Week.14.txt

- April 9th, 2021

    - Virtual Private Networks (VPNs)

        - Last class we talk about network layer security, in particular virtual private network give companies or the remote site of the company and their employees that are access from their remote network the kind of illusion that they are actually connecting through their own private network, but doing so physically, it's very difficult. That's why VPN implementation is based on overlay over public Internet.

        - There are 2 basic types.
            - One is site-to-site VPN. Typically, you have IPsec routers that interconnect different sites or headquarters of an organization, and from the host point of view, they are just serving the network as if just without the VPN; it's IPsec router that perform all the difficult job of negotiating the keys and encryption/ decryption, etc.
            - Another form is called remote access VPN, which allows a employee to be able to access company network remotely through the virtual private network, and in this case typically you will need some kind of software to be installed on the person's device

    - IPsec Services

        - So IPsec provides 4 basic services:

            - It enable data integrity

            - It authenticate the origin of messages

            - It have some mechanism to prevent reply attack

            - It provide confidentiality of the message that is exchanged between the remote sites

    - IPsec: Tunneling Mode

        - There are 2 basic modes in IPsec. One is called the tunnel mode. So in tunnel mode we have, we can have 2 settings. We can have in this site-to-site VPN that you have routers that are IPsec enabled and they provide some kind of private network to their respective devices in the, in their own local area network. So when some data is been sent from a host, when it arrives at the IPsec enabled router, we're going to see that the IPsec router would actually put the data that's received from client into the as part of the payload of new IP packet. And in the IP packet it will utilize the, the router's IP

address both of the destination router and the source router's IP address as part of the IP header. It also includes a IPsec header and then in the payload that corresponds to the original IP packet, and this portion (IP header and IP payload) will be encrypted. At the receiving router, it simply take out the payload and also decrypt the message that's been encapsulated in the payload, utilizing the keys that has previously been negotiated, and the forward the data to the respective destination host.

- So the reason its called tunnel is because from, between the IPsec router you can think there is kind of virtual tunnel that's been established and all the packet coming from host will go through this tunnel without, and enjoy the security properties provided by IPsec

- Another setup is like when we access campus network utilizing Mac VPN. So this corresponds to remote access VPN. So here you actually run your as a client, as a host you may run your local IPsec client such as cisco anyconnect. And this cisco anyconnect kind of works as a end point for you to, for your local host to negotiate IP stack keys and perform the data encryption, etc. on behalf of the client. And the IPsec client will actually talk to a IPsec gateway that sits on your, on campus network. So let's say that this part is actually like McMaster campus network, and you are trying to access a service on a particular server on campus. So at the, your Cisco anyconnect gonna talk to IPsec gateway that has been config'd in your client. And between your, between the host, a remote host and the IPsec gateway, you will see there's a secure tunnel that's been established for data exchange between the remote host and a server on campus. So once the data has been received by the IPsec gateway, then the payload will be taken out or be forwarded to the destination host.

- So most cases correspond to the tunnel mode, so this is in contrast to a second way of operating IP sockets called transport mode. Before I explain transport mode, does this sound familiar to you? Have we encountered this notion of tunneling in a previous lecture? Alice mentioned that we utilize tunneling to do IPv6 over IPv4. So you may have multiple sites that actually operating their IPv6 site and you have IPv6 site. But you want to communicate over a public Internet that runs IPv4. So you can do so by utilizing a gateway that can essentially encapsulate IPv6 packets in IPv4 packet, which will be sent through the Internet. The other example, for instance: in a way you can think of network translation have the kind of flavor to that, that the difference in network translation is that you actually change the transport layer header with a port and also change the IP header will not really encapsulating the packet per say.

- IPsec: Transport Mode

- So most cases correspond to the tunnel mode, so this is in contrast to a second way of operating IP sockets called transport

mode.

- So this is different from the case that you actually do not require any infrastructure but just relies on the software that runs on the server and client to provide similar capability. So this is called IPsec transport mode. So here you have a client and server that each run its own, so it kind of you can think of it as if they are acting as their own IPsec gateway. So when they want to communicate with the other entity, then they need to negotiate keys, etc. But the difference is that the, so unlike the tunneling mode, so the data that's being encapsulated is simply the payload of the IP packet and to construct IPsec packet it would include the destination address and the source address of those corresponding hosts. So you do not really change the IP address field because you are actually acting on the behalf as IPsec gateway. And then you include the additional IP header. So if you compare this figure and the previous figure here, the key difference is really whether the, in the outer IP packet, whether you utilize a router IP either as a destination IP address or source IP address, or both. And whether you have actually included the, in the inner, the payload of this outer IP packet, this IP header that some of the original IP packet that you want to transmit. So in the transport mode you just put the payload there, and use the same IP header, but you add IPsec header. So this happens when you actually do not have a like a IPsec enabled gateway that can help you to do such kind of encapsulation.

- So if you compare those two modes, because of the field of the original IP header in the tunnel mode is also encrypted, so attacker will not be able to see at least in the case of site-to-site VPN it won't even be able to see what is the originator and the destination of the IP packet. However, in the transport mode, because IP header remains the same the attacker will be able to see where the data actually originated from or destined to so in that sense that IP tarnsport mode, IPsec transport mode is little less secure than the tunnel mode

- IPsec: Protocols

- So we cannot go through a lot of details in IPsec, but I'm just gonna discuss some high level concept. So if you think about to enable the services we just discussed: data integrity, origin authentication, prevention of replay attack and confidentiality, they're obviously, you should have a mechanism allow you to be able to negotiate the cipher suite, the kind of too setup session keys and message integrity key, etc. So this part is taken care of by a protocol called IKE or IKE2. So this, as a result of this protocol, so before you actually can exchange data between the endpoints, you have to go through this kind of negotiation as a result of the negotation the, some information associated with a particular session will be stored on the IP gateway; it could be in your Cisco anyconnect client

or actually on the router that are IPsec enabled. So once you have such information, this information will be utilized for subsequent communication of actual data packet between the endpoints. You can choose to utilize two protocols for data cleanup operation. You can use something called authentication header protocol that only provide source authentication, data integrity, but not confidentiality. What means is that the data packet themselves, the payload in the IP packet is not encrypted. So but this authentication header protocol provide the ability to be able to authenticate whether the data has been tempered with and will be able to authenticate whether the source is corresponds to that what has previously negotiated through the IKE, and stored in the security associations.

- A more popular one for the data plane operation is the encapsulation security protocol. So this one actually not only provides source authentication, data integrity but also provide data confidentiality. This is more widely used, so if you think about the combination we just discussed, we have 2 modes. One is transport mode, also called host mode. The other is tunnel mode, and here you have 2 data plane protocols. One is authentication header, the other is ESP (Encapsulation Security Protocol). The combination of having tunnel mode with ESP is the most widely used IPsec VPN so in the subsequent discussion we're gonna assume that we use ESP and operate as in tunnel mode

- Security Associations (SAs)

- So we mention that before actually data communication can proceed we need to utilize IKE to negotiate the policy that the cipher suite and also setup the keys. So the information that as a result of that negotiation will be stored in the security associations and this security association will be stored on the IPsec gateway; can be a router or could a client. One thing that you need to be aware is that security associations are simplex. What that means is that it is associated with one direction of communication. If you have bi-directional communication, for instance: you have a TCP connection that want to exchange data between the client and server, then you need to have a security association for each side of the communication. So this means that you, the SAs are actually, it's actually you need to have 2 SAs that between the end host that communicate with one another so it could be like in the case that you have, if you go through IPsec, through some kind of IPsec gateway through site-to-site VPN, then corresponding to each host you're gonna have respective let's say from host A to host B here, you're gonna have 1 SA, and then from host B to host A, the connection coming back would also have another SA. So, but this is network layer tunnel, so you could have multiple say TCP connections in different applications that go through those tunnels; so they all associate depending on which way the communication goes from A to B, it is essentially SA_1. From B to A it is associated with SA_2.

- SA needs to be setup for each direction of a bi-directional connection between like security gateway or between a IPsec enabled client and the security gateway

- Example SA From R1 To R2

- So here is an example of a site-to-site VPN. So in the, in this example that we have IPsec enabled router, R1 and R2, with their respective IP addresses and in this address, the R1 has IP address of 200.168.1.100 and R2 has interface 193.168.2.23. So suppose we want to establish we have, we want to establish a tunnel from R1 to R2 and this direction will be associated with 1 SA, so this is only for this particular direction; so this is why SA is considered simplex. So for in this SA, it gonna store the following information. You gonna have the, there will be identifier that's indexed by the so called security parameter index, so each SA would have a unique SPI, a unique security parameter index. It will also store the origin of the SA interface that corresponds to the IP address of this router interface and it also store the destination SA interface that is IP address of the destination. And then it will store information such as what kind of encryption will be utilized like 3-DES with CBC, or AES. And then it store the encryption key and type of integrity check that's used, for example: using MD5 or SHA-1, SHA-2. And then it also store the authentication key that's used for message integrity. As you can see that are necessary for subsequent encryption and message integrity check of the data that's being exchanged from the end hosts that communicate through those two IPsec gateways. So this is just one side, if you have another, in order to carry data from R2 to R1, for example: you have host that sent some data over then you would have another security association. In that case, it will have a different security parameter index and it will have like the source and destination IP addresses will be swapped and it may or may not associate with the same encryption method and different keys for encryption as well as message integrity. So, yeah the question. This is for site-to-site

- Again all those keys are the result of policy and security key management during the, through the IKE protocol

- Security Association Database (SAD)

- So SA are actually stored in a database on the IPsec gateways. So this database will contain all the SAs that has been utilized for different connections, for instance: if you think about you have a situation that you have two sites like one headquarter and one branch office that use site-to-site VPN, then there should be one SA for each direction of the communication and then for in the case that you have remote access VPN that you have multiple salesperson that try to access the headquarter network through the VPN, then if

you have 'n' salesperson at different locations for each side of the connection that you have for salesperson you have one SA, and so for bi-directional connections that you gonna have '2 * n' SAs that's been stored in R1's SAD. So this database will be looked up upon reception of a message at the router, so the router will look up the SAD and determine by looking at the information that's included in the IPsec header send by the, from the client, then it will be able to determine what kind of, what the corresponding security association is, and then use such, use the key that's stored in security association for encrypting and to ensure the integrity of the particular message.

- IPsec Datagram

  - So this is what IPsec datagram looks like. Again here we talk about the tunnel mode with ESP. So this ESP we actually would not only ensure message integrity but also wanna ensure the data confidentiality. That means that the payload will be encrypted. With tunnel mode, that means that I would actually not use the old IP header. Instead, I will actually include new IP header that may be associated with the IP address of the IPsec gateway as source or destination IP address, and put the original IP packet in a encrypted way as a payload of that new IP packet. So let's look at the specific structure of the IPsec datagram in this setting.

  - So we, so this whole thing is, we call it, this IPsec datagram, we kind of as an analgous to enchilada, so it's actually something with some kind of wrapping. So in this datagram you're gonna have a new IP header and you will followed by ESP header and then you have the original IP packet that's sent by the end host. So this part will include original IP header and original IP datagram payload, and you would add some called a ESP trailer at the end. And this whole thing will be encrypted and you, and after that include something called the ESP authentication code. So this allows, this will protect everything from the ESP hdr to the, as well as encrypted IP, original IP packet and ESP trailer, and this (ESP authentication code) will serve the purpose of integrity check. So we want to make sure that this whole part will not be tampered with, so the attacker will not be able to modify any of the contents here.

- R1: Convert Original Datagram To IPsec Datagram

  - So let's look at how this IP packet can be constructed and step-by-step. So, again, we are in a situation that we have, we use IPsec tunnel mode and the router will receive regular IP packet from the end host. So this IP packet coming from one of the host will simply have your original IP header and original IP payload that could be, say TCP segment. So the router, the IP enabled router will first add a trailer at the end of this IP packet. So this trailer would include information like typically add some padding, some additional information that'll make up the content that needs to be encrypted. So

the encryption will be done utilizing the algorithm and the key that's specified by the SA. So this here we already assumed that the SA has been set up and stored on this router, R1. So next the router would append the ESP header to the, infront of the encrypted IP packet, and then it will create a authentication Mac code message, authentication code after the ESP trailer. So this, again, utilize algorithm and key that's specified in the SA. And finally, it will add a new IP header like in this case it may use the IP address of R1 as the source address of this IP packet, and the destination router R2 as a destination IP address. If this actually tried to reach remote access host, then the destination IP address would be the IP address of the host.

- Inside The Enchilada

  - So here's a little more detail about each of the fields we just mentioned. For ESP header it include the SPI; this is a index that's used to look up the security association at the router. It also include a sequence number. So this one is used for the, the sequence number will be incremented for every single subsequent IPsec message that's exchanged for this connection. So this is useful to prevent replay attack. And in the trailer field it will include some kind of padding information on the lens of the pad; how many beds you include in the padding? And some information about the next header.

- IPsec Sequence Numbers

  - So the sequence number will be initially for new SA, it will be initiated to zero to start with. But for subsequent datagram that sent on this direction of the connection, it will be incremented and this information will be included in the header, in the ESP header. So this allows the ability to prevent replay attack or you know actually packet sniffing is not particularly harmful in this case, because it's actually already encrypted.

  - At the receiver side if it see some kind of duplicate sequence number it simply discard such kind of datagram. So if you think about this from end-to-end point-of-view, so just think about we have this figure here, so if you have in this site-to-site VPN, so you have like the a SA that's been setup from router 1 to the router 2, then all the message that go through this tunnel will have, will be associated with the same SA, but you're gonna have the different datagram may have its own different sequence numbers. So think about let's say your TCP is performing some retransmission from TCP point-of-view, TCP will maintain its own sequence number for the purpose reliable data, but from the IPsec perspective at the network layer, this retransmission is actually different IP datagram, so that will be associated with a different sequence number

  - Because at the network layer it doesn't really

differentirate whether this is original TCP segment or retransmission of TCP previous segment

- Security Policy Database (SPD)

  - The last piece in IPsec VPN is something called the security policy database (SPD). So unlike the SAD, the security association database, that store the collection of the security associations which actually serve the purpose of when you actually have the, so the SAD actually stores the security association that stores the type of keys that you use and what kind of cipher utilized. In SPD that actually will store the information that's associate with what to do with individual datagram. So this actually is, actually allows the secure IPsec gateway to determine which SA it will use. So in a way you can think of SA tells you what to do, what kind of key to use, what kind of cipher to use, but the security policy database is kind of, it help you to determine which SA to utilize. This is a little bit, the concept is a little bit like for firewalls that in firewalls we discussed previously that you have some policy, some rules to say, "ok if the source IP address is such and such, and the port number is such and such, then I may decide whether I wanna block this message or allow this message to go through". So this SPD is kind of store information like for what kind of action you need to do, and what kind of security association you may utilize for the particular IP datagram, that's IPsec gateway receives.

  - So when you combine the SPD and SAD, then you will be able to fully determine what you do with arriving datagram.

  - That's a very high level overview, and definitely do not have sufficient details

- Q/A
  - Question: What does confidentiality mean?
    - Answer: It means that your data will not be, it's only visible to the communicating party but not too third party. So when, the main mechanism to ensure confidentiality is to encrypt your data