# Final Exam

Network Security Autumn 2017

8 February 2018

**Surname**, Given Names (*e.g.*, Turing, Alan Mathison): _____

Student Identification Number (*e.g.*, 15-123-456): _____

Student Signature: _____

## Rules and guidelines:

- Place your identification card on your desk. An assistant will check your identity during the exam.
- Once the exam starts, make sure you have received **all** pages of the exam. The exam should have **16 pages total**, including a page for extra space. **Do not** separate the exam sheets.
- Do not forget to fill in your **name, student identification number and signature** on this page.
- You **must** answer questions using **black or blue ink**. Illegible answers may not get any credit.
- The use of notes, textbooks or other written materials is **not** allowed. You are allowed to use a **scientific calculator** during the exam. Any other device that provides communication or document storage capabilities is **not** allowed (this includes smart watches).
- You have **90 minutes** to complete this exam.
- As a general guideline, one point should correspond to one minute. Thus you should write answers that are **clear and concise**. Generally, you do not need to completely fill the space provided for solutions.
- You are **not** required to score all points to get the maximum grade.
- When answering questions, always **explain your reasoning**. If a question asks, for instance, whether A is more secure than B, a plain "yes" or "no" answer will not be awarded any points.
- For questions during the exam, **raise your hand** and an assistant will come to answer your question.
- If you need extra space to answer a question, use the page provided at the back of the exam.
- At the end of the exam, please **remain seated** while we collect the exams. You may hand in your exam before the end, except in the last 10 minutes of the exam. Please **hand in all exam sheets**: if any sheet is missing, the examination will be marked with grade 1.0 and counts as failed.

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|---|---|---|---|---|---|---|---|---|
| Points: | 9 | 10 | 7 | 16 | 13 | 15 | 11 | 9 | 90 |
| Score: | | | | | | | | | |

# 1. TLS, PKIs, Certificate Transparency (9 points)

(a) (3 points) An attacker is trying to attack the company Wahoo and its users. Assume that users always visit Wahoo's website with an HTTPS connection, using fixed Diffie-Hellman and AES encryption. (You may assume that Wahoo does not use certificate pinning.)

For each of the following attack scenarios, mark *all* of the options that an attacker could achieve in that attack scenario. (Partially correct answers and incorrect answers get 0 points.)

i. (1 point) If the attacker obtains a copy of Wahoo's certificate, the attacker could:

- ☐ impersonate the Wahoo web server to a user
- ☐ discover some of the plaintext of data sent during a past connection between a user and Wahoo's website
- ☐ discover all of the plaintext of data sent during a past connection between a user and Wahoo's website
- ☐ successfully replay data that a user previously sent to the Wahoo server over a prior HTTPS connection
- ☐ none of the above

ii. (1 point) If the attacker obtains the private key of a certificate authority trusted by users of Wahoo, the attacker could:

- ☐ impersonate the Wahoo web server to a user
- ☐ discover some of the plaintext of data sent during a past connection between a user and Wahoo's website
- ☐ discover all of the plaintext of data sent during a past connection between a user and Wahoo's website
- ☐ successfully replay data that a user previously sent to the Wahoo server over a prior HTTPS connection
- ☐ none of the above

iii. (1 point) Suppose the attacker obtains the private key that was used by Wahoo's server during a past connection between a user and Wahoo's server, but not the current private key. Also, assume that the user has learned that the certificate corresponding to the old private key has been revoked and is no longer valid. This attacker could:

- ☐ impersonate the Wahoo web server to the user
- ☐ discover all of the plaintext of data sent during a current connection (one where the current private key is used) between the user and Wahoo's website
- ☐ discover all of the plaintext of data sent during a past connection (one where the old private key was used) between the user and Wahoo's website
- ☐ none of the above

(b) (2 points) For server efficiency reasons, many banks prefer not to set up TLS connections for the page containing username/password entry, but instead create a non-TLS page that contains Javascript code that encrypts the username/password before sending it to the server. Provide a brief argument on whether this approach is secure or not.

(c) (4 points) Recently, a study has shown that many governments are forcing Certificate Authorities to issue them forged certificates which will enable them to carry out man-in-the-middle attacks against TLS-secured web sites. Consider the following hypothetical example of such an attack:

The CIA wants to monitor the activities carried out by the customers of Commercial Bank of Dubai (whose servers reside in Dubai). The CIA forces VeriSign to generate a valid certificate for the name "Commercial Bank of Dubai" (whose actual certificate is issued by Etisalat, UAE). Now, the CIA can perform a man-in-the-middle attack against all the US users who access the Commercial Bank of Dubai from within US. This attack is undetectable by current browsers as VeriSign is already a trusted CA.

This example shows what is known as the *compelled certificate creation attack*. Answer the following questions about this attack:

i. (2 points) What is the risk taken by the government (or by the CA that issues the second valid certificate) when they try to carry out such an attack against users?

ii. (2 points) Can Certificate Transparency be used to make such an attack publicly detectable? If yes, under what circumstances? If no, why not?

## 2. DoS and Botnets (10 points)

Consider an attacker that wants to perform a DDoS attack on a victim server. The attacker has created a botnet with 10,000,000 compromised IoT devices but is unable to spoof source addresses.

(a) (1 point) The attacker uses the botnet devices to directly flood the victim server. If each IoT bot device simply sends 10 TCP SYN packets per second towards the victim, what is the approximate aggregate bandwidth that reaches the victim? Give your answer in the unit of bits per second.

(b) (3 points) Consider a scenario where the victim server uses an address filtering approach. Botnet nodes are blacklisted (based on their regular connection requests), and the associated packets are rejected. What is an advantage of this approach and a limitation?

(c) (3 points) Consider a network device that can store up to 10,000 address filtering rules. How could this device be used to mitigate the attack described in part (a)? Describe a disadvantage of the suggested approach?

(d) (3 points) Now assume that botnets are capable of spoofing addresses. Consider a service which publishes a list of misbehaving hosts, categorized as those performing 20 consecutive SYN requests to one of the servers in the network controlled by the service. Any server may periodically download the blacklist and refuse connections by listed hosts. Describe two disadvantages of this approach.

## 3. Intrusion Detection, Firewalls and Evasion (7 points)

(a) (2 points) Next generation firewalls utilize application and protocol semantics to filter malicious traffic. Describe a possible negative impact of such devices being pervasively deployed in the Internet.

(b) (3 points) Consider an IDS that uses a hash table to store flow identifiers. The hash table is implemented with an MD5 cryptographic hash function applied on the flow's 5-tuple which returns an array index, for example (where % represents the modulo operator):

```
index = md5(src_ip, src_port, dest_ip, dest_port, protocol) % array_size
```

Each entry of the array contains a linked list of pointers to handle colliding entries.

i. (2 points) How could an attacker exploit this data structure to perform a DoS attack on the IDS?

ii. (1 point) How can this attack be mitigated?

(c) (2 points) Your company currently deploys an IPS system on the office network. Some of employees of your company have requested to be allowed to connect to their home networks via VPN from the office. Why is this not safe to allow? Describe a mitigation for the perceived risk.

# 4. Anonymous Communication Systems (16 points)

(a) (4 points) **Tor: compromised directory.** Directory authorities are very powerful entities in Tor, since they regularly publish the list of all Tor relays and their status. The authorities work by a consensus algorithm such that a majority of directory authorities can change the list that is being published. Assume that an adversary is able to compromise 6 out of the 10 directory authorities in Tor:

i. (2 points) Describe one way in which this adversary can de-anonymize a large fraction of the Tor network.

ii. (2 points) Which entity (or entities) may detect the attack you described? (Motivate your answer.)

(b) (4 points) **Tor: compromised website.** With reference to the previous question, consider now a different adversary which is able to obtain the private key for the certificate of `torproject.org`, the website used to distribute the Tor client (browser) and all the updates to it. Describe (at least) one advantage and one disadvantage of this adversary's capabilities compared to the previous adversary which could compromise 6 out of 10 directory authorities. Which adversary is stronger, and why?

(c) (8 points) **Tor: long circuits.** In Tor, a `create` cell is sent to a Tor relay to establish new keys and make that relay be part of a circuit that is being created. For instance, a `create` cell is sent by a client to an entry guard to establish the first hop of a new circuit. To add a hop to a circuit that is being created, a client sends a `relay_early` cell, which carries an `extend` command to the (current) last hop. This last relay then uses the information in the `extend` command to send a `create` cell to the relay that should be added to the circuit.

The use of `relay_early` cells instead of the normal `relay` cells (used to transport normal data) has been added to Tor to prevent the creation of very long circuits. Relays only accept `extend` commands when they are contained in `relay_early` cells, and any relay will accept only up to 8 `relay_early` cells for a single circuit.

For the questions below, we define the *amplification factor* of a circuit as the number of (data) cell processing actions performed by honest relays divided by the number of (data) cells sent by a malicious sender. (For example, a default circuit consisting of three honest relays has an amplification factor of 3.)

i. (2 points) What is the maximum amplification factor for a simple circuit? (Without using any tricks; assuming all Tor relays are honest.) Briefly explain your reasoning.

ii. (3 points) Do you see ways to increase this factor? If so, by how much? (Assume that all Tor relays are honest.)

iii. (3 points) Mallory thinks she has found a great way to circumvent the circuit length limitation mechanisms and perform powerful DoS attacks against Tor. She sets up a malicious relay, and builds a circuit that has her malicious relay as the 8th hop. Now on the first few hops she only uses `relay` cells, even when they contain `extend` commands, and she instructs her malicious relay to change the type of such cells from `relay` to `relay_early`. She believes this can lead to arbitrary circuit length if she places her malicious relay also as the 16th, 24th, ..., hops on the circuit. Is she correct? Would this lead to an arbitrarily high amplification factor? Explain your reasoning.

## 5. Probabilistic Traffic Monitoring (13 points)

(a) (7 points) A *wireless ISP* charges the users (with individual IP addresses) of its *WiFi* service offering based on their bandwidth usage. As the number of users is large, the company decides to forgo precisely billing every user to enable the use of commodity hardware. Using Netflow, they instead sample every $k$-th packet, and bill the sampled customers according to the inferred usage. While aware that some users may luckily avoid a charge, they anticipate that users which use high bandwidth or the network for prolonged period will likely be charged.

i. (1 point) Give a definition of "flow" useful in the billing process.

ii. (2 points) Can this billing approach be considered fair for paying users with respect to their usage? Why or why not?

iii. (2 points) Describe an attack that could be launched against such a system.

iv. (2 points) The company instead decides to focus on billing the users consuming the most bandwidth. Suggest an approach for collecting flow throughput.

(b) (6 points) You are responsible for the network of a company which serves two resources, $X$ and $Y$, from two different web servers. Resource $X$ is 2 MB whereas resource $Y$ is 5 KB. You have asked for the outgoing link of 100 Gbps to be upgraded, as flows serving resource $X$ dominate the outgoing link resulting in an increased latency for flows serving resource $Y$. Unfortunately, you do not have access to the application logs to prove this to your supervisor.

i. (2 points) Describe an approach for measuring the consumed throughput of each resource, assuming the router does not have sufficient resources to individually track each flow.

ii. (4 points) Consider the case when both resources instead reside on the same server. How can you determine the exact number of flows for resource $X$ and at least an approximate of the number of flows for resource $Y$ within an interval $T$?

# 6. DNS and DNSSEC (15 points)

(a) (4 points) **Caching resolvers.** Recursive resolvers have an important role in DNS: by caching the replies they obtain, they significantly lower the average lookup latency for end hosts using the resolvers. However, the use of caching is not without its drawbacks.

i. (2 points) Name and describe one drawback in terms of *privacy*.

ii. (2 points) Why would an adversary prefer a world in which caching resolvers are used extensively (as they are in our world) over a world in which all end hosts' clients perform the full recursive look-ups on their own?

(b) (4 points) **Resolvers and DNSSEC.** Recall that DNSSEC is typically used only between recursive resolvers and authoritative name servers. Assuming DNSSEC is fully deployed, consider an end host that fully trusts the recursive resolvers provided by its ISP (so it also trusts that the resolver will always use DNSSEC).

i. (2 points) What attack undermining the authenticity of the query replies should the end host still worry about, and why?

ii. (2 points) What countermeasure (if any) could be adopted to prevent or significantly mitigate the attack you mentioned? Briefly explain your answer.

(c) (7 points) **Authenticated denial of existence.** DNSSEC is not only concerned with the authentication of the usual DNS records; it also aims to authenticate the *absence* of certain records. At a high level, the strategy DNSSEC adopts to do this is the following: to authenticate the non-existence of, e.g., `medicine.ethz.ch`, the authoritative server needs to provide a signed non-existence statement containing the pair of existing names which alphabetically come respectively right before and right after the non-existing name. For instance, this pair could be (`mavt.ethz.ch`, `mtec.ethz.ch`). Concretely, DNSSEC introduces the NSEC record, which could look as follows:

| NAME | TTL | CLASS | TYPE | NEXT DOMAIN NAME | TYPE BIT MAP |
|---|---|---|---|---|---|
| mavt.ethz.ch | 3600 | IN | NSEC | mtec.ethz.ch | (A, MX, RRSIG, MX) |

i. (2 points) The *Type Bit Map* shown above specifies what record types exist for `mavt.ethz.ch`: why is this field needed/useful in the context of authenticated denial of existence?

ii. (2 points) What other records from the `ethz.ch` domain, besides the NSEC record shown above, are needed to verify the non-existence of `medicine.ethz.ch`?

iii. (3 points) A much more trivial solution would be for the authoritative name servers to sign non-existence statements on the fly for each query they receive for non-existent records. Name *two* significant drawbacks of this solution, and briefly explain your answer.

# 7. Broadcast authentication (11 points)

(a) (1 point) Is it possible to perform authenticated broadcast communication, assuming that the set of receivers may change over time, that the receivers have no prior information about the sender, and that any trusted third party is offline (i.e., a PKI CA)?

(b) (1 point) Alice wants to send messages to mutually untrusted receivers Bob, Carol, and Dave. Alice picks a random key K and sends it to the receivers over a secure channel providing secrecy and authenticity. To broadcast message m, Alice sends out m, MAC(K, m). Since each receiver knows key K, each receiver can verify the authenticity of the message by verifying the MAC. Assuming the MAC function is secure, is this protocol secure?

(c) (9 points) Answer the following questions about the TESLA protocol.

i. (2 points) The sender has been broadcasting with TESLA to a large group, and it turns out that the sender used all keys of the original hash chain. Is it possible to extend the one-way hash chain and continue the TESLA broadcast operation efficiently? Justify your answer.

ii. (4 points) An important TESLA parameter is the key disclosure delay. Although the choice of the disclosure delay does not affect the security of the system, it is an important performance factor. As we discussed in class, a short disclosure delay will cause delayed packets to lose their safety property, so receivers will discard them, and a long disclosure delay leads to a long authentication delay for receivers.
As an alternative, the sender may include in each packet the time $t_p$ at which it is going to disclose the key for this packet. With this method, the receiver only needs to know the bound $D_t$ on the clock skew and $T_0$, the sender's local time at the initiation of the session. Then the receiver records the local time $T$ when the packet has arrived, and verifies that $T \leq T_0 + D_t + t_p$. Else the packet is considered unauthenticated. Is this secure? Justify your answer.

iii. (3 points) Instead of operating on a time basis, sender $S$ decides to operate TESLA on a packet basis. $S$ now broadcasts the packet $P_i$ along with the key $K_i$ and the message authentication code (MAC) of $P_i$ computed with the key $K_{i+1}$ as follows:

$$S \rightarrow * : P_i, K_i, MAC_{K_{i+1}}(P_i)$$

The receiver must wait for the next packet to validate the MAC of the packet $P_i$.
Is this secure? Justify your answer.

# 8. SCION (9 points)

In this question, we will explore the possibility of various attacks against SCION. NOTE: For those who do not remember the details of the SCION protocol from the lecture slides, we have reproduced the relevant data structures on the next page.

(a) (2 points) **Hop field spoofing attack**. Consider a malicious host that desires to send packets down a link for which it does not have a valid hop field. How can it create a packet that traverses that link?

(b) (3 points) **Reflection attack.** In this question, we explore if a reflection attack can be launched in SCION.

i. (1 point) Can an adversary launch a reflection attack against a victim that is in the same AS? Justify your answer.

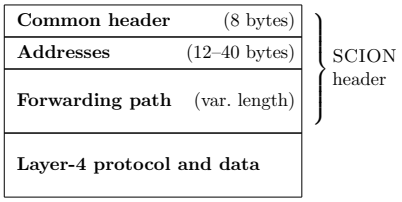
ii. (2 points) Can an adversary launch a reflection attack against a victim in another AS? Justify your answer.

(c) (2 points) **Path alteration attack** occurs when a malicious entity, such as a compromised border router in an AS, modifies the SCION path in a SCION data packet, and the packet reaches the intended destination via the modified path. Is SCION resilient against a path alteration attack? Justify your answer.

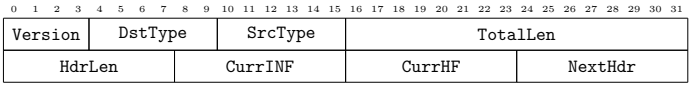
(d) (2 points) **Wormhole attack** occurs when two ASes collude to announce a bogus link between them. This consequently creates shorter paths to attract traffic. Does SCION defend against a wormhole attack? Why or why not? Justify your answer.

# Appendix: SCION Information

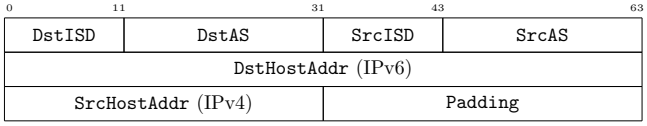
A **SCION data packet** has the header structure as shown below, and we provide more details for the three fields (i.e, **SCION Common Header, Addresses, Forwarding paths**) that constitute the **SCION header**.



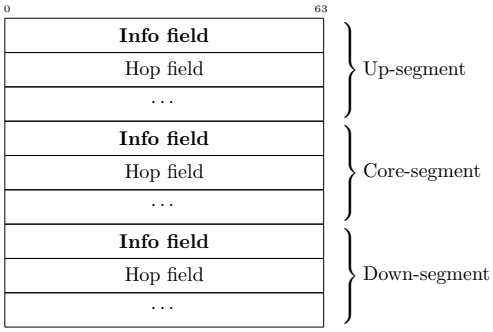

The **SCION common header** has the following structure:



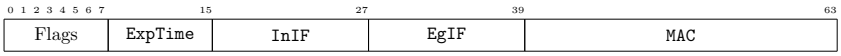

The **Addresses** field indicates the addresses of the source and destination hosts. Assuming that the source and destination ASes are an IPv6 and an IPv4 networks respectively, the addresses field has the structure as shown below. Note that a complete SCION address is defined by a (ISD Number, AS Number, and Local Address)-triplet.




The **Forwarding Path** field contains information to forward packets across ASes and has the following high-level layout:




A **Hop Field** in a forwarding path has the following structure:




The only cryptographic aspects in a SCION header are the MACs in the hop fields in a forwarding path. Each MAC, which is 24 bits, is created by each AS during the beaconing process and is computed as follows:

$$\sigma_H = \mathsf{MAC}_K(TS \parallel Flags \parallel ExpTime \parallel InIF \parallel EgIF \parallel HF')$$

In this equation, TS refers to the timestamp (in the Info field), Flags to the hop field flags, ExpTime to the expiration time, InIF and EgIF to the ingress and egress interfaces, and HF' is the hop field of the AS from which the beacon was received (it is empty when the core AS creates the first hop field).

## Extra Page

Please use this page in case you run out of space elsewhere in the exam.