Final exam in

Web Security EITF05

Department of Electrical and Information Technology
Lund University

October 27th, 2017

- You may answer in either Swedish or English.
- If any data is lacking, make and state reasonable assumptions.
- Use legible hand writing. If your answers cannot be read, you will receive zero points on that problem.
- Grading is done as follows.
  Grade 3 = 20–29 points,
  Grade 4 = 30–39 points,
  Grade 5 = 40–50 points.

Good luck!

Paul

Problem 1. On October 18th 2017, Svenska Dagbladet wrote that a consultant with security clearance used his home computer to publicly post sensitive source code from Pensionsmyndigheten and/or Försäkringskassan. The code was exposed for public scrutiny for six days before this security breach was detected and mitigated.

Assuming the worst, every single hacker now has access to this code and has analysed it for weaknesses. Assume that the code consists of badly written PHP code.

a) What kind of attack might a hacker typically attempt to employ in order to steal sensitive information from an internal database at Försäkringskassan?
b) What is the best countermeasure for such attacks?
c) Would the database content be secure in enemy hands (unreadable to the hacker) if it were encrypted with a modern (unbroken) symmetric cipher? Motivate.

Answer

a) An SQL injection.
b) Prepared statements.
c) Probably not, since the encryption/decryption key is most likely stored in the source code.

(3 points)

Problem 2. Consider the session functionality in PHP.

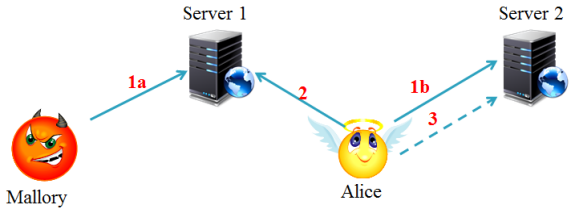
a) How do PHP sessions work? (Explain the interplay between the session id in a cookie and the global variable `_SESSION`.)
b) How does PHP prevent session prediction attacks?

Answer

a) See illustration in lecture slides.
b) PHP (pseudo-)randomly generates high-entropy session ids. (hard to guess)

(2+1 points)

Problem 3. Consider the following illustration of a CSRF attack.



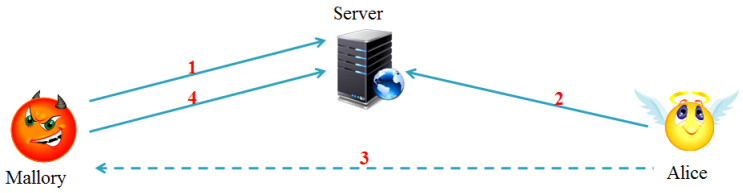
Why do CSRF tokens efficiently protect your website from CSRF attacks?

Answer
CSRF token is embedded in a hidden field that Mallory cannot access (available only to Alice), so Mallory cannot construct a valid request (step 3). (3 points)

Problem 4. Give a regular expression that matches an email address. You may assume that the name parts of the email address contain alphanumerical characters only.

Answer
There are many possibilities; `^[A-Za-z0-9]+\.[A-Za-z0-9]+@[A-Za-z0-9]+\.com$`. (3 points)

Problem 5. Consider the following illustration of an XSS attack with three involved entities; Mallory, Server and Alice.



a) What can the Server do to efficiently prevent XSS attacks?

b) What can Alice do to efficiently prevent XSS attacks?

Answer

a) Apply CSP. Filter user input to avoid script injection.

b) She can log out when her browsing session is finished, this destroys the session id. She can also switch off client-side scripting, but that would break the functionality of many web sites, which is not really a good option.

(1.5+1.5 points)

Problem 6. On October 17th 2017, security specialist Brian Krebs wrote that a new massive IoT botnet sporting more than one million IoT-devices was identified. This massive botnet was harvested using a variation of the 'Mirai' code called 'IoT Reaper'.

Furthermore, on October 24th 2017, Sydsvenskan wrote about a distributed denial of service (DDoS) attack against servers at Skånetrafiken. The ticketing system was unavailable for three hours.

These two events were most likely not related, but botnets are very useful for implementing powerful DDoS-attacks.

a) What measure, other than using a powerful botnet, can you take to make a DDoS-attack as efficient as possible? Briefly explain how it works.

b) Estimate the traffic load that you can attain using the IoT Reaper botnet after including your improvement in a). Make and state reasonable assumptions.

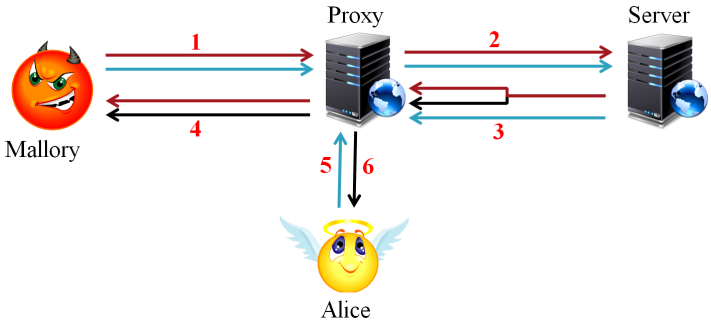
Answer

a) DNS amplification. Send UDP packet with DNS query to DNSSEC-enabled DNS server, write victims IP as destination address. The answer can be about 50 times larger than the original query.

b) No given answer, but it is a good idea to clearly present your formula.

(1+2 points)

Problem 7. Consider the following illustration of an HTTP response splitting attack.



a) Referring to the illustration, which arrows (state number and color) contain the spoofed page that Mallory wants to serve to Alice?

b) Can Mallory control the TTL of the spoofed page (how long it is cached)? Motivate.

Answer

a) 1 red, 2 red, 3 black, 4 black, 6 black.

b) Yes, the entire page including HTTP headers can be spoofed.

(1.5+1.5 points)

Problem 8. There are many different ways of protecting against spam. Explain each of the following terms, relating them to spam protection.

a) Greylisting.

b) Bayesian filtering.

Answer

a) A spam protection method that utilizes that spammers often do not follow the SMTP protocol. A mail server can 'greylist' a new senders by default, refusing to deliver the mail. If the sender resends (follows the protocol), she will become whitelisted, and all future mails will be forwarded without delay.

b) Statistical filtering is often used as the core of spam detection, for classifying incoming mail as either spam or legitimate. The basic functionality, as explained in the lecture notes, considers the mail according to the bag-of-words paradigm.

(1.5+1.5 points)

Problem 9. You live in a dystopic society that has started to randomly censor literature. Unfortunately, they have censored your favorite book, The Hitchhiker's Guide to the Galaxy. Fortunately for you, however, they are using Base64 to do this. What are the censored words in the following quotes? Choose any **one**.

- Don't `UGFuaWM=`.
- There is an art, it says, or rather, a knack to flying. The knack lies in learning how to throw yourself at the ground and `bWlzcw==`.
- The ships hung in the sky in much the same way that `YnJpY2tz` don't.
- Forty-two, said Deep Thought, with infinite majesty and `Y2FsbQ==`.
- It is a mistake to think you can solve any major problems just with `cG90YXRvZXM=`.

Hint: Decimal representation of ASCII characters:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122

The Base64 alphabet:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51

0	1	2	3	4	5	6	7	8	9	+	/
52	53	54	55	56	57	58	59	60	61	62	63

Note: The first (top) quote above also happens to be very good advice for you on this exam. (3 points)

Answer

- Don't Panic.
- There is an art, it says, or rather, a knack to flying. The knack lies in learning how to throw yourself at the ground and miss.
- The ships hung in the sky in much the same way that bricks don't.
- Forty-two, said Deep Thought, with infinite majesty and calm.
- It is a mistake to think you can solve any major problems just with potatoes.

Problem 10. HTTP digest authentication (RFC2617) is a challenge response protocol in which the client calculates the digest (the response) according to

$$\text{MD5}(\ \text{MD5}(A1) : \mathit{nonce} : \mathit{nc} : \mathit{cnonce} : \mathit{qop} : \text{MD5}(A2)\ ),$$

with

$$\begin{aligned} A1 &= \mathit{username} : \mathit{realm} : \mathit{password}, \\ A2 &= \begin{cases} \mathit{method} : \mathit{URI} & \text{if } \mathit{qop} = \mathit{auth}, \\ \mathit{method} : \mathit{URI} : \text{MD5}(\mathit{entity\text{-}body}) & \text{if } \mathit{qop} = \mathit{auth\text{-}int}. \end{cases} \end{aligned}$$

a) Explain the usage and purpose of the *realm* parameter.

b) Is the protocol vulnerable to TMTO-attacks? Motivate.

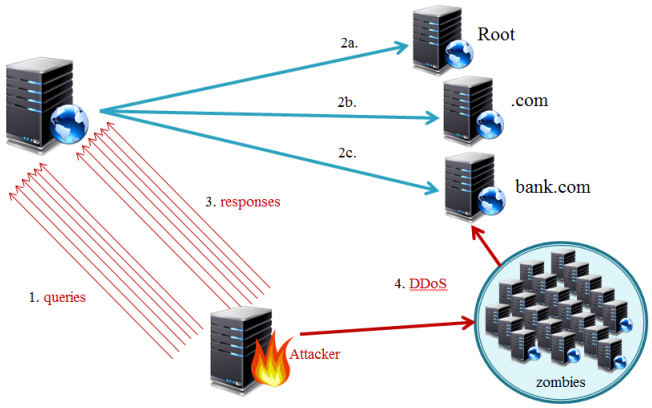
Answer

a) A text string telling the user which password to enter (same site can use different ones for access to different resources).

b) The parameter *cnonce* protects against TMTO attacks. A MITM with the ability to alter messages could perform a TMTO attack if the *cnonce* parameter is not present, replacing all randomness generated by the server with her own pre-determined random looking string that she has pre-built her TMTO tables for. This does not work if the *cnonce* parameter is present, because it is generated by the client and used when preparing the digest, and it is revealed to the MITM *afterwards*. Pre-generate tables therefore become meaningless, as the MITM would need to use different tables for every request, leaving the MITM with brute-force as a better but infeasible option.

(1+2 points)

Problem 11. Consider the following illustration of a DNS cache poisoning attack using a very large IoT Reaper botnet.



a) If 64-bit transaction IDs were used (instead of 16-bit), roughly how many responses would the attacker need to send in step 3 to complete a successful DNS cache poisoning attack with high probability. Make and state relevant assumptions. You may, for example, ignore the randomness in the port number.

b) Can the attacker choose to target the responses from the `Root` or `.com` servers in steps 2a or 2b instead of the responses from `bank.com` at step 2c? Motivate.

Answer

a) Roughly 2^{32}, due to the birthday paradox.

b) Yes, this is the same set-up with randomly generated transaction IDs. The attacker should redirect the botnet to target those machines in that case, for slower response times.

(2+3 points)

Problem 12. A DKIM signature header of an email is given below.

```
DKIM-Signature:
v=1;
a=rsa-sha256;
c=simple/relaxed;
d=gmail.com;
s=gamma;
h=received:message-id:date:from:to:subject:mime-version:content-type;
bh=9gicsZnlcLK7yYh6VIrgyAMMRZiWsSbWqSPIhc78RRk=;
b=k4ofvpHPkaQmvuSoGVhRrnCsPK+JEuv9KUrZO7aiypvf/6Y1N2iIatvLvdzwOnZX
  /W6Kxyx6Z4Ybuk8Dqk/vNTIE7Jpy+GQUUHFvM0NFtmZo1CbGRvo8DdHnXRBB/qWw
  lV+Z6wxw/mq7lNuJknVprOAaTLws5mwcZ+AWL8KwHg0=
```

a) Explain the purpose and usage of the selector.

b) DKIM is used as a component in DMARC. Briefly, what other components are used in DMARC?

Answer

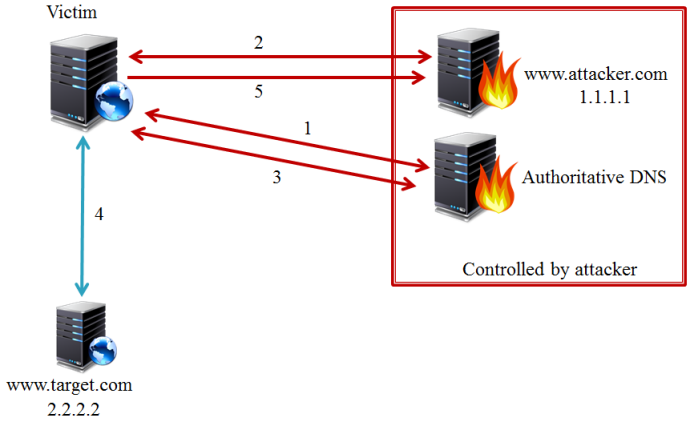
a) The selector specifies which (of possibly many) public domain key to use to verify the signature.

b) Apart from DKIM, the central components are SPF records and the concept of alignment. In addition, a feedback system is put in place for the configuring party to be able to analyze how well the current configuration is performing.

(2+3 points)

Problem 13. Consider the following illustration of a DNS rebinding attack.

a) Briefly explain how the attack works.
b) Does the victim need to have JavaScript enabled for the attack to work? Motivate.



Answer

a) See lecture notes/slides.
b) For the attack to be useful, yes. Otherwise it is difficult to complete step 5, deliver the retrieved content to the attacker.

(3+2 points)

Problem 14. Briefly explain the following terms and acronyms.

a) CORS
b) DNSSEC
c) MX record
d) Hashcash
e) Rainbow table

Answer

a) Cross-origin resource sharing. See lecture notes.
b) Domain Name System Security Extensions. Provides authenticated DNS responses.
c) DNS entry for mail servers.
d) A spam protection mechanism. See lecture notes.
e) A technique for reversing hash functions. See lecture notes.

(5 points)
