Solutions to Selected Problems

Guide to Internet Cryptography

Companion Material

February 24, 2026

Preface This document provides solutions to selected problems from the book *Guide to Internet Cryptography: Security Protocols and Real-World Attack Implications*. The material is intended for educational use in courses and self-study.

Book website: https://link.springer.com/book/10.1007/978-3-031-19439-9

1 Chapter 19: Email: Protocols and SPAM

Problem 19.1 POP3 and IMAP

(a) Check the configuration of your e-mail client for SMTP and IMAP. Is TLS used? If yes, how is TLS activated?

(b) Suppose your e-mail client uses POP3 with the challenge-and-response protocol from RFC 1939 but without TLS. Can you think of how to compute a low-entropy common secret *pw* efficiently?

(c) Which identity of the client should the Kerberos ticket from Figure 19.2 contain?

Solution

(a) This is a practical exercise.

(b) Just record $RAND$ and the response RES. Then, for any password candidate pw' from the dictionary, compute $MD5(RAND, pw')$ and compare it to RES.

(c) The client's email address.

Problem 19.2 SMTP-over-TLS

One of your fellow students proposes a new Internet Draft on SMTP-over-TLS. He proposes to make the use of TLS transparent to the recipient of an e-mail by adding a flag "TLS=BOOLEAN" to each `Recieved` header. If TLS was used when receiving the email, this flag is set to TRUE by the receiving SMTP server; if not, it is set to FALSE. Can the recipient in Figure 19.3 trust this mechanism?

Solution

No, because these headers are not protected, and any attacker accessing the email can change a value "TLS=FALSE" to "TLS="TRUE".

Problem 19.3 Bayes filters

(a) Why are Bayes filters trained on words and not on sentences?

(b) Can you imagine a fully automated machine learning algorithm for SPAM detection, where a neural network scans all e-mails and automatically classifies them into SPAM and HAM?

(c) Suppose you want to minimize the False Negative rate of a SPAM filter. How can you do that?

Solution

(a) Because words are simple to measure in this simple machine-leraning approach. Todays LLM-based solutions may be trained on sentences, but they are much more complex.

(b) No. For classification, you need precise criteria when to classify an email as SPAM. Any spammer learning about these precise criteria could build his SPAM mails such that the avoid classification.

(c) You can set a lower threshold for the probability of an email being SPAM. Or you can assume that SPAM has a much greater percentage in overall traffic than 50%.

Problem 19.4 Bayes filters

60 SPAM mails and 120 HAM emails are training sets for a Bayes filter. The following table shows the frequency of four keywords in the training sets.

	SPAM (60 mails)	HAM (120 mails)
Tide (TI)	3	12
Persil (PE)	6	21
Bitcoin (BI)	36	6
Stablecoin (ST)	24	9

(a) The assumption for the probability of SPAM and HAM is Pr(SPAM) $= 1/2 =$ Pr(HAM), and the threshold for classifying it as spam is 80%. In the table, see the frequency of occurrence of the TI, PE, BI, and ST in SPAM and HAM emails in different-sized training sets. Now calculate for each of these words the probability that a mail containing this word is SPAM and then decide whether it will be classified as SPAM by the filter.

(b) Calculate the probabilities Pr(SPAM | TI AND PE) and Pr(SPAM | BI AND ST) and decide in each case whether emails containing these two words are classified as SPAM.

(c) Now make the assumption Pr(SPAM) $= 4/5$ and Pr(HAM) $= 1/5$. Using these new a priori probabilities, compute Pr(SPAM | TI AND BI) and indicate whether mail containing these two words will be classified as spam.

Solution

Naïve Bayes Spam Classification

Training data:

$$|\text{SPAM}| = 60, \qquad |\text{HAM}| = 120$$

(a) Single Words, $P(\textbf{SPAM}) = P(\textbf{HAM}) = \frac{1}{2}$

Bayes formula:

$$P(S \mid w) = \frac{P(w \mid S)P(S)}{P(w \mid S)P(S) + P(w \mid H)P(H)}$$

Since priors are equal, they cancel:

$$P(S \mid w) = \frac{P(w \mid S)}{P(w \mid S) + P(w \mid H)}$$

Tide (TI)

$$P(TI \mid S) = \frac{3}{60} = 0.05$$

$$P(TI \mid H) = \frac{12}{120} = 0.10$$

$$P(S \mid TI) = \frac{0.05}{0.05 + 0.10} = \frac{0.05}{0.15} = 0.333$$

Persil (PE)

$$P(PE \mid S) = \frac{6}{60} = 0.10$$

$$P(PE \mid H) = \frac{21}{120} = 0.175$$

$$P(S \mid PE) = \frac{0.10}{0.10 + 0.175} = \frac{0.10}{0.275} = 0.364$$

Bitcoin (BI)

$$P(BI \mid S) = \frac{36}{60} = 0.60$$

$$P(BI \mid H) = \frac{6}{120} = 0.05$$

$$P(S \mid BI) = \frac{0.60}{0.60 + 0.05} = \frac{0.60}{0.65} = 0.923$$

Stablecoin (ST)

$$P(ST \mid S) = \frac{24}{60} = 0.40$$

$$P(ST \mid H) = \frac{9}{120} = 0.075$$

$$P(S \mid ST) = \frac{0.40}{0.40 + 0.075} = \frac{0.40}{0.475} = 0.842$$

(b) Two Words (Naïve Independence)

$$P(S \mid w_1, w_2) = \frac{P(w_1 \mid S)P(w_2 \mid S)}{P(w_1 \mid S)P(w_2 \mid S) + P(w_1 \mid H)P(w_2 \mid H)}$$

TI and PE

$$P(TI, PE \mid S) = 0.05 \cdot 0.10 = 0.005$$

$$P(TI, PE \mid H) = 0.10 \cdot 0.175 = 0.0175$$

$$P(S \mid TI, PE) = \frac{0.005}{0.005 + 0.0175} = \frac{0.005}{0.0225} = 0.222$$

BI and ST

$$P(BI, ST \mid S) = 0.60 \cdot 0.40 = 0.24$$

$$P(BI, ST \mid H) = 0.05 \cdot 0.075 = 0.00375$$

$$P(S \mid BI, ST) = \frac{0.24}{0.24 + 0.00375} = \frac{0.24}{0.24375} = 0.985$$

(c) New Priors: $P(S) = \frac{4}{5}$, $P(H) = \frac{1}{5}$

$$P(S \mid w_1, w_2) = \frac{P(w_1 \mid S)P(w_2 \mid S)P(S)}{P(w_1 \mid S)P(w_2 \mid S)P(S) + P(w_1 \mid H)P(w_2 \mid H)P(H)}$$

TI and BI

$$P(TI, BI \mid S) = 0.05 \cdot 0.60 = 0.03$$

$$P(TI, BI \mid H) = 0.10 \cdot 0.05 = 0.005$$

$$\text{Numerator} = 0.03 \cdot 0.8 = 0.024$$

$$\text{Denominator} = 0.024 + (0.005 \cdot 0.2) = 0.024 + 0.001 = 0.025$$

$$P(S \mid TI, BI) = \frac{0.024}{0.025} = 0.96$$

Problem 19.5 DKIM

(a) Why don't OpenPGP and S/MIME sign the e-mail headers?

(b) Would `h=From:From:To:To:From:To:Subject:Subject` be a valid parameter in the DKIM header? Suppose this e-mail contains three `FROM`, two `TO`, and one `SUBJECT` header – in which sequence would they be hashed?

(c) Canonicalize the following header with the `relaxed` method:
`sUBjeCT<SP>:<HTAB>ALeX<SP><SP><CRLF><HTAB>is<SP>cool<CRLF>`

(d) Why must the value of the parameter `bh` always be shorter than the value of the parameter `b`?

Solution

(a) This was a design choice to simplify implementations. Signing the body of an email can be done by simply choosing the byte sequence between the empty line separating RFC 822 header and body, and the end of the email (and optionally applying canonicalization to this). Signing headers implies implementing a selection mechanisms to extract certain substrings from the total header string based on ASCII keywords, and canonicalizing it.

However, there is no cryptographic reason for not signing static email headers. For S/MIME, all cryptographic operations are focused on MIME objects, and this philosophy prevents the inclusion of other headers than MIME headers.

(b) YES.
The string to be hashed consists of canonicalized versions of

- the last and second last FROM headers,
- the last and second last TO headers,
- the first FROM header,
- an empty string, since there is no third TO header,
- the last SUBJECT header, and an empty string since there is no second SUBJECT header.

(c) `subject:ALeX<SP>is<SP>cool<CRLF>`

(d) Because the hash value contained in `bh` must be embeddable in the algebraic structure (DSA group, RSA ring) in which the signature contained in `b` is computed.

Problem 19.6 SPF

(a) Suppose a botnet sends SPAM e-mails using the accounts available on the infected computers. Can SPF detect this attack?

(b) The `include:otherdomain.test` directive includes the SPF whitelist for `otherdomain.test` in the current SPF whitelist. Consider the following two SPF policies:
`shop.com` IN TXT "v=spf1 ip4:192.0.2.2 include:shop.co.uk -all"
`shop.co.uk` IN TXT "v=spf1 ip4:192.0.2.1 include:shop.com -all"
Which problem might occur when evaluating one of these policies?

Solution

(a) No, since these SPAM emails will be sent from legitimate, but hacked email accounts.

(b) This is an endless loop, because the two `include` directives reference each other.

Problem 19.7 DMARC

Consider the following DMARC policy for `joergschwenk.com`:
`"v=DMARC1; p=reject; pct=100; rua=mailto:dmarc-report@joergschwenk.com; ruf=mailto:dmarc-report@joergschwenk.com; adkim=r; aspf=r;"`
Suppose that the DKIM and SPF checks were OK. For the alignment, we have
822.From=`mail@joergschwenk.com`,
821.MailFrom=`mailer@mail.joergschwenk.com`, and the DKIM parameter
`d=googlemail.com`.
Which alignment will be OK?

Solution

The alignement for SPF is OK. Since `aspf=r`, the 822.From (5322.From) and 821.MailFrom (5321.MailFrom) addresses only must have an organisatorial domain on common, and this is `joergschwenk.com`.

The alignment for DKIM is not OK. There is no DKIM signature that contains the domain `joergschwenk.com` from 822.From, or any subdomain of it, in the `d=` parameter.