Web Security and Special Topics

Question 1 ***Boogle*** ()

Boogle is a social networking website that's looking into expanding into other domains. Namely, they recently started a map service to try their hand at fusing that with social media. The URL for the main website is https://www.boogle.com, and they want to host the map service at https://maps.boogle.com.

(a) Describe how to make a cookie that will be sent to only Boogle's map website and its subdomains.

Solution: Set the domain parameter of the cookie to .maps.boogle.com

(b) How can Boogle ensure that cookies are only transmitted encrypted so eavesdroppers on the network can't trivially learn the contents of the cookies?

Solution: Set the secure flag on each cookie.

(c) Boogle adds the ability for users to check in to locations on maps.boogle.com, but discovers an XSS vulnerability that slipped through QA. Name a hotfix they can do to prevent scripts from stealing cookies using XSS.

Solution: Set the HTTPOnly flag on cookies.

(d) Some of the XSS attacks are scraping sensitive information from the map site, like user emails. The security team wants to know the scope of the vulnerability. Can attackers use XSS to also scrape sensitive information from the main site, https://www.boogle.com? Explain why or why not.

Solution:

No, the two sites have different origins so https://maps.boogle.com cannot read anything from https://www.boogle.com.

(e) Boogle wants to be able to host websites for users on their servers. They decide to host each user's website at https://[username].boogle.com. Why might this not be a good idea?

Solution: A malicious user could set cookies that would be sent to other users' sites as well as the entire `.boogle.com` domain. Also, any cookies meant for `boogle.com` will go to the malicious user.

(f) Propose an alternate scheme so that Boogle can still host other users websites with less risk, and explain why this scheme is better.

Note: It is okay if the user sites interfere with each other, as long as they cannot affect official Boogle websites.

Solution:

Boogle should create a new domain exclusively for user hosted content, like `https://[username].boogleusercontent.com`. This way, user sites cannot set cookies that will affect all boogle domains due to the cookie setting policy. This is known as a cookie tossing attack, and is one of the reasons why github hosts user sites on github.io instead of github.com (see `https://blog.github.com/2013-04-09-yummy-cookies-across-domains/`).

Question 2 ***Tracking*** ()

Let's say the web-page at `http://cute-puppies.com` looks like the following:

```
<html>
    <body>
        <p>Here is a GIF of puppies</p>
        <img src="https://image-host.com/yvxSMjw.gif">
        <script type="text/javascript"
                src="http://yahoo.com/analytics.js"></script>
        <script type="text/javascript"
                src="https://google.com/analytics.js"></script>
    </body>
</html>
```

Note that `google.com` is loaded over HTTPS, whereas `yahoo.com` is loaded over HTTP.

Alice uses Mozilla Firefox on her laptop running Microsoft Windows. In her first browser tab, she has `https://berkeley.edu` open. In a second tab, she opens `http://cute-puppies.com`. In a third tab, she opens `http://cute-puppies.com` once again.

Assume that no two entities share information out of band and `cute-puppies.com` and `image-host.com` keep cookies for users. Each of the parts below are independent.

(a) Assuming Alice does not use any tracking protection, which entities know that the same person visited `cute-puppies.com` twice?

- ■ `cute-puppies.com` operators
- ■ `yahoo.com` operators
- ■ `google.com` operators
- ■ `image-host.com` operators
- □ Microsoft
- □ Mozilla
- ■ Alice's ISP
- □ UC Berkeley

Solution:

Most sites will see this, as Alice's browser makes requests to all of these sites twice.

Alice's ISP sends off all of Alice's traffic, so of course they can see what sites Alice visits.

Neither Microsoft nor Mozilla collect information about what sites you visit in your browser.

UC Berkeley's site is in a different origin, so it cannot see information about the `cute-puppies.com` tab.

(b) Assume Alice opted in for a privacy service run by her ISP. This privacy service

blocks analytics scripts based on a URL-based blocklist (not host-based). Which entities know that the same person visited `cute-puppies.com` twice?

- ■ `cute-puppies.com` operators
- ☐ `yahoo.com` operators
- ■ `google.com` operators
- ■ `image-host.com` operators
- ☐ Microsoft
- ☐ Mozilla
- ■ Alice's ISP
- ☐ UC Berkeley

Solution:

The ISP service can block the HTTP connection, but not the HTTPS connection since HTTPS hides the path of the URL we are visiting. Even though the ISP can see that the user is requesting `google.com`, they do not know if this is an analytics URL.

(c) Assume Alice uses a browser plugin. The browser-plugin blocks the analytics scripts based on a URL-based blacklist (not host-based). Which entities know that the same person visited `cute-puppies.com` twice?

- ■ `cute-puppies.com` operators
- ☐ `yahoo.com` operators
- ☐ `google.com` operators
- ■ `image-host.com` operators
- ☐ Microsoft
- ☐ Mozilla
- ■ Alice's ISP
- ☐ UC Berkeley

Solution:

The browser plugin sees all requests, so it can also block HTTPS requests.

(d) Assume Alice uses a VPN run by UC Berkeley. Which entities know that the same person visited `cute-puppies.com` twice?

- ■ `cute-puppies.com` operators
- ■ `yahoo.com` operators
- ■ `google.com` operators
- ■ `image-host.com` operators
- ☐ Microsoft
- ☐ Mozilla
- ☐ Alice's ISP
- ■ UC Berkeley

Solution:

Alice's connection to the VPN is encrypted, so her ISP does not know which sites she is visiting. However now UC Berkeley sees all of her traffic instead.

Question 3 ***A Tour of Tor*** ()

As a reminder, when connecting to a normal website through Tor, your computer first queries the Tor "consensus" to get a list of all Tor nodes, and using this information it connects to the first Tor node and, from there, creates a circuit through the Tor network, eventually ending at an exit node.

(a) (4 min) Consider the scenario where you are in a censored country and the censor choses not to block Tor, the censor is the adversary, and no Tor relays exist within this country. How many Tor relays must your traffic pass through, including the exit node, to prevent the censor from blocking your traffic.

- ● One
- ○ Two
- ○ Three
- ○ Four
- ○ Tor doesn't stop this adversary

Solution: The censor doesn't block Tor and the relay is outside of the country, so one hop will get you safely past the censor.

(b) (4 min) Consider the scenario where you are the only user of Tor on a network that keeps detailed logs of all IPs contacted. You use Tor to email a threat. The network operator is made aware of this threat and that it was sent through Tor and probably originated on the operator's network. How many Tor relays must your traffic pass through, including the exit node, to guarantee the network operator can't identify you as the one who sent the threat?

- ○ One
- ○ Two
- ○ Three
- ○ Four
- ● Tor doesn't stop this adversary

Solution: Since you are the only Tor user, the network operator can look through the list of IPs and see that you contacted a Tor relay regardless of how many relays you use.

(c) (4 min) Consider the scenario where there is a single hostile Tor node but you don't know that node's identitity, and that node can be an exit node. You want to keep confidential from this node what HTTP sites you are visiting through Tor. How many Tor relays must your traffic pass through, including the exit node, to guarantee this adversary can't know what sites you visit?

- ○ One
- ● Two
- ○ Three
- ○ Four
- ○ Tor doesn't stop this adversary

Solution: If you only use a single relay, then if that relay is hostile they will be able to see your request. If you use two relays, the first relay cannot see your request, and the second can see your request but doesn't know who it's from. So in either cases, you are protected.

(d) (4 min) Consider the scenario where there are mulitple independent hostile Tor nodes but you don't know their identities, and these nodes can be exit nodes. You want to keep confidential from all these nodes what HTTP sites you are visiting through Tor. How many Tor relays must your traffic pass through, including the exit node, to guarantee that every independent hostile node can't know what sites you visit?

- ○ One
- ● Two
- ○ Three
- ○ Four
- ○ Tor doesn't stop this adversary

Solution: Same as before, no individual node can ever both your identity and the request as long as you use at least two relays.

(e) (4 min) Consider the scenario where there are multiple colluding hostile Tor nodes but you don't know those nodes identities, and these nodes can be exit nodes. You want to keep confidential from all these nodes what HTTP sites you are visiting through Tor. How many Tor relays must your traffic pass through, including the exit node, to guarantee that the colluding system of hostile nodes can't know what sites you visit?

- ○ One
- ○ Two
- ○ Three
- ○ Four
- ● Tor doesn't stop this adversary

Solution: You can never guarantee this since any set of relays you pick could be all colluding and hostile.

Note that in real life, using three relays makes the probability of this happening negligible (assuming a certain amount of randomness in relay selection).

(f) (4 min) Consider the scenario where there is a single hostile Tor node but you don't know that node's identity, and that node can be an exit node. You want to have data integrity for the HTTP sites you are visiting through Tor. How many Tor relays must your traffic pass through, including the exit node, to guarantee this adversary can't manipulate the data you receive from the sites you visit?

- ○ One
- ○ Two
- ○ Three
- ○ Four
- ● Tor doesn't stop this adversary

Solution: The exit node can always tamper with HTTP traffic without detection