

Question 2 *Tracking*

(15 min)

- (a) Sam is researching which pair of headphones to buy, and he visits a few different blogs outlining the pros and cons of each model. He then visits his favorite social media site, FaceSpace, and notices incredibly targeted ads, which advertise specific headphone models. He goes back to each of the blogs that he visited, and sees that each one had an iframe tag containing an embedded FaceSpace like button for that page. Looking closer, he sees that each iframe source URL is structured as `facespace.com/like_button?id=<ID>`, and that each blog page he visited has a different ID. How did FaceSpace figure out that Sam is interested in purchasing a pair of headphones?
- (b) Sam figures that he can maintain his privacy from FaceSpace simply by removing any FaceSpace like buttons embedded onto the webpages he views. So he writes a small extension to his browser that removes all FaceSpace like buttons before loading the page. Content, he continues browsing, this time comparing different graphics cards. Unfortunately, when Sam goes back to FaceSpace, his page is filled with graphics card ads. Sam concludes that some of the sites he visited must be cooperating in some way with FaceSpace, but isn't sure about the details. What are some ways that FaceSpace could've figured out that Sam is interested in graphics cards?

(c) Sam is now done with FaceSpace and their invasive tracking. He decides to clear all of his cookies and go back to browsing different types of headphones. He reads a blog comparing two headphone models, and clicks a link from that blog to a RedFeed post about cats, only to find another advertisement about headphones on the page. But when Sam reviewed his traffic log, no cookies were sent to RedFeed in his request, and no extra information was passed in the URL. How did RedFeed figure out that Sam is interested in headphones?

(d) What are some ways of avoiding web-based tracking? List some pros and cons for each.

Question 3 *Intrusion Detection*

(10 min)

FooCorp is deciding which intrusion detection method to employ in a few target scenarios. In each part, consider which of the intrusion detection methods learned in class would be most appropriate, and justify why. Try to be as specific as possible.

- (a) FooCorp wants to detect attacks for a specific vulnerability that may exist in some of their web servers.

- (b) FooCorp is using HTTPS, but all of their services use the same modular web framework. They are interested in detecting any time their servers receive arguments that are suspicious, in real-time.

- (c) FooCorp is a diverse company, with a wide variety of different web servers built on top of different web frameworks, offering different services. They wish to detect suspicious arguments for all of their services. Every service uses HTTP and not HTTPS, and FooCorp has a low budget for security, but they want real-time detection.

- (d) FooCorp again has many different web servers built on different web frameworks, but each uses the same logging format. They are using HTTPS, and do not need real-time detection.

Question 4 *Detection Tradeoffs***(10 min)**

Suppose that S is a network-based intrusion detector that works by passively analyzing individual UDP and TCP packets. Suppose that A is a host-based intrusion detector that is a component of the browser that processes and analyzes individual URLs before they are loaded by the browser. Suppose S has false positive rate S_P and false negative rate S_N , and A has false positive rate A_P and false negative rate A_N .

Your company decides to build a hybrid scheme for detecting malicious URLs. The hybrid scheme works by combining scheme S and scheme A , running both in parallel on the same traffic. The combination could be done in one of two ways. Scheme H_E would generate an alert if for a given network connection either scheme S or scheme A generates an alert. Scheme H_B would generate an alert only if both scheme S and scheme A generate an alert for the same connection. (Assume that there is only one URL in each network connection.)

- (a) Assuming that decisions made by S and A are well-modeled as independent processes, and ignoring any concerns regarding evasion, what can you say about the false positives and false negatives of H_B and H_E ? In terms of S_P, S_N, A_P, A_N , what are the false positive and false negative rates for H_B and H_E .
- (b) If deploying the hybrid scheme in a new environment, is one of H_E and H_B clearly better? If not, what environment parameters would help determine whether H_E or H_B is better, and for each parameter p , increasing p favors which hybrid scheme?