of huge content. How are the registration and CD keys generated? Usually by some fairly complicated mathematical algorithm, but not usually not so complicated that someone with enough free time, can't figure out what it is. So, also in the world of wares are CD cracks and program cracks, which are either little pieces of software that analyze your dot EXE and change the one line of programming code that prompt for CD key. They figuratively change it to don't prompt for CD key. And that has effectively cracked the software. Alternatively, some of these programs will figure out what the algorithms are for generating the keys, and rather that touching you program, it will just generate a CD key for you.

STUDENT: Inaudible

Yes, there is a percentage of people who will certainly figure out how to circumvent these restrictions to that they can make their own legal copies of CDs. There is also a bigger percentage of people that say they are hacking or cracking into software so that they can make legal copies of their disks. That is the sort of goodie-2-shoes argument that is often put forth, honestly, against industry when they say this is appropriate because you always have folks saying this is my legal right to copy data that I have already paid for and make up legitimate backups of. It holds some water, but that's perhaps not strong ground to stand on when it comes to this stuff.

Other questions or concerns?

STUDENT: Inaudible

Yes, Dan has a concern. Come on up.

Is this working? Really? Just use this record? Ok. I will demonstrate this, and if you want to stand near me while I plug this in. Dan has an announcement to make.

DAN: So, of course, we have our usual sections and workshops. Section this week is for how to disinfect your PC. I stole, because you know I am going to come to class and flash in the background and say this is how you do it. And then… No, I'm kidding. We will go over a number of topics such as anti-virus software, safe mode, how to rid your PC of various malicious software, and the like. It is very good especially if you have ever been infected with a virus or a worm. Then in workshop this Saturday… Last year we started digital photography workshop and it was by far the most popular workshop, [Inaudible] it was on the podcast it was more popular than most of his lectures, and so…

DAN: Certainly come by and we will talk about digital cameras. We will talk a little bit about how they work. WE will get into some of the real nitty-gritty. We are going to pack a lot into 2 hours. So, hopefully you will come and bring your questions. And we will have a good time. Yes, bring your cameras if you want and we will have a camera party.

Dan is right, truth be told, that over 8,000 people downloaded his podcasted digital audio workshop last year. So, if that isn't endorsement enough, you've got to be there this weekend. He is our resident photographer.

In conclusion what Dan had also set up for us as we spoke was this screen. I was asked last week if we could give you a sense of what a packet sniffer looked like. Here is what a packet sniffer looks like. This is ethereal, which is a very common program. Its name has recently been changed to Wire Shark. It's freely available. With a program like this, you are connected to a hub. You can sniff all of the traffic going through that hub. If you are on a wireless that does not use any kind of encryption, like WEP or WEPA; similarly, can you sniff all of that wireless traffic. What Dan did while we were talking earlier, was sniff some of the wireless traffic that has not been going on through all the instant messages and e-mails some of you guys have been sending tonight. But just to protect your privacy, he only sniffed my traffic. And whoever else might have been connected to Belcan54G, so, if you connected to Belcan54G, shouldn't have done that since we have all of your traffic here right now. The take-away for tonight though will be, as arcane as all of this looks, you should see things that look familiar in structure. In the left couple of columns, what do you clearly see in columns 2 and 3?

STUDENT: Inaudible

Source and destination IP addresses laid out in source and destination. This is to say, that Dan's program sniffed every packet that left my computer and went to some other server. And I think Chris was connected to this router as well, so all of her traffic was similarly sniffed. In the protocol column, what protocols do you recognize?

STUDENT: Inaudible

DNS, HTTP, HTTP being web traffic. If for instance if I right click on this one, and I say follow TPC stream. What you will see is that one of these guys, be it Chris or Dan or someone else, requested a web page of Google. So, notice that this is essentially the contents of the request they sent to Google. What page did they request? It looks like here they requested the home page. Look at the very top were it says get /, well / obviously refers to just the root (the home page of the site.) One of these guys visited Google. If we look at one other line, we will see here… let's follow this one. This is the same one. Let me go back real quick, and clear the filter. If we scroll down instead to non-HTTP traffic, follow TCP stream... here we get another page. It looks like someone in the room, not only visited Google, they happened, while we were talking, to search Computer Science E-1. This just gives you a taste of what kind of information that even now, as those little green lights flicker on the access point, is flowing back and forth in this room… had we wanted to, for the couple of you who have been using your laptops tonight, it would have been trivial for us (ethically, or unethically) to copy every piece of data that was going from your computer to that access point. The only real protection you might have had is if you were using HTTPS, or you were using a VPN server, or some kind of proxy server. I would wager that most of you were not doing such: in which case one sitting in this room, maybe one of you, could have been sitting here, all this time, watching every instant message, every web page, and every e-mail going across the wireless network.

So, with that said, goodnight.