EP2500 Networked Systems Security
Homework 1 Problem Set

Total: 200 points. Required to pass: 110 points.

November 17, 2015

Deadline: 23:59 (UTC +1), December 1, 2015

Please send your solutions by e-mail to papadim@kth.se in PDF format from ONE of the team members. Please include in the PDF your names and a brief (5 lines max.) explanation of the contributions of each team member, as you see fit. Subject line of the email: `[NSS 2015 HW1] <your name 1, your name 2>`

Contents

1 Basic Cryptographic Primitives and Protocols

Exercise 1 Key establishment (20 pt.)

Alice and Bob want to establish a two-way secure channel. Please answer the following two questions for the case that (a) they use symmetric key cryptography, and (b) they use asymmetric key cryptography:

1. How many keys do they need?
2. Who needs to know what key?
3. Propose a simple key transport protocol (with a one-way transmission from Alice to Bob), such that Alice can leverage Bob's public key and provide him a new shared symmetric key and authenticate herself to Bob. Assume a sequence number S_i that they both remember from their previous key refresh and that Bob already knows Alice's public key.

Answer of exercise 1

1. They need 1 symmetric key. Both need to know the key.
2. At least two pairs. Assume K_a and K_b are the private keys, and PK_a and PK_b the public keys. Then Alice needs to know K_a, PK_a and PK_b; and Bob needs to know K_b, PK_b and PK_a.
3. $E_{PK_b}\{K_{AB}\}, S_{i+1}, Sig_{K_a}\{S_{i+1}, K_{AB}\}$

Exercise 2 XOR encryption (10pt.)

Alice and Bob wish to exchange 2 messages, M_1 and M_2. Assume that Alice wishes to send M_1 to Bob, and Bob responds with M_2. They encrypt the messages using the XOR operation, using a key of the same length with the message. The encrypted messages are sent over the channel:

$$C_1 = M_1 \oplus K_1$$
$$C_2 = M_2 \oplus K_2$$

Unfortunately, Bob forgets to use the second key K_2 after the first transmission and reuses K_1 ($K_1 == K_2$). Assume Eve is listening the channel and reads:

$$C_1 = 001101000100110001101111$$
$$C_2 = 001010010100110001111000$$

Assume that Eve is able to understand the content of M_1, e.g., she guessed that it is the binary representation of the ASCII characters **net**. Write the binary and ASCII representation of M_1, M_2, and K_1 and explain how Eve can obtain M_2 and K_1 (to convert ASCII text to binary you can use the following site: http://www.roubaixinteractive.com/PlayGround/Binary_Conversion/Binary_To_Text.asp).

Answer of exercise 2

$$M_1 = 011011100110010101110100(net)$$
$$M_2 = 011100110110010101100011(sec)$$
$$K = 010110100010100100011011$$

Exercise 3 Symmetric Key (15 pt.)

Consider the following protocol (Figure 1) which Alice and Bob use in order to *mutually authenticate* each other, i.e., convince each other that "they are who they say they are". Assume that Alice and Bob share a secret key K.

In this protocol, Alice first sends an unpredictable random number R_A. In the second step, Bob encrypts this message to prove knowledge of the key K and also sends a random number R_B. In the third step, Alice decrypts $E_K(R_A)$. If the result is not her original number she aborts the protocol otherwise she encrypts R_B and sends it to Bob. Bob performs a similar check and if everything is ok, he's convinced he's talking to Alice. Find two attacks in which an attacker can impersonate some of them to the other.

(*Assume that the key is not compromised, so nobody can use it to create fake messages.*)

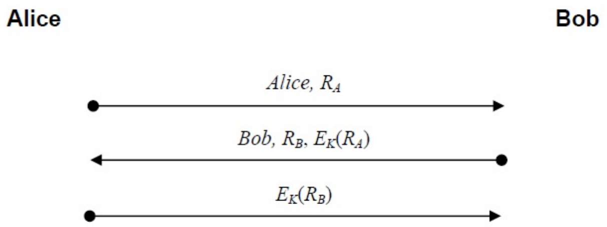


Figure 1: Mutual Authentication protocol

Answer of exercise 3

- When Eve received the challenge R_A from Alice she has to encrypt and send it back. She starts another session with Bob and pretend she is Alice to him by forwarding the same challenge R_A to him. Bob will respond back with $E_K(R_A)$ which she then forwards to Alice. Then she drops the session with Bob and continues the protocol with Alice.
- When Eve receives R_A from Alice, instead of opening a new connection with Bob, she reflects the challenge back to Alice. She just opens a new connection with Alice and sends the challenge to her.

Exercise 4 Block Cipher (25pt.)

Consider the following encryption algorithm that operates on 128-bit keys and messages of size 64 bits. F is another secure block cipher such as AES that encrypts 128-bit messages.

```
$E_K(m)$                          → $K$ is the key, $m$ is a message of 64 bits
    Set $R$ equal to a random 64-bit string
    Set $C = F_K(R||m)$           → where "||" denotes concatenation
return(C)
```

a) Explain how the recipient of an encrypted message c, can recover the original message.

b) Assume that the attacker (Eve) has the capability to choose arbitrary plaintexts and obtain the corresponding ciphertexts. This is done through an *oracle* where the adversary can submit a number of encryption queries. For instance, Eve sends two messages m_0 and m_1 of her choice to the oracle. The oracle answers back with the encryption c of one of them. The goal of Eve is to discover which of the two messages was encrypted. Such an attack is called *chosen-plaintext attack* (CPA).
In our context, each query is a request for encrypting one pair of messages. By sending q pairs of messages, the oracle will respond with the encryption of the first message of all pairs or the second message of all the pairs.
Based on the response see if the attacker can infer something about the algorithm. The number of queries should be large but carefully selected so that some advantage is gained. Be careful not to select your queries based on previous answers. The queries cannot be adaptive in the CPA game...You just construct a set of q pairs of messages, you give it to the oracle and the oracle decides to encrypt either the first message of all the pairs or the second one.
(*You may find useful the following fact (also known as birth paradox): the probability* of a collision when throwing q balls into b buckets is approximately $q^2/2b$)

c) Is the scheme secure?

Answer of exercise 4

a) $\mathsf{Decrypt}_K(c) = F_K^{-1}(c) \bigoplus R$

b) Our strategy begins by sending the oracle the following pairs: $(1, 0), (2, 0), (3, 0), ..., (q, 0)$, i.e., the first message of each pair is some 64-bit number i in binary and the second message is always 0. Based on the CPA game, the oracle will encrypt either the different i's (case 1) or just the 0's (case 2). To encrypt a message m the oracle uses the expression $AES_K(R||m)$, where R is some random 64-bit number. This is what happens in the two cases:

- Case 1: The oracle produces the ciphertexts $c_i = AES_K(R_i||i)$, for $i = 1, 2, 3, ..., q$. Since $i \neq j$, we get $R_i||i \neq R_j||j$, no matter what R_i and R_j are. Therefore $AES_K(R_i||i) \neq AES_K(R_j||j)$ for all $i \neq j$. In other words, in this case $c_i \neq c_j$ for all $i \neq j$.
- Case 2: The oracle produces the ciphertexts $C_i = AES_K(R_i||0)$, for $i = 1, 2, 3, ..., q$. Clearly $c_i = c_i$ if and only if $R_i = R_j$. Thus two ciphertexts will be the same only if the oracle picks the same two random numbers to encrypt the sequence of zeros. If we make sure that the number of queries q is large enough, this can happen with high probability. In fact, based on the birthday paradox, the probability is proportional to $q^2/2^{65}$. Thus if we pick the number of queries q equal to 2^{33}, then this probability is quite large.

To conclude, we send to the oracle $q = 2^{33}$ queries of the previous form. If no collision is observed, we bet on case 1. If a collision is detected we bet on case 2.

c) The scheme is not secure.

2 Physical Layer security

Exercise 5 Jamming (20 pt.)

Consider a sender A and a receiver B such that A can transmit messages to B over any of $C = 10$ available wireless channels. One message can be transmitted within $T = 1$ sec. The adversary has rather limited capabilities and jams $C_{jam} < C$ out of the C channels.

i. First, assume that the jamming channels are fixed throughout the attack. In fact, $C_{jam} = 4$ channels are chosen randomly at the beginning and kept throughout the attack. Moreover, assume that A pseudo-randomly chooses a channel among the C available ones, giving each of the channels the same probability. The transmitter sends its message sequences without changing the transmission channel. This could be because it is not aware of the presence of a jammer.

   What is the probability that a transmission that lasts 5 seconds is unjammed? What can you say about the probability that at least 60% of a transmission lasting 5 sec is jammed?

ii. Now assume that the jammer chooses a new set of C_{jam} channels every T_{jam} seconds. Every new channel set is randomly chosen and independent of the previous choices. Every channel set with C_{jam} elements has the same probability of being chosen. Also assume that the transmitter chooses a new channel for transmission every $T = 1$ sec. This choice is random, independent over time and from the jammer's choice, and every channel has the same probability of being chosen by the transmitter.

   (a) If $T_{jam} = 1$ sec, what is the probability that a transmission that lasts 5 seconds is unjammed? What is the probability that at least 60% of a transmission lasting 5 sec is jammed?

   (b) If $T_{jam} = 2$ sec, should the sender randomly change its channel every T sec or every $2T$ sec?

iii. Assume that after every time slot of duration $T = 1$ sec, the transmitter obtains feedback from the receiver whether the message it tried to transmit in this time slot arrived successfully or whether it was jammed. If the message was jammed, the transmitter chooses one of the remaining $C - 1$ channels with equal probability. The jammer also chooses a new set of C_{jam} channels to be jammed, independently from the previous jammed channel set and from the sender's choice, and giving equal probability to every channel set with C_{jam} elements. What is the probability that two messages are successfully transmitted within 2 sec?

Answer of exercise 5

i) The probability that one message is unjammed is

$$1 - \frac{C_{jam}}{C} = \frac{2}{5}.$$

As the sender does not change the channel through which it transmits and jammer the does not change the channel set it jams, the probability that a transmission that lasts 5 sec is unjammed equals 2/5 as well. Thus the probability that 60% of a transmission lasting 5 sec is jammed equals 3/5, as either 100% of the transmission are jammed (with probability 3/5) or 0% are jammed, which happens with probability 2/5.

ii) a) Now, as the choices are independent over time, the probability of 5 unjammed transmission in a row equals

$$\left(1 - \frac{C_{jam}}{C}\right)^5.$$

That at least 60% of a transmission lasting 5 sec, i. e. 3 out of 5 attempts of transmitting a single message, are jammed, has a probability of

$$\begin{aligned}
&\sum_{k=3}^{5} \binom{5}{k} \left(\frac{C_{jam}}{C}\right)^k \left(1 - \frac{C_{jam}}{C}\right)^{5-k} \\
&= \binom{5}{3} \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right)^2 + \binom{5}{4} \left(\frac{2}{5}\right)^4 \frac{3}{5} + \binom{5}{5} \left(\frac{2}{5}\right)^5 \\
&= \frac{992}{3125} \approx 0.3174.
\end{aligned}$$

b) If the transmitter changes the channel every $T = 1$ sec, then

$$\begin{aligned}
P_T[S_2 = 0] = P_T[\text{no success over 2 seconds}] &= \left(\frac{C_{jam}}{C}\right)^2, \\
P_T[S_2 = 1] = P_T[\text{exactly one success over 2 seconds}] &= 2\frac{C_{jam}}{C}\left(1 - \frac{C_{jam}}{C}\right), \\
P_T[S_2 = 2] = P_T[\text{2 successes over 2 seconds}] &= \left(1 - \frac{C_{jam}}{C}\right)^2.
\end{aligned}$$

If the transmitter changes the channel every $2T$ seconds, then

$$\begin{aligned}
P_T[S_2 = 0] = P_{2T}[\text{no success over 2 seconds}] &= \frac{C_{jam}}{C}, \\
P_T[S_2 = 1] = P_{2T}[\text{exactly one success over 2 seconds}] &= 0, \\
P_T[S_2 = 2] = P_{2T}[\text{two successes over 2 seconds}] &= 1 - \frac{C_{jam}}{C}.
\end{aligned}$$

On average, within two attempts, there are thus

$$0 \cdot P_T[S = 0] + 1 \cdot P_T[S = 1] + 2 \cdot P_T[S = 2] = 2\left(1 - \frac{C_{jam}}{C}\right)$$

successes if the transmitter changes its channel every second. If it changes its channel every 2 seconds, the average number of successes within two attempts equals

$$0 \cdot P_{2T}[S = 0] + 1 \cdot P_{2T}[S = 1] + 2 \cdot P_{2T}[S = 2] = 2\left(1 - \frac{C_{jam}}{C}\right).$$

Thus both methods have the same average throughput over two transmission attempts. As blocks of two subsequent transmission attempts are independent of each other, the average throughput over $2n$ attempts is just the sum over n different 2-second attempts, which in both cases equals

$$2n\left(1 - \frac{C_{jam}}{C}\right).$$

iii) If there is no success in the first step, then there cannot be two transmission successes. The probability of success in the first step equals

$$1 - \frac{C_{jam}}{C}.$$

The second transmission attempt is successful as well if the jammer does not choose a set containing the channel chosen by the sender for the first attempt. Because of the independence of the jammer's choices over time, this probability also equals $1 - C_{jam}/C$. Hence the probability of two successes equals

$$\left(1 - \frac{C_{jam}}{C}\right)^2.$$

3 Denial of Service (DoS)

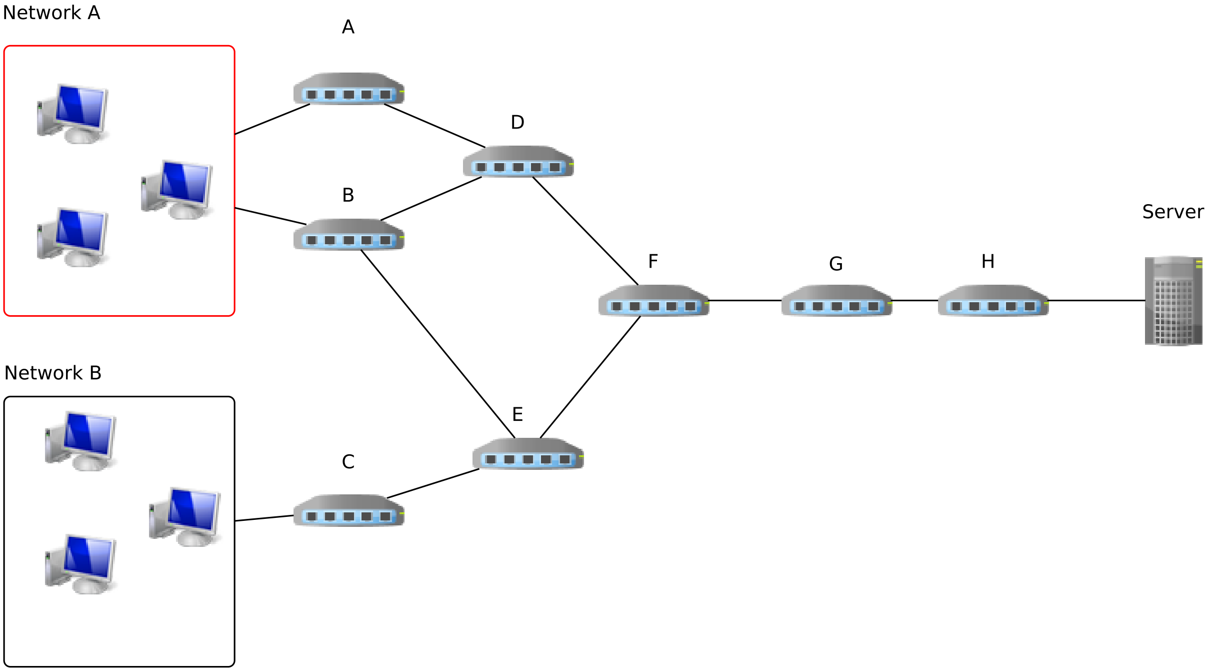


Figure 2: Network topology.

Exercise 6 Flooding (20 pt.)

The topology presented in Figure 2 is composed of two networks (A and B), routers (A,B,C,D,E,F,G,H) and a server (Server). Network A contains some zombies (bots) in addition to a large number of legitimate hosts. The bots send SYN messages to the Server in an effort to exhaust its resources. Based on your knowledge and experience, do you agree or disagree with the following statements? Please elaborate on the effectiveness or the inappropriateness of the following suggested countermeasures.

- As we intent to serve only legitimate requests (Network B), we should allocate buffers with higher capacity for requests originating from Network B compared to the buffer size that is used to serve requests originating from network A. This way most of the server resources will be used to serve legitimate requests.
- Set up a Firewall in front of the Server that can examine the source address of each packet and drop packets with IPs belonging to Network A. This way no malicious traffic coming from Network A will make it to the Server and as a result all the resources of the Server will be used to serve legitimate requests.
- Have border routers (A,B) block all TCP packets with the SYN flag set to "1".
- We can use packet marking mechanisms so that each packet puts a marking as it forwards packets. For example, a packet that follows a Path A→D→F→G→H→Server will receive a marking (A,D,F,G,H). Packets that follow the path B→D→F→G→H→Server will receive the marking

(B,D,F,G,H). Assume that the available header space suffices for only three marking. This means that if router D receives a packet with a marking (A,B,C) and decides to mark it, its marking will be (D,B,C). Based on this marking scheme we can identify traffic that originates from Network A.

Answer of exercise 6

- This will not help. Due to the fact that the *source address* is not authenticated, *IP spoofing* will render this suggestion ineffective. The Bots of Network A can still spoof their IPs to match addresses of Network B and as a result still manage to overwhelm the buffers available for serving Network B.
- The problem behind this suggestion is the same like the one in the previous question. Bots can spoof their IP addresses to match the ones of Network B. As a result, the firewall will inspect the *source address* field of the IP header, see that it carries an IP of Network B and relay it to the server.
- This is not a feasible solution. Many of the students answered that this way we manage to block all traffic originating from legitimate hosts of Network A. This is a totally valid point. In addition, the end-to-end principle [1] dictates that routers are used only for packet switching purposes. As a result, a router is aware only of the IP layer and is agnostic of the higher level application protocols that might be in place. As a result, an inspection of the TCP flags cannot be performed by a router.
- Here we have deterministic packet marking. This means that all routers along the path will mark packets as they route them to the next-hop router. Since we have space only for three markings, all packets will be marked with (F,G,H) markings. This way, the server cannot use the marking in order to differentiate between traffic originating from Network A or B. Of course, a probabilistic marking scheme would me much more useful.

[1] http://en.wikipedia.org/wiki/End-to-end_principle

4 Unauthorized Access

Password Strength can be expressed in terms of entropy. Entropy is a measure of disorder and unpredictability in bits. This means that the higher the entropy is, the stronger the password is. The mathematical formula to compute entropy is $H = L \cdot \log_2 N$, where L is the length of the password and N is the size of the alphabet used (different characters).

Exercise 7 Password Entropy (20pt.)

a) What is the entropy per symbol of alphanumeric characters (symbols form 0-9 and a-z including capital letters)?

b) Compute the entropy of a password that consists of 8 random symbols chosen from the extended ASCII table (assume an ASCII table with 256 characters available).

c) What is the difference between a dictionary and a brute-force attack? Be brief (5 lines max.).

d) Why isn't entropy enough to guarantee security against dictionary attacks? Give a simple example

e) How many combinations approximately would it require to brute force the entire key space of a password that consists of 16 random symbols chosen from the ASCII table (ASCII table has a total of 256 symbols)

f) How much time would a machine performing 100.000 computations per second need to to crack the entire key space of a password that consists of 6 random symbols chosen from a pool of 16 different choices? What is the expected time for an adversary to succeed in the brute-force attack? Compute the same for the previous question

Answer of exercise 7

a) 62 symbols in total. Entropy is 5.95 according to the formula

b) Entropy is 64

c) In a brute-force attack the adversary tries every possible combination to discover a password. In a dictionary attack the adversary uses a comprehensive list of commonly used passwords, such as "password"

d) Increased password entropy means additional effort needed by an adversary to brute-force the password. However, increased entropy does not necessarily mean increased security against dictionary attacks. Consider the case of passwords similar to "1111111111" or "abcdefg123456789" that have high entropy. Similar passwords are highly vulnerable to dictionary attacks

e) $256^{16} = 3.4 \cdot 10^{38}$ combinations approximately

f) $16^6 \cdot 10^{-5} = 167.77$ seconds or a little bit less than 3 minutes. And $3.4 \cdot 10^{38} \cdot 10^{-5} = 6.4687976 \cdot 10^{27}$ years. The expected time is half in both cases

Exercise 8 Passwords Storage (10pt.)

a) Is it secure to store passwords in clear-text in databases?

b) Is it secure to store them hashed in a database?

c) What else would you suggest for storing passwords securely?

Answer of exercise 8

a) No. If the database is compromised then so are the passwords. Moreover, why would you completely trust a database administrator?

b) No, it is also not secure to store hashed passwords in the database. Rainbow tables are very useful to recover the original passwords. Moreover if passwords are only hashed, then same passwords (and obviously weak) used by users will result in the same hash, which eventually exposes them all to the adversary.

c) The solution is to use a salt value. A salt consists of a few random bytes concatenated to the password before it is hashed (more flavors exist). Each user should be associated to a different salt, generated at the time of the password creation. Salt values can be stored on the server , on a different machine than the actual password database.

5 Secure Routing

Exercise 9 Secure Routing Theory (15pt.)

1. Why do you think Border Gateway Protocol (BGP) routers have to frequently send updates of their routing tables to their neighbors (normally every 30 seconds)? (Hint: Consider false route advertisement)

2. Do you think that by increasing the frequency of updates, the impact of false route or link advertisement could be prevented? (Beyond BGP, consider link state advertisements too)

3. What is the average cost when using Routing Information Protocol (RIP), to connect two routers which are 20 hops away assuming that there are two paths from the source and each link has a cost of 2?

4. If each BGP router issues its own private-public key pair, then they can authenticate themselves securely to every other BGP router. Do you agree? Why?

5. How would you explain the count-to-infinity problem of the RIP protocol in no more than two short sentences?

6. Why BGP doesn't suffer from the count-to-infinity problem?

Answer of exercise 9

1. The links between the routers are dynamic and change over time. Thus the routers have to be up-to-date with any changes in the topology. Routers could send out updates every 5, but that would mean that the topology would be changing faster than the updates. As a result false information would be kept in the routing tables.

2. No. If the attacker has compromised or introduced a BGP router, he can also keep advertising the false routes at a higher frequency.

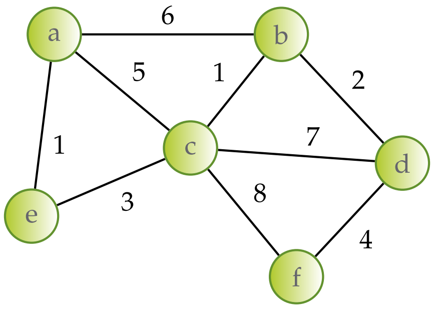
3. RIP does not support routes more than 16 hops away (16 denotes unreachable).

4. No, unless there is a trusted third party such as a Public Key Infrastructure (PKI). The router's public-key need to be certified otherwise the attacker could pretend to be a router and still be able to send digitally signed messages.

5. In RIP routing loops may occur. If router B tells router A that it has a path to another router, there is no way for A to know if it is part of that path.

6. In BGP when a router receives routing information from a neighbour, it also receives the whole path to the destination that the neighbour uses. This way a router can understand if a path makes sense, by checking if it is part of it.

Exercise 10 Routing Information Protocol (RIP) (25 pt.)



The figure shows an example of a network where the routers use Routing Information Protocol (RIP) to build the routing tables.

The original version of RIP has no built-in authentication, and the information provided in a RIP packet is often used without verification. The version 2 of RIP was enhanced with a simple password authentication algorithm, which makes RIP attacks harder to happen.

Assume that the network nodes use the RIP protocol version 2 and they all have the same pre-shared key.

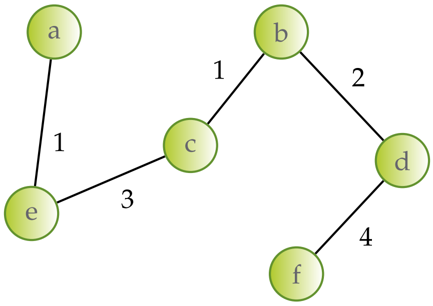
1. Could the node **a** attract the traffic (or part of it) intended for **f**? If yes, how and how many other nodes would be affected? If not, why? (5 pt.)

2. How would you enhance the RIP protocol to provide a (more) *secure* and *scalable* schema for distributing the updates?

   a) Describe the steps of a new version of the protocol which still uses a simple password authentication algorithm, but with *different keys* for each router. (10 pt.)

   b) Describe the steps of a new version of the protocol that combines *Public Key Infrastructure (PKI)* (assume a PKI is available) and *hash-chains*. (10 pt.)

Answer of exercise 10

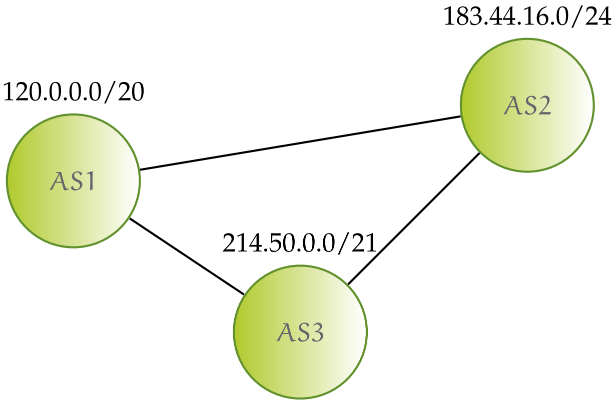


1. Yes. Since they all have the same pre-shared key, node *a* can advertise a shorter path to node *f*, thus attracting the traffic, intended for node *f*.

2.a) The two end points can authenticate the updates using a simple password authentication scheme. They can authenticate the message by calculating the hash of the message and the password (H(msg || pwd)). But, it is not scalable.

2.b) Assuming a PKI in place, one can sign the updates using its private key corresponding to its certificate and attach it to facilitate the verification on the recipient side. Using the hash chain for authentication is more efficient, but it cannot be implemented alone. The hash chain anchor needs to be signed.

Exercise 11 BGP (20pt.)



Assume the network described in the figure above. The BGP router of Autonomous System (AS)-2 gets compromised. Consider that an important aspect of BGP is the updates sent by the routers. According to what has been covered in the lectures, explain:

1. How could the attacker use the updates to attract traffic originating from AS-1 and heading to AS-3?
2. How can the system be secured? Explain your answers in brief.

Answer of exercise 11

1. Prefix hijacking is an attack that can be used to attract all traffic from AS1 with AS3 as their destination. In this type of attack, AS2 would have to advertise a more specific range of addresses that AS3 would normally advertise.

   The reason comes from the routing classless routing algorithms and it is called longest prefix matching. Therefore, the adversary could advertise 214.50.0.0/24. By forwarding the packets to the destination (AS3), it would be hard to understand that something is wrong.

2. As in Secure BGP (S-BGP), Address Attestations (AA) and Route Attestations (RA) can solve this issue. Using AAs, each router can only advertise the IP range it is assigned. This is achieved using a PKI structure, and with the Certificate Authority (CA) issuing certificates containing among other parameters the public key of the router and the IP range it can advertise. On the other hand, RAs are added in the BGP updates to certify that a neighboring AS can advertise IPs (propagate the update), belonging to the issuer of the update.