

An online version is available at <https://textbook.cs161.org>.

Textbook by [David Wagner](#), [Nicholas Weaver](#), [Peyrin Kao](#),
Fuzail Shakir, Andrew Law, and [Nicholas Ngai](#)

Additional contributions by Noura Alomar, Sheqi Zhang, and [Shomil Jain](#)

Last update: August 23, 2021

Contact for corrections: cs161-staff@berkeley.edu

It would not be too much of a stretch to say that much of today's world is built upon the Internet. Many of the services that run on top of the Internet come with their own class of vulnerabilities and defenses to match. In particular, we will be focusing on web security, which covers a class of attacks that target web pages and web services.

17 SQL Injection

17.1 Code Injection

SQL injection is a special case of a more broad category of attacks called code injections.

As an example, consider a calculator website that accepts user input and calls `eval` in Python in the server backend to perform the calculation. For example, if a user types `2+3` into the website, the server will run `eval("2+3")` and return the result to the user.

If the web server is not careful about checking user input, an attacker could provide a malicious input like

```
2+3"); os.system("rm -rf /
```

When the web server plugs this into the `eval` Python function, the result looks like

```
eval("2+3"); os.system("rm -rf /")
```

If interpreted as code, this statement causes the web server to delete all its files!

The general idea behind these attacks is that a web server uses user input as part of the code it runs. If the input is not properly checked, an attacker could create a special input that causes unintended code to run on the server.

17.2 SQL Injection Example

Many modern web servers use SQL databases to store information such as user logins or uploaded files. These servers often allow users to interact with the database through HTTP requests.

For example, consider a website that stores a SQL table of course evaluations named `evals`:

id	course	rating
1	cs61a	4.5
2	cs61b	4.4
3	cs161	5.0

A user can make an HTTP GET request for a course rating through a URL:

```
http://www.berkeley.edu/evals?course=cs61a
```

To process this request, the server performs a SQL query to look up the rating corresponding to the course the user requested:

```
SELECT rating FROM evals WHERE course = 'cs61a'
```

Just like the code injection example, if the server does not properly check user input, an attacker could create a special input that allows arbitrary SQL code to be run. Consider the following malicious input:

```
garbage'; SELECT password FROM passwords WHERE username = 'admin
```

When the web server plugs this into the SQL query, the resulting query looks like

```
SELECT rating FROM evals WHERE course = 'garbage';  
SELECT password FROM passwords WHERE username = 'admin'
```

If interpreted as code, this causes the query to return the password for the `admin` user!

17.3 SQL Injection Strategies

Writing a malicious input that creates a syntactically valid SQL query can be tricky. Let's break down each part of the malicious input from the previous example:

- `garbage` is a garbage input to the intended query so that it doesn't return anything.
- `'` closes the opening quote from the intended query. Without this closing quote, the rest of our query would be treated as a string, not SQL code.
- `;` ends the intended SQL query and lets us start a new SQL query.
- `SELECT password FROM passwords WHERE username = 'admin` is the malicious SQL query we want to execute. Note that we didn't add a closing quote to `'admin`, because the intended SQL query will automatically add a closing quote at the end of our input.

Consider another vulnerable SQL query. This time, we have a `users` table that contains the `username` and `password` of every user.

When the web server receives a login request, it creates a SQL query by plugging in the username and password from the request. For example, if you make a login request with username `alice` and password `password123`, the resulting SQL query would be

```
SELECT username FROM users WHERE username = 'alice'
AND password = 'password123'
```

If the query returns more than 0 rows, the server registers a successful login.

Suppose we want to login to the server, but we don't have an account, and we don't know anyone's username. How might we achieve this using SQL injection?

First, in the username field, we should add a dummy username and a quote to end the opening quote from the original query:

```
SELECT username FROM users WHERE username = 'alice'
' AND password = '_____'
```

Next, we need to add some SQL syntax so that this query returns more than 0 rows (since we don't know if `alice` is a valid username). One trick for forcing a SQL query to always return something is to add some logic that always evaluates to true, such as `OR 1=1`:

```
SELECT username FROM users WHERE username = 'alice' OR 1=1
' AND password = '_____'
```

Next, we have to add some SQL so that the rest of the query doesn't throw a syntax error. One way of doing this is to add a semicolon (ending the previous query) and write a dummy query that matches the remaining SQL:

```
SELECT username FROM users WHERE username = 'alice' OR 1=1;
SELECT username FROM users WHERE username = 'alice' AND password = '_____'
```

The second query might not return anything, but the first query will return a nonzero number of entries, which lets us perform a login. The last step is to add some garbage as the password:

```
SELECT username FROM users WHERE username = 'alice' OR 1=1;
SELECT username FROM users WHERE username = 'alice' AND password = 'garbage'
```

Thus, our malicious username and password should be

```
username = 'alice' OR 1=1; SELECT username FROM users WHERE username = 'alice'
password = 'garbage'
```

Another trick to make SQL injection easier is the `--` syntax, which starts a comment in SQL. This tells SQL to ignore the rest of the query as a comment.

In our previous example, we can instead start a comment to ignore parts of the query we don't want to execute:

```
SELECT username FROM users WHERE username = 'alice' OR 1=1
--' AND password = 'garbage'
```

Thus, another malicious username and password is

```
username = 'alice' OR 1=1--
password = 'garbage'
```

Further reading: [SQL Injection Attacks by Example](#)

17.4 Defense: Escape Inputs

One way of defending against SQL injection is to escape any potential input that could be used in an attack. Escaping a character means that you tell SQL to treat this character as part of the string, not actual SQL syntax.

For example, the quote `'` is used to denote the end of a string in SQL. However, the escaped quote `\'` is treated as a literal quote character in SQL, and it does not cause the current string to end.

By properly replacing characters with their escaped version, malicious inputs such as the ones we've been creating will be treated as strings, and the SQL parser won't try to run them as actual SQL commands.

For example, in the previous exploit, if the server replaces all instances of the quote `'` and the dash `-` with escaped versions, the SQL parser will see

```
SELECT username FROM users WHERE username = 'alice\' OR 1=1\-\-'
AND password = 'garbage'
```

The escaped quote won't cause the `username` string to end, and the escaped dashes won't cause a comment to be created. The parser will try to look up someone with a username `alice' OR 1=1--` and find nothing.

However, we have to be careful with escaping. If an attacker inputs a backslash followed by a quote `\'`, the escaper might “escape the escape” and give the input `\\'` to the SQL parser. The parser will treat the two backslashes `\\` as an escaped backslash, and the quote won't be escaped!

The key takeaway here is that building a good escaper can be tricky, and there are many edge cases to consider. There is almost no circumstance in which you should try to build an escaper yourself; secure SQL escapers exist in SQL libraries for almost every programming language. However, if you are running SQL statements with raw user input, escapers are often an ineffective solution, because you need to ensure that every call is properly escaped. A far more robust solution is to use parameterized SQL.

17.5 Defense: Parameterized SQL/Prepared Statements

A better defense against SQL injection is to use parameterized SQL or prepared statements. This type of SQL compiles the query first, and then plugs in user input after the query has

already been interpreted by the SQL parser. Because the user input is added after the query is compiled and interpreted, there is no way for any attacker input to be treated as SQL code. Parameterized SQL prevents all SQL injection attacks, so it is the best defense against SQL injection!

In most SQL libraries, parameterized SQL and unsafe, non-parameterized SQL are provided as two different API functions. You can ensure that you've eliminated *all* potential SQL vulnerabilities in your code by searching for every database query and replacing each API call with a call to the parameterized SQL API function.

The biggest problem with parameterized SQL is compatibility. SQL is a (mostly) generic language, so SQL written for MySQL can run on Postgres or commercial databases. Parameterized SQL requires support from the underlying database (since the processing itself happens on the database side), and there is no common standard for expressing parameterized SQL. Most SQL libraries will handle the translation for you, but switching to prepared statements may make it harder to switch between databases.

In practice, most modern SQL libraries support parameterized SQL and prepared statements. If the library you are using does not support parameterized SQL, it is probably best to switch to a different SQL library.

Further reading: [OWASP Cheat Sheet on SQL Injection](#)

18 Intro to the Web

18.1 URLs

Every resource (webpage, image, PDF, etc.) on the web is identified by a URL (Uniform Resource Locator). URLs are designed to describe exactly where to find a piece of information on the Internet. A basic URL consists of three mandatory parts:

`http://www.example.com/index.html`

The first mandatory part is the *protocol*, located before `://` in the URL. In the example above, the protocol is `http`. The protocol tells your browser how to retrieve the resource. In this class, the only two protocols you need to know are HTTP, which we will cover in the next section, and HTTPS, which is a secure version of HTTP using TLS (refer to the networking unit for more details). Other protocols include `git+ssh://`, which fetches a git archive over an encrypted tunnel using `ssh`, or `ftp://`, which uses the old FTP (File Transfer Protocol) to fetch data.

The second mandatory part is the *location*, located after `://` but before the next forward slash in the URL. In the example above, the location is `www.example.com`. This tells your browser which web server to contact to retrieve the resource.

Optionally, the location may contain an optional *username*, which is followed by an `@` character if present. For example, `evanbot@www.example.com` is a location with a username `evanbot`. All locations must include a computer identifier. This is usually a domain name

such as `www.example.com`. Sometimes the location will also include a port number, such as `www.example.com:81`, to distinguish between different applications running on the same web server. We will discuss ports a bit more when we talk about TCP during the networking section.

The third mandatory part is the *path*, located after the first single forward slash in the URL. In the example above, the path is `/index.html`. The path tells your browser which resource on the web server to request. The web server uses the path to determine which page or resource should be returned to you.

One way to think about paths is to imagine a filesystem on the web server you're contacting. The web server can use the path as a filepath to locate a specific page or resource. The path must at least consist of `/`, which is known as the “root”¹ of the filesystem for the remote web site.

Optionally, there can be a `?` character after the path. This indicates that you are supplying additional arguments in the URL for the web server to process. After the `?` character, you can supply an optional set of *parameters* separated by `&` characters. Each parameter is usually encoded as a key-value pair in the format `key=value`. Your browser sends all this information to the web server when fetching a URL. See the next section for more details on URL parameters.

Finally, there can be an optional *anchor* after the arguments, which starts with a `#` character. The anchor text is not sent to the server, but is available to the web page as it runs in the browser.

The anchor is often used to tell your browser to scroll to a certain part of the webpage when loading it. For example, try loading https://en.wikipedia.org/wiki/Dwinelle_Hall#Floor_plan and https://en.wikipedia.org/wiki/Dwinelle_Hall#Construction and note that your browser skips to the section of the article specified in the anchor.

In summary, a URL with all elements present may look like this:

`http://evanbot@www.cs161.org:161/whoami?k1=v1&k2=v2#anchor`

where `http` is the protocol, `evanbot` is the username, `www.cs161.org` is the computer location (domain), `161` is the port, `/whoami` is the path, `k1=v1&k2=v2` are the URL arguments, and `anchor` is the anchor.

Further reading: [What is a URL?](#)

18.2 HTTP

The protocol that powers the World Wide Web is the Hypertext Transfer Protocol, abbreviated as HTTP. It is the language that clients use to communicate with servers in order to fetch resources and issue other requests. While we will not be able to provide you with a full overview of HTTP, this section is meant to get you familiar with several aspects of the protocol that are important to understanding web security.

¹It is called the root because the filesystem can be treated as a tree and this is where the tree starts.

18.2.1 The Request-Response Model

Fundamentally, HTTP follows a request-response model, where clients (such as browsers) must actively start a connection to the server and issue a request, which the server then responds to. This request can be something like “Send me a webpage” or “Change the password for my user account to `foobar`.” In the first example, the server might respond with the contents of the web page, and in the second example, the response might be something as simple as “Okay, I’ve changed your password.” The exact structure of these requests will be covered in further detail in the next couple sections.

The original version of HTTP, HTTP 1.1, is a text-based protocol, where each HTTP request and response contains a *header* with some metadata about the request or response and a *payload* with the actual contents of the request or response. HTTP2, a more recent version of HTTP, is a binary-encoded protocol for efficiency, but the same concepts apply.

For all requests, the server generates and sends a response. The response includes a series of headers and, in the payload, the body of the data requested.

18.2.2 Structure of a Request

Below is a very simple HTTP request.

```
GET / HTTP/1.1
Host: squigler.com
Dnt: 1
```

The first line of the request contains the method of the request (`GET`), the path of the request (`/`), and the protocol version (`HTTP/1.1`). This is an example of a `GET` request. Each line after the first line is a request header. In this example, there are two headers, the `DNT` header and the `Host` header. There are many HTTP headers defined in the HTTP spec which are used to convey various pieces of information, but we will only be covering a couple of them through this lab.

Here is another HTTP request:

```
POST /login HTTP/1.1
Host: squigler.com
Content-Length: 40
Content-Type: application/x-url-formencoded
Dnt: 1
```

```
username=alice@foo.com&password=12345678
```

Here, we have a couple more headers and a different request type: the `POST` request.

18.2.3 GET vs. POST

While there are quite a few methods for requests, the two types that we will focus on for this course are `GET` requests and `POST` requests. `GET` requests are generally intended for

“getting” information from the server. POST requests are intended for sending information to the server that somehow modifies its internal state, such as adding a comment in a forum or changing your password.

In the original HTTP model, GET requests are not supposed to change any server state. However, modern web applications often change server state in response to GET requests in query parameters.

Of note, only POST requests can contain a body in addition to request headers. Notice that the body of the second example request contains the username and password that the user `alice` is using to log in. While GET requests cannot have a body, it can still pass query parameters via the URL itself. Such a request might look something like this:

```
GET /posts?search=security&sortby=popularity
Host: squigler.com
Dnt: 1
```

In this case, there are two query parameters, `search` and `sortby`, which have values of `security` and `popularity`, respectively.

18.3 Elements of a Webpage

The HTTP protocol is designed to return arbitrary files. The response header usually specifies a [media type](#) that tells the browser how to interpret the data in the response body.

Although the web can be used to return files of any type, much of the web is built in three languages that provide functionality useful in web applications.

A modern web page can be thought of as a distributed application: there is a component running on the web server and a component running in the web browser. First, the browser makes an HTTP request to a web server. The web server performs some server-side computation and generates and sends an HTTP response. Then, the browser performs some browser-side computation on the HTTP response and displays the result to the user.

18.3.1 HTML

HTML (Hypertext Markup Language) lets us create structured documents with paragraphs, links, fillable forms, and embedded images, among other features. You are not expected to know HTML syntax for this course, but some basics are useful for some of the attacks we will cover.

Here are some examples of what HTML can do:

- Create a link to Google: `Click me`
- Embed a picture in the webpage: ``
- Include JavaScript in the webpage: `<script>alert(1)</script>`
- Embed the CS161 webpage in the webpage: `<iframe src="http://cs161.org"></iframe>`

Frames pose a security risk, since the outer page is now including an inner page that may be from a different, possibly malicious source. To protect against this, modern browsers enforce frame isolation, which means the outer page cannot change the contents of the inner page, and the inner page cannot change the contents of the outer page.

18.3.2 CSS

CSS (Cascading Style Sheets) lets us modify the appearance of an HTML page by using different fonts, colors, and spacing, among other features. You are not expected to know CSS syntax for this course, but you should know that CSS is as powerful as JavaScript when used maliciously. If an attacker can force a victim to load some malicious CSS, this is functionally equivalent to the attacker forcing the victim to load malicious JavaScript.

18.3.3 JavaScript

JavaScript is a programming language that runs in your browser. It is a very powerful language—in general, you can assume JavaScript can arbitrarily modify any HTML or CSS on a webpage. Webpages can include JavaScript in their HTML to allow for dynamic features such as interactive buttons. Almost all modern webpages use JavaScript.

When a browser receives an HTML document, it first converts the HTML into an internal form called the DOM (Document Object Model). The JavaScript is then applied on the DOM to modify how the page is displayed to the user. The browser then renders the DOM to display the result to the user.

Because JavaScript is so powerful, modern web browsers run JavaScript in a sandbox so that any JavaScript code loaded from a webpage cannot access sensitive data on your computer or even data on other webpages.

Most exploits targeting the web browser itself require JavaScript, either because the vulnerability lies in the browser's JavaScript engine, or because JavaScript is used to shape the memory layout of the program for improving the success rate of an attack.

Almost all web browsers implement JavaScript as a Just In Time compiler, dynamically converting JavaScript into machine code². Many modern desktop applications (notably Slack's desktop client) are actually written in the Electron framework, which is effectively a cut down web browser running JavaScript.

19 Same-Origin Policy

Browsing multiple webpages poses a security risk. For example, if you have a malicious website (www.evil.com) and Gmail (www.gmail.com) open, you don't want the malicious website to be able to access any sensitive emails or send malicious emails with your identity.

²Trivia: Running JavaScript fast is considered so important that ARM recently introduced a dedicated instruction, FJCVTZS (Floating-point Javascript Convert to Signed fixed-point, rounding toward Zero), specifically to handle how JavaScript's math operates.

Modern web browsers defend against these attacks by enforcing the same-origin policy, which isolates every webpage in your browser, except for when two webpages have the same origin.

19.1 Origins

The origin of a webpage is determined by its protocol, domain name, and port. For example, the following URL has protocol `http`, domain name `www.example.com`, and port 80.

`http://www.example.com/index.html`

To check if two webpages have the same origin, the same-origin policy performs string matching on the protocol, domain, and port. Two websites have the same origin if their protocols, domains, and ports all exactly match.

Some examples of the same origin policy:

- `http://wikipedia.org/a/` and `http://wikipedia.org/b/` have the same origin. The protocol (`http`), domain (`wikipedia.org`), and port (none), all match. Note that the paths are not checked in the same-origin policy.
- `http://wikipedia.org` and `http://www.wikipedia.org` do not have the same origin, because the domains (`wikipedia.org` and `www.wikipedia.org`) are different.
- `http://wikipedia.org` and `https://wikipedia.org` do not have the same origin, because the protocols (`http` and `https`) are different.
- `http://wikipedia.org:81` and `http://wikipedia.org:82` do not have the same origin, because the ports (81 and 82) are different.

If a port is not specified, the port defaults to 80 for `http` and 443 for `https`. This means `http://wikipedia.org` has the same origin as `http://wikipedia.org:80`, but it does not have the same origin as `http://wikipedia.org:81`.

19.2 Exceptions

In general, the origin of a webpage is defined by its URL. However, there are a few exceptions to this rule:

- JavaScript runs with the origin of the page that loads it. For example, if you include `<script src="http://google.com/tracking.js"></script>` on `http://cs161.org`, the script has the origin of `http://cs161.org`.
- Images have the origin of the page that it comes from. For example, if you include `` on `http://cs161.org`, the image has the origin of `http://google.com`. The page that loads the image (`http://cs161.org`) only knows about the image's dimensions when loading it.
- Frames have the origin of the URL where the frame is retrieved from, not the origin of the website that loads it. For example, if you include

`<iframe src="http://google.com"></iframe>` on `http://cs161.org`, the frame has the origin of `http://google.com`.

JavaScript has a special function, `postMessage`, that allows webpages from different origins to communicate with each other. However, this function only allows very limited functionality.

Further reading: [Same-origin policy](#)

20 Cookies and Session Management

HTTP is a stateless protocol, which means each request and response is independent from all other requests and responses. However, many features on the web require maintaining some form of state. For example, when you log into your email account, you can stay logged in across many requests and responses. If you enable dark mode on a website and make subsequent requests to the website, you want the pages returned to have a dark background. If you're browsing an online shopping website, you want the items in your cart to be saved across many requests and responses. Browser and servers store HTTP cookies to support these features.

At a high level, you can think of cookies as pieces of data stored in your browser. When you make a request to enable dark mode or add an item to your shopping cart, the server sends a response with a **Set-Cookie** header, which tells your browser to store a new cookie. These cookies encode state that should persist across multiple requests and responses, such as your dark mode preference or a list of items in your shopping cart. In future requests, your browser will automatically attach the relevant cookies to a request and send it to the web server. The additional information in these cookies helps the web server customize its response.

20.1 Cookie Attributes

Every cookie is a name-value pair. For example, a cookie `darkmode=true` has name `darkmode` and value `true`.

For security and functionality reasons, we don't want the browser to send every cookie in every request. A user might want to enable dark mode on one website but not on another website, so we need a way to only send certain cookies to certain URLs. Also, as we'll see later, cookies may contain sensitive login information, so sending all cookies in all requests poses a security risk. These additional cookie attributes help the browser determine which cookies should be attached to each request.

- The **Domain** and **Path** attributes tell the browser which URLs to send the cookie to. See the next section for more details.
- The **Secure** attribute tells the browser to only send the cookie over a secure HTTPS connection.
- The **HttpOnly** attribute prevents JavaScript from accessing and modifying the cookie.

- The `expires` field tells the browser when to stop remembering the cookie.

20.2 Cookie Policy: Domain and Path

The browser sends a cookie to a given URL if the cookie's `Domain` attribute is a domain-suffix of the URL domain, and the cookie's `Path` attribute is a prefix of the URL path. In other words, the URL domain should end in the cookie's `Domain` attribute, and the URL path should begin with the cookie's `Path` attribute.

For example, a cookie with `Domain=example.com` and `Path=/some/path` will be included on a request to `http://foo.example.com/some/path/index.html`, because the URL domain ends in the cookie domain, and the URL path begins with the cookie path.

Note that cookie policy uses a different set of rules than the same origin policy. This has caused problems in the past.

20.3 Cookie Policy: Setting Domain and Path

For security reasons, we don't want a malicious website `evil.com` to be able to set a cookie with domain `bank.com`, since this would allow an attacker to affect the functionality of the legitimate bank website. To prevent this, the cookie policy specifies that when a server sets a cookie, the cookie's domain must be a URL suffix of the server's URL. In other words, for the cookie to be set, the server's URL must end in the cookie's `Domain` attribute. Otherwise, the browser will reject the cookie.

For example, a webpage with domain `eecs.berkeley.edu` can set a cookie with domain `eecs.berkeley.edu` or `berkeley.edu`, since the webpage domain ends in both of these domains.

This policy has one exception: cookies cannot have domains set to a top-level domain, such as `.edu` or `.com`, since these are too broad and pose a security risk. If `evil.com` could set cookies with domain `.com`, the attacker would have the ability to affect all `.com` websites, since this cookie would be sent to all `.com` websites. The web browser maintains a list of top-level domains, which includes two-level TLDs like `.co.uk`.

The cookie policy allows a server to set the `Path` attribute without any restrictions.³

Further reading: [Cookies](#)

20.4 Session Management

Cookies are often used to keep users logged in to a website over many requests and responses. When a user sends a login request with a valid username and password, the server will generate a new session token and send it to the user as a cookie. In future requests, the

³The lack of restriction on the `Path` attribute has caused problems in the past, as cookies are presented to the server and JavaScript as an unordered set of name/value pairs, but is stored internally as name/-path/value tuples, so if two cookies with the same name and host but different path are present, both will be presented to the server in unspecified order.

browser will attach the session token cookie and send it to the server. The server maintains a mapping of session tokens to users, so when it receives a request with a session token cookie, it can look up the corresponding user and customize its response accordingly.

Secure session tokens should be random and unpredictable, so an attacker cannot guess someone else's session token and gain access to their account. Many servers also set the `HttpOnly` and `Secure` flags on session tokens to protect them from being accessed by XSS vulnerabilities or network attackers, respectively.

It is easy to confuse session tokens and cookies. Session tokens are the values that the browser sends to the server to associate the request with a logged-in user. Cookies are how the browser stores and sends session tokens to the server. Cookies can also be used to save other state, as discussed earlier. In other words, session tokens are a special type of cookie that keep users logged in over many requests and responses.

21 Cross-Site Request Forgery (CSRF)

Using cookies and session tokens to keep a user logged in has some associated security risks. In a cross-site request forgery (CSRF) attack, the attacker forces the victim to make an unintended request. The victim's browser will automatically attach the session token cookie to the unintended request, and the server will accept the request as coming from the victim.

For example, suppose a website has an endpoint `http://example.com/logout`. To log out, a user makes a GET request to this URL with the appropriate session token attached, and the server checks the session token and performs the logout. If an attacker can trick a victim into clicking this link, the victim will be logged out of the website without their knowledge.

CSRF attacks can also be executed on URLs with more malicious actions. For example, a GET request to `https://bank.com/transfer?amount=100&recipient=mallory` with a valid session token might send \$100 to Mallory. An attacker could send an email to the victim with the following HTML snippet:

```

```

This will cause the browser to try and fetch an image from the malicious URL by making a GET request. Because the browser automatically attaches the session token to the request, this causes the victim to unknowingly send \$100 to Mallory.

It is usually bad practice to have HTTP GET endpoints that can change server state, so this type of CSRF attack is less common in practice. However, CSRF attacks are still possible over HTTP POST requests. HTML forms are a common example of a web feature that generates HTTP POST requests. The user fills in the form fields, and when they click the Submit button, the browser generates a POST request with the filled-out form fields. Consider the following HTML snippet on an attacker's webpage:

```
<form name=evilform action=https://bank.com/transfer>
  <input name=amount value=100>
  <input name=recipient value=mallory>
```

```
</form>
<script>document.evilmform.submit();</script>
```

When the victim visits the attacker’s website, this HTML snippet will cause the victim’s browser to make a POST request to `https://bank.com/transfer` with form input values that transfer \$100 to Mallory. Like before, the victim’s browser automatically attaches the session token to the request, so the server accepts this POST request as if it was from the victim.

21.1 Defense: CSRF Token

A good defense against CSRF attacks is to include a CSRF token on webpages. When a legitimate user loads a webpage from the server with a form, the server will randomly generate a CSRF token and include it as an extra field in the form. (In practice, this field often has a hidden attribute set so that it’s only visible in the HTML, so users don’t see random strings every time they submit a form.) When the user submits the form, the form will include the CSRF token, and the server will check that the CSRF token is valid. If the CSRF token is invalid or missing, the server will reject the request.

To implement CSRF tokens, the server needs to generate a new CSRF token every time a user requests a form. CSRF tokens should be random and unpredictable so an attacker cannot guess the CSRF token. The server also needs to maintain a mapping of CSRF tokens to session tokens, so it can validate that a request with a session token has the correct corresponding CSRF token. This may require the server to store a large amount of state if it expects heavy traffic.

If an attacker tries the attack in the previous section, the malicious form they create on their website will no longer contain a valid CSRF token. The attacker could try querying the server for a CSRF token, but it would not properly map to the victim’s session token, because the victim never requested the form legitimately.

21.2 Defense: Referer Validation

Another way to defend against CSRF tokens is to check the Referer⁴ field in the HTTP header. When a browser issues an HTTP request, it includes a Referer header which indicates which URL the request was made from. For example, if a user fills out a form from a legitimate bank website, the Referer header will be set to `bank.com`, but if the user visits the attacker’s website and the attacker fills out a form and submits it, the Referer header will be set to `evil.com`. The server can check the Referer header on each request and reject any requests that have untrusted or suspicious Referer headers.

Referer validation is a good defense if it is included on every request, but it poses some problems if someone submits a request with the Referer header left blank. If a server accepts

⁴Yes, the “Referer” field represents a roughly three decade old misspelling of referrer. This is a silly example of how “legacy”, that is old design decisions, can impact things decades later because it can be very hard to change things.

requests with blank Referer headers, it may be vulnerable to CSRF attacks, but if a server rejects requests with blank Referer headers, it may reduce functionality for some users.

In practice, Referer headers may be removed by the browser, the operating system, or a network monitoring system for privacy issues. For example, if you click on a link to visit a website from a Google search, the website can know what Google search you made to visit its website from the Referer header. Some modern browsers also have options that let users disable sending the Referer header on all requests. Because not all requests are guaranteed to have a Referer header, it is usually only used as a defense-in-depth strategy in addition to CSRF tokens, instead of as the only defense against CSRF attacks.

Further reading: [OWASP Cheat Sheet on CSRF](#)

22 Cross-Site Scripting (XSS)

XSS is a class of attacks where an attacker injects malicious JavaScript onto a webpage. When a victim user loads the webpage, the user's browser will run the malicious JavaScript.

XSS attacks are powerful because they subvert the same-origin policy. Normally, an attacker can only run JavaScript on websites they control (such as `https://evil.com`), so their JavaScript cannot affect websites with origins different from `https://evil.com`. However, if the attacker can inject JavaScript into `https://google.com`, then when a user loads `https://google.com`, their browser will run the attacker's JavaScript with the origin of `https://google.com`.

XSS attacks allow malicious JavaScript to run in the user's browser with the same origin as a legitimate website. This allows the attacker to perform any action the user can perform at `https://google.com` or steal any user secrets associated with Google and send them back to the attacker.

There are two main categories of XSS attacks: stored XSS and reflected XSS.

22.1 Stored XSS

In a stored XSS attack, the attacker finds a way to persistently store malicious JavaScript on the web server. When the victim loads the webpage, the web server will load this malicious JavaScript and display it to the user.

A classic example of stored XSS is a Facebook post. When a user makes a Facebook post, the contents of the post are stored on Facebook's servers, so that other users can load their friends' posts. If Facebook doesn't properly check user inputs, an attacker could make a post that says

```
<script>alert("XSS attack!")</script>
```

This post is now stored in Facebook's servers. If another user loads the attacker's posts, they will receive an HTML page with this script on it, and the browser will run the script and trigger a pop-up that says **XSS attack!**

22.2 Reflected XSS

In a reflected XSS attack, the attacker finds a vulnerable webpage where the server receives user input in an HTTP request and displays the user input in the response.

A classic example of reflected XSS is a Google search. When you make an HTTP GET request for a Google search, such as `https://www.google.com/search?&q=cs161`, the returned webpage with search results will include something like

You searched for: `cs161`

If Google does not properly check user input, an attacker could create a malicious URL `https://www.google.com/search?&q=<script>alert("XSS attack!")</script>`. When the victim loads this URL, Google will return

You searched for: `<script>alert("XSS attack!")</script>`

The victim's browser will run the script and trigger a pop-up that says `XSS attack!!`

22.3 Defense: Sanitize Input

A good defense against XSS is checking for malicious input that might cause JavaScript to run, such as `<script>` tags. However, it is very difficult to write a good detector that catches all XSS attacks. For example, the following input causes JavaScript to run without ever using `<script>` tags:

```
<img src=1 href=1 onerror="JavaScript:alert("XSS attack!")" />
```

Just like SQL input escaping, sanitizing potentially dangerous input can be very tricky. For example, consider an escaper that searches for all instances of `<script>` and `</script>` and removes them. An attacker could provide this malicious input:

```
<scr<script>ipt>alert("XSS attack!")</scr<script>ipt>
```

After the escaper removes the two `<script>` tags it sees, the result is `<script>alert("XSS attack!")</script>`, and the attacker can still execute JavaScript!

Another way to escape input is to replace potentially dangerous characters with their HTML encoding. For example, the less than (`<`) and greater than (`>`) signs are encoded as `<`; and `>`; , respectively. These encodings cause less than and greater than signs to display on the webpage, without being interpreted as HTML.

Fortunately, there is a standardized set of sanitizations that is known to be robust.

22.4 Defense: Content Security Policy

Another XSS defense is using a content security policy (CSP) that specifies a list of allowed domains where scripts can be loaded from. For example, `cs161.org` might allow scripts that are loaded from `*.cs161.org` or `*.google.com` and disallow all other scripts, including any inline scripts that are injected by the attacker.

CSPs are defined by a web server and enforced by a browser. In the HTTP response, the server attaches a **Content-Security-Policy** header, and the browser checks any scripts against the header.

If you enable CSP, you can no longer run *any* scripts that are embedded directly in the HTML document. You can only load external scripts specified by the `script` tag and an external URL. These scripts can only be fetched from the sites specified in the CSP. This prevents an attacker from directly injecting scripts into an HTML document or modifying the HTML document to fetch scripts from the attacker's domain.

Further reading: [OWASP Cheat Sheet on XSS](#)

23 Clickjacking/User Interface (UI) Attacks

Many of the web attacks we've seen involve forcing a victim to click on an attacker-generated link (reflected XSS), or forcing a victim to visit an attacker-controlled website (CSRF). How might an attacker achieve this?

UI attacks (or *clickjacking attacks*) are a category of attacks that try to fool a victim into inadvertently clicking on an attacker-supplied input. The end goal of these attacks is to “steal” a click from the user, so that the user loads something controlled by the attacker (possibly for a further attack). Many UI attacks rely on clever visual tricks to deceive the user.

Download buttons are a classic example of clickjacking. When you visit a website to download a file, you might see many different download buttons with different shapes and colors. One of these is the true download button, and the others are malicious download buttons that actually take you to attacker-controlled websites or perform other malicious actions in your browser. An unwitting user might click on the wrong download button and be sent to the attacker website. The malicious download buttons could be added to the website through a different web exploit (e.g. stored XSS) or as a paid advertisement.

Depending on how much control the attacker has over the page, more sophisticated clickjacking attacks are possible:

- The attacker could manipulate an HTML form so that the user sees a payment of \$5, but the underlying form will actually submit a payment of \$50.
- The attacker could draw a fake cursor on the page. The user sees the fake cursor over a legitimate button and clicks, but their real cursor has actually clicked on a malicious link.
- The attacker could draw an entire browser on the page. The user sees an address bar and clicks, but they have actually clicked on a fake address bar generated by the attacker (with a malicious link behind the address bar).

23.1 Clickjacking Defenses

There are many ways to defend against clickjacking attacks. The general idea is to force the user to make sure that they're clicking on what they intended to click.

Confirmation pop-ups: If the user clicks on a link or button that will perform some potentially dangerous activity (e.g. opening a website, executing Javascript, downloading a file), display a pop-up asking the user to confirm that this is their intended action. However, users might still click on the pop-up without reading it, especially if they're too frequent. Remember to consider human factors!

UI randomization: Randomize the location of certain elements on a website. For example, a submit button could alternately be located at the left side of the screen and the right side of the screen. This makes it harder for attackers to draw a fake submit button over the real submit button, because they won't know where it's located. However, webpages that look different every time could pose usability problems.

Direct the user's attention to their click: This can be done by freezing the rest of the screen besides the area directly around the user's cursor, or by highlighting the user's cursor. This will make the user less likely to be fooled by a fake cursor and force them to focus on where their real cursor is pointing. The user's clicks can also be invalidated if the user tries to click outside of a relevant portion of the screen.

Delay the click: Force the user to hover over the desired button for some amount of time before allowing the user to click the button. This forces the user to spend some time looking at where they're clicking before they actually perform the click.

24 CAPTCHAs

Consider the following scenario: you've created a website that allows users to upload a picture. Your server will scan the picture for text using a compute-intensive algorithm and return the text to the user. An adversary wants to mount a denial-of-service (DoS) attack on your website by uploading lots of bogus images, forcing your server to run the expensive algorithm on all the bogus images.

Consider another scenario: Your website has a login page. An adversary wants to steal a legitimate user's account, so they try to brute-force the user's password by submitting login requests with every possible password.

Generally, when we're building websites, we'd like to build websites for people: we don't want robots. CAPTCHAs are a test that ask the fundamental question: *Is this a human?* Consequently, when we design CAPTCHAs, we want to choose problems that are easy for humans, but difficult for computers.

CAPTCHAs are primarily focused on machine vision problems, which are traditionally difficult for computers to solve. Historically, CAPTCHAs consist of a series of distorted letters or words. There are a wide variety of CAPTCHAs: some with color, some with low contrasts, some with merged-together letters, etc. A more recent example you may be familiar with

is Google's reCAPTCHA algorithm, which shows you some images and asks you to identify the objects in the pictures (e.g. "Select all images with boats.")

24.1 Issues with CAPTCHAs

There's an inherent arms race present here: as solving algorithms get better, our defense deteriorates. The reason why CAPTCHAs have gotten so much harder over the last decade is because individuals have spent time creating much better solving algorithms - and we're reaching a point where it's becoming more and more difficult for humans to solve CAPTCHAs quickly.

Of course, those implementing CAPTCHAs often miss the original motivation behind their development. The original CAPTCHA paper included the subtitle "How Lazy Cryptographers do AI" as the intent was to force attackers to solve harder problems in machine vision. Now modern CAPTCHAs such as Google ReCAPTCHA are focused on getting humans to provide training data for AI systems which means the CAPTCHAs are inherently self-defeating for those deploying the CAPTCHA.

In some cases, it's necessary to provide an alternative, accessible CAPTCHA method, such as an audio-based spoken phrase that a human is required to transcribe. In this case, we've unintentionally opened up a new attack vector: attackers may now break target the audio-based CAPTCHA, which may be easier to solve than the traditional image-based CAPTCHA.

If you search "crack CAPTCHA" on Google, you'll likely find many CAPTCHA solving services for as low as \$0.10 cents per CAPTCHA. These services use humans to do the actual work. These days, a CAPTCHA no longer asks the question of "Is this a human or a bot?" Instead, it says "Is this a human, or a bot willing to spend a fraction of a penny?"

The takeaway: if something is worth \$0.10 or more to an attacker, CAPTCHAs do not work.