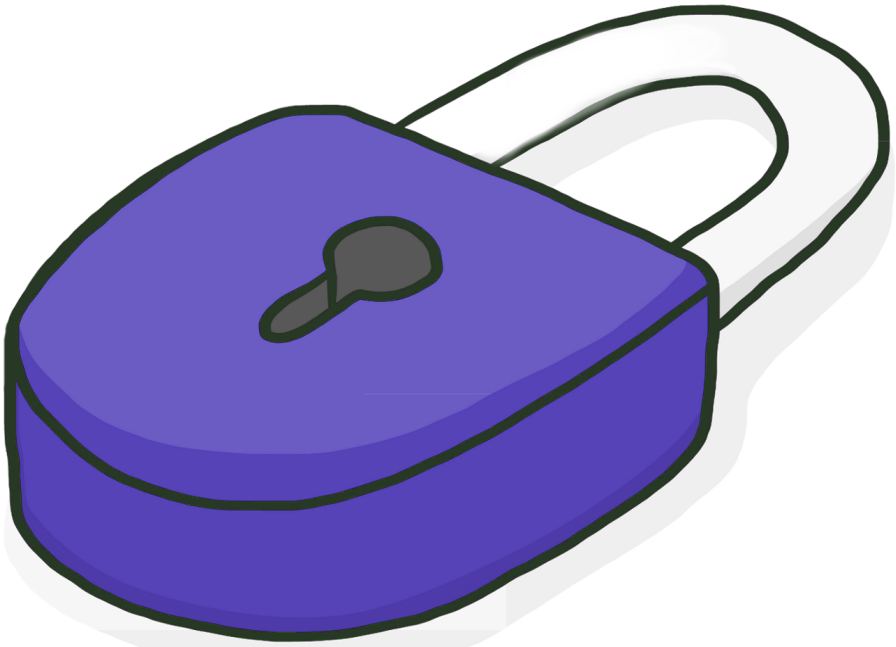
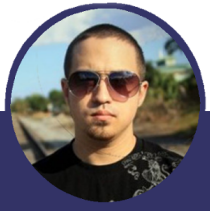
BUILDING SECURE PHP APPS





a practical guide

Ben Edmunds

Building Secure PHP Apps

is your PHP app truly secure? Let's make sure you get home on time and sleep well at night.

Ben Edmunds

This book is for sale at
http://leanpub.com/buildingsecurephpapps

This version was published on 2014-05-05

This is a Leanpub book. Leanpub empowers authors and publishers with the Lean Publishing process. Lean Publishing is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

Tweet This Book!

Please help Ben Edmunds by spreading the word about this book on Twitter!

The suggested hashtag for this book is #buildingsecurephpapps.

Find out what other people are saying about the book by clicking on this link to search for this hashtag on Twitter:

https://twitter.com/search?q=#buildingsecurephpapps

Contents

Constructor

Several years ago I was writing a web application for a client in the CodeIgniter PHP framework, *shudder*, but CodeIgniter didn't include any type of authentication system built in. I, of course, did what any good/lazy developer would do and went on the hunt for a well made library to supply authentication capabilities. To my chagrin I discovered that there weren't any clean, concise libraries that fit my needs for authentication in CodeIgniter. Thus began my journey of creating Ion Auth, a simple authentication library for CodeIgniter, and a career long crusade for securing web applications as well as helping other developers do the same.

Here we are years later, a lot of us have moved on to other frameworks or languages, but I still repeatedly see basic security being overlooked. So let's fix that. I want to make sure that you'll never have to live the horror of leaking user passwords, have someone inject malicious SQL into your database, or the suite of other "hacks" that could have been easily avoided. Let's make sure we all get home on time and sleep well at night.

This book will be a quick read with handbook style references to specific items you can act on. It is meant to be something you can read in a couple hours and then reference later as needed. I'll also try to make sure we have some fun in the process.

Format

All code samples in the indented blocks can be assumed to be in PHP unless otherwise noted.

Lines starting with a dollar sign

```
$ ls -al
```

are examples of using the command line as a normal user.

Lines starting with a pound sign

```
# ls -al
```

are examples of using the command line as the root user.

Server command line examples will assume some type of *nix (centos, redhat, ubuntu, osx, etc) operating system.

I'm trying to keep the code examples from wrapping where possible so method arguments will be on their own lines. This may seem odd but it is much easier to read than wrapped code with this book format.

Errata

If you find any errors don't hesitate to get in touch with me via email[1].

[1] feedback@buildsecurephpapps.com

Sample Code

All of the examples are in PHP unless otherwise noted. I will use native PHP code where possible, even if it creates more boilerplate. If something requires too much work to succinctly explain in native PHP I will use the Laravel framework since it has an elegant syntax and should be easy to understand.

Some of the code examples are broken up for explanation. To view complete code examples you can reference the Github repository[2].

Let's do this.

[2] https://github.com/benedmunds/Building-Secure-PHP-Apps-Examples

About the Author

Ben Edmunds[3] leads development teams to create cutting-edge web and mobile applications. He is an active leader, developer, and speaker in various development communities. He has been developing software professionally for over 10 years and in that time has worked on everything from robotics to government projects.

PHP Town Hall podcast co-host. Portland PHP Usergroup co-organizer. Open source advocate.

[3] http://benedmunds.com

Chapter 1 - Never Trust Your Users. Sanitize ALL Input!

Let's start with a story. Mike is the system admin for a small private school in Oklahoma. His main responsibility is keeping the network and computers working. Recently he started automating various tasks around the school by building a web application for internal use. He doesn't have any formal training and just started programming about a year ago, but he feels pretty good about his work. He knows the basics of PHP and has built a pretty stable customer relationship manager for the school. There are still a ton of features to add, but the basics are covered. Mike even received kudos from the superintendent for streamlining operations and saving the school money.

Everything was going well for Mike until a particular new student started. The student's name is Little Bobby Tables[4]. One day, Jon from the admin office called Mike to ask why the system was down. After inspecting, Mike found that the table containing all the students' information was missing entirely. You see, Little Bobby's full name is actually "Robert'); DROP TABLE students;–". There aren't any backups of the database; it has been on Mike's "to do" list for a while, but he hadn't gotten around to it yet. Mike is in big trouble.

[4] http://xkcd.com/327/

SQL Injection

Real World

While it's unlikely a real child's name will contain damaging SQL code, this kind of **SQL injection vulnerability** happens in the real world all the time:[5]

- In 2012, LinkedIn leaked over 6 million users' data due to an undisclosed SQL injection vulnerability
- In 2012, Yahoo! exposed 450,000 user passwords
- In 2012, 400,000 passwords were compromised from Nvidia
- In 2012, 150,000 passwords were compromised from Adobe
- In 2013, eHarmony had roughly 1.5 million user passwords exposed

How SQL Injection Works

If you use input directly from your users without modification, a malicious user can pass unexpected data, and fundamentally change your SQL queries.

If your code looks something like this:[6]

[5]For most of these precise details were undisclosed, so we can't be certain these were due to SQL injection attacks. Chances are the majority were though.

[6]The `mysql_*` extension and it's methods are officially deprecated. Please don't use them.

```
mysql_query('UPDATE users
  SET first_name="' . $_POST['first_name'] . '"
  WHERE id=1001');
```

You would expect the generated SQL to be:

```
UPDATE users set first_name="Liz" WHERE id=1001;
```

But if your malicious user types their first name as:

```
Liz", last_name="Lemon"; --
```

The generated SQL then becomes:

```
UPDATE users
SET first_name="Liz", last_name="Lemon"; --"
WHERE id=1001;
```

Now all of your users are named Liz Lemon, and that's just not cool.

How To Guard Against It

The single requirement for guarding against SQL injection is to **sanitize input** (also known as **escaping**). You can escape each input individually, or use a better method known as **parameter binding**. Parameter binding is definitely the way I recommend, as it offers more security. Using PHP's `PDO` class[7], your code now becomes:

[7] http://us1.php.net/manual/en/intro.pdo.php

```
$db = new PDO(...);
$query = $db->prepare('UPDATE users
  SET first_name = :first_name
  WHERE id = :id');

$query->execute([
  ':id'         => 1001,
  ':first_name' => $_POST['first_name']
]);
```

Using bound parameters means that each value will be escaped, quoted properly, and only one value is expected. Keep in mind, bound parameters protect your query, but they don't protect the input data after it enters your database. Remember, *any* data can be malicious. You will still need to strip out and/or escape data that will be displayed back to the user. You can do this when you save the data to the database, or when you output it, but don't skip this very important step. We'll cover this more in the "Sanitizing Output" section coming up.

Your code is now a little longer, but it's safe. You won't have to worry about another Little Bobby Tables screwing up your day. Bound parameters are pretty awesome right? You know what else is awesome, Funyuns are awesome.

Best Practices and Other Solutions

Stored procedures are another way to protect against SQL injection. A stored procedure is a function built in your database. Using a stored procedure means you're less likely to be susceptible to SQL injection, since your data isn't passed

directly as SQL. In general, stored procedures are frowned upon. The main reasons for which include:

1. Stored procedures are difficult to test
2. They move the logic to another system outside of the application
3. They are difficult to track in your version control system, since they live in the database and not in your code
4. Using them can limit the number of people on your team capable of modifying the logic if needed

Client-side JavaScript is NOT a solution for validating data, ever. It can be easily modified or avoided by a malicious user with even a mediocre amount of knowledge. Repeat after me: I will NEVER rely on JavaScript validation; I will NEVER EVER rely on JavaScript validation. You can certainly use JavaScript validation to provide instant feedback and present a better user experience, but for the love of your favorite deity, check the input on the back end to make sure everything is legit.

Mass Assignment

Mass assignment can be an incredibly useful tool that can speed up development time, or cause severe damage if used improperly.

Let's say you have a `User` model that you need to update with several changes. You could update each field individually, or you could pass all of the changes from a form and update it in one go.

Your form might look like this:

```
<form action="...">
  <input name="first_name" />
  <input name="last_name" />
  <input name="email" />
</form>
```

Then you have back end PHP code to process and save the form submission. Using the Laravel framework, that might look like this:

```
$user = User::find(1);
$user->update(Input::all());
```

Quick and easy right? But what if a malicious user modifies the form, giving themselves administrator permissions?

```
<form action="...">
  <input type="text" name="first_name" />
  <input type="text" name="last_name" />
  <input type="text" name="email" />
  <input type="hidden" name="permissions" value="{\
'admin':'true'}" />
</form>
```

That same code would now change this user's permissions erroneously.

This may sound like a dumb problem to solve, but it is one that a lot of developers and sites have fallen victim to. The most recent, well-known exploit of this vulnerability was when a user exposed that Ruby on Rails was susceptible to this. When Egor Homakov originally reported to the Rails team that new Rails installs were insecure, his bug report was rejected. The core team thought it was a minor concern that would be easier for new developers to leave enabled by default. To get attention to this issue, Homakov hilariously "hacked" Rails' GitHub account (GitHub is built on Rails) to give himself administrative rights to their repositories. Needless to say, this proved his point, and now Rails (and GitHub) are protected from this attack by default.

How do you protect your application against this? The exact implementation details depend on which framework or code base you're using, but you have a few options:

- Turn off mass assignment completely
- Whitelist the fields that are safe to be mass assigned
- Blacklist the fields that are not safe to be mass assigned

Depending on your implementation, some of these may be used simultaneously.

In Laravel you add a `$fillable` property to your models to set the whitelist of fields that are mass assignable:

```
class User extends Eloquent {

  protected $table = 'users';

  protected $fillable = ['first_name', 'last_name'\
, 'email'];
```

This would stop the "permissions" column from being mass assigned. Another way to handle this in Laravel is to set a blacklist with the `$guarded` property:

```
class User extends Eloquent {

  protected $table = 'users';

  protected $guarded = ['permissions'];
```

The choice is up to you, depending on which is easier in your application.

If you don't use Laravel, your framework probably has a similar method of whitelisting/blacklisting mass assignable fields. If you use a custom framework, get on implementing whitelists and blacklists!

Typecasting

One additional step I like to take, not just for security but also for data integrity, is to typecast known formats. Since PHP is a dynamically typed language[8], a value can be any type: string, integer, float, etc. By typecasting the value, we can verify that the data matches what we expect. In the previous example, if the ID was coming from a variable it would make sense to typecast it if we knew it should always be an integer, like this:

```
1  $id = (int) 1001;
2
3  $db    = new PDO(...);
4  $query = $db->prepare('UPDATE users
5    SET first_name = :first_name
6    WHERE id = :id');
7
8  $query->execute([
9    ':id'         => $id, //we know its an int
10   ':first_name' => $_POST['first_name']
11 ]);
```

In this case it wouldn't matter much since we are defining the ID ourselves, so we know its an integer. But if the ID came from a posted form or another source, this would give us additional peace of mind.

PHP supports a number of types that you can cast to, they are

[8] http://stackoverflow.com/questions/7394711/what-is-dynamic-typing

```
$var = (array) $var;
$var = (binary) $var;
$var = (bool) $var;
$var = (boolean) $var;
$var = (double) $var;
$var = (float) $var;
$var = (int) $var;
$var = (integer) $var;
$var = (object) $var;
$var = (real) $var;
$var = (string) $var;
```

This is helpful not only when dealing with your database, but throughout your application. Just because PHP is dynamically typed doesn't mean that you can't enforce typing in certain places. Yeah science!

Sanitizing Output

Outputting to the Browser

Not only should you take precautions when saving the data you take in, you should sanitize / escape any user-generated data that is output back to the browser.

You can modify and escape your data prior to saving to the database, or in between retrieving it and outputting to the browser. It usually depends on how your data is edited and used. For example, if the user is editing the data later, it usually makes more sense to save it as-is, and sanitize upon output.

What security benefits come from escaping user-generated data that you output? Suppose a user submits the following JavaScript snippet to your application, which saves it for outputting later:

```
<script>alert('I am not sanitized!');</script>
```

If you don't sanitize this code before you echo it out to the browser, the malicious JavaScript will run normally, as if you wrote it yourself. In this case it's a harmless `alert()`, but a hacker won't be nearly as kind.

Another popular place for this type of exploit is in an image's XIFF data. If a user uploads an image and your application displays the XIFF data, it will need to be sanitized as well. Anywhere you are displaying data that came into your app from the outside, you need to sanitize it.

If you're using a templating library or a framework that handles templating, escaping may happen automatically, or

there is a built-in method for doing so. Make sure to check the documentation for your library / framework of choice to determine how this works.

For those of you handling this yourself, PHP provides a couple of functions that will be your best friends when displaying data in the browser: `htmlentities()`[9] and `htmlspecialchars()`[10]. Both will escape and manipulate data to make it safer before rendering.

`htmlspecialchars()` should be your go-to function in 90% of cases. It will look for characters with special meaning (e.g., `<`, `>`, `&`) and encode these characters to HTML entities.

`htmlentities()` is like `htmlspecialchars()` on steroids. It will encode any character into its HTML entity equivalent if one exists. This may or may not be what you need in many cases. Make sure to understand what each one of these functions does exactly, then evaluate which is best for the type of data you are sending to the browser.

[9] http://us1.php.net/htmlentities

[10] http://us1.php.net/htmlspecialchars

Echoing to the Command Line

Don't forget to sanitize the output of any command line script you are running. The functions for this are `escapeshellcmd()`[11] and `escapeshellarg()`[12].

They are both pretty self-explanatory. Use `escapeshellcmd()` to escape any commands that you are calling. This will prevent arbitrary commands from being executed. `escapeshellarg()` is used to wrap arguments to ensure they are escaped correctly, and don't open your application up to manipulating the structure of the commands.

[11] http://us1.php.net/escapeshellcmd

[12] http://us1.php.net/escapeshellarg

Chapter Two - HTTP-S/SSL/BCA/JWH/SHA and Other Random Letters; Some of Them Actually Matter.

Once again, it's time for a little story. In October 2010 Eric Butler released a Firefox extension named Firesheep to highlight a huge problem on the web that most people hadn't been paying enough attention to. Firesheep allowed any regular ol' user to watch the non-encrypted traffic on their local network and then hijack other user's sessions. Firesheep exploits a type of man in the middle attack, sidejacking. Sound scary? It should, because it is. Maybe you're thinking, well this is conjecture. Alright fine, facts in. Let's walk through an illustration to make the point.

It's December 2010, Jane is out of town on a work trip for Achme Inc and is staying at a Hilton Garden Inn, it just so happens to be the same hotel that John is staying at. John is in the running for a position that Jane is also trying to get. Jane recently heard about Firesheep on the news and is in a mischievous mood. She logs on to the hotel wifi and runs Firesheep. Luckily for Jane, John is using the wifi and she sees that he has an unsecured connection to their company web

email portal. With one click she is now logged in to John's email account. Just take a second and think of the trouble she could cause him, the private things she has access to, the general control/chaos email can exert in someone's life.

This type of exploit, session hijacking via unencrypted network traffic (aka sidejacking), has always been possible by those that knew what they were doing. Now with the release of Firesheep this is possible by anyone that knows how to download an extension and click a button.

While you go download Firesheep, (yea thats right, I know what you're doing you jerk) you might be thinking that this is a horrible thing to happen. Quite the opposite actually, this has spurred web companies to finally get off of their respective laurels and take HTTPS seriously. Gmail, Facebook, and Twitter now all default to using HTTPS throughout their entire site. Previously the standard had been to only encrypt login pages, which secured the user's login credentials but left their current session open to hijacking as in our example above.

What is HTTPS

Normal interweb traffic is transferred over HTTP, when you type "http://www.google.com" into your browser you're using HTTP, notice the "http://" at the beginning there. Normal HTTP traffic uses port 80, HTTPS on the other hand uses port 443. HTTP is not secure in the least, every thing you do is sent free and clear for anyone listening to see what you're doing. HTTPS is "HTTP Secure" or "HTTP on SSL", acronym semantics can be argued but they both mean the same thing. HTTP using SSL to secure it.

I'm only going to cover how HTTPS works at a very high level since the details won't matter to most people. If you're interested in learning more please do, google.com is a good place to start ;)

A real life example to explain how SSL works is a diplomatic bag[13]. The contents are secured and can only be opened on either end of the transfer by the person with the proper credentials. The bag is secured by international law, as well as physical means, just as the SSL encrypted message body is protected by a strong algorithm and keys.

A certificate authority will sign your website's certificate to prove that it is valid. The user's web browser already knows the major certificate authorities and will verify the sites certificate against the root certificate that the certificate authority provides. The traffic will then be encrypted with this key on both ends, so the only traffic going across the network is encrypted traffic. If you've ever used SSH with public keys for authentication you are already familiar with

[13] http://en.wikipedia.org/wiki/Diplomatic_bag

the process. You have a public and private key that is used to verify your identity with a remote server.

This will protect you from man in the middle attacks, including the session hijacking we mentioned above if all of your site is encrypted with HTTPS.

Limitations

There are a few limitations when using HTTPS that may make it infeasible in certain circumstances.

Virtual Hosts

Under normal configurations virtual hosts can not be used with SSL. This is a problem if you're using shared hosting or simply running multiple sites on the same server. The reason for this is because the server can't determine the host header until the connection has been completed, which requires the SSL authentication. Since certificates can only have one host this means it will simply not work. The easiest way around this is to setup multiple IP addresses and use IP based hosts instead of the name based host resolution you're probably used to. I usually recommend setting up a separate server for secure sites though, if you need HTTPS you are probably at the point of needing a dedicated server as well.

There are however some hosting providers with shared certificates that can be used across the sites hosted with them. This can enable you to quickly and cheaply support HTTPS. The main issue with this is that the domain would need to reflect the hosting provider's domain name. For example instead of

```
https://yourApp.com/login
```

the URL would be something like

```
https://yourHost.com/yourApp/login
```

This may or may not be a concern depending on your application and branding needs.

Speed

HTTPS connections require SSL handshakes to establish the connection, thus making the overall transfer slower. Once that initial handshake is performed additional connections only require the encryption and decryption of the content, meaning that once the initial connection is complete, subsequent connections aren't much slower. The performance impact is ***incredibly*** low though, this is not a valid reason to discredit the use of HTTPS.

Caching

Cheddar. Fat stacks. Dead Presidents. Cash money. Nah, actually we're talking about cache. The secret sauce behind your super quick load times. You have to say it with a british accent. Modern browsers will cache HTTPS content the same as HTTP content so there is no disconnect there. To cause older browser to support caching set the Cache-Control header, for example

```
header('Cache-Control: max-age=31536000');
```

would tell the browser to cache for one year.

The real issue comes with proxy caching. Proxy caching might come from an ISP or a service meant to speed up

connections. This is mostly used in rural parts of the world that have slow internet connection speeds. Using HTTPS, this type of caching is impossible since all the traffic the proxy sees is encrypted. This is not a major issue for most sites but if you have a large global userbase, or an application that targets users in remote locations, this should be considered carefully.

Another thing to think about, there is a good chance that there are parts of your site that should NOT be cached. This means that you shouldn't just let the browser cache everything, sit down and plan out which parts of your application should be cached and for how long. For example, CSS and JavaScript should probably be cached for a significant amount of time; whereas the user's timeline view should update very often.

Certificate Types

There are two types of SSL certificates.

Domain Validated Certificates do not verify as much information as their counter parts but they are substantially cheaper. Usually starting around fifty dollars, they will likely be the best option for small sites. The main down side from a user perspective is that there is usually some distinction in the browser between the two, for example a Domain Validated Certificate might only show a lock symbol in the address bar while an Extended Validation Certificate will show the full green address bar.

Extended Validation Certificates are the gold standard of SSL certificates. They not only validate that you are the owner of the domain but also verify the identify and legitimacy of the domain owner. Since this usually requires a personal effort on the part of the Certificate Authority these certificates are

significantly more expensive. Usually Extended Validation Certificates start around five hundred dollars. This will be the certificate of choice for most large and reputable companies. Browsers will display the full green address bar when an Extended Validation Certificate is in use, giving users more peace of mind.

When to use HTTPS

The traditional view has been to use HTTPS anywhere credentials or other sensitive data is passed to the server. For many years this has meant that login pages and shopping carts were all that was encrypted. These are still valid and necessary places to use encryption but will leave the rest of the user's session open to man in the middle attacks. Recently there has been a movement to use HTTPS everywhere. Which is just a marketed way of stating that every page of your site would be encrypted on HTTPS. This is a good rule in many cases, the limitations of HTTPS should be considered though, don't just blindly implement HTTPS everywhere without evaluating the trade-offs. If you determine that the limitations discussed above are offset by the enhanced security throughout for your specific application then using HTTPS on each page is strongly recommended.

Are you thinking that at this point it'd be easier to just forget about this whole HTTPS thing? Okay. Okay. Let's just slow down. Slow down. Regardless of you're constraints you have an obligation to your users to implement the best security you possibly can. If you run a shopping cart or collect credit cards for instance, HTTP is not even an option. More and more even for what isn't considered sensitive data, like a social media account, it is becoming standard to encrypt. Don't be left behind, use HTTPS whenever you can.

Implementing HTTPS

What kind of SSL Certificate do I need?

The main question to ask yourself is do you need to secure subdomains or not. If you need to secure multiple subdomains, eg

```
api.yourApp.com
docs.yourApp.com
yourMom.yourApp.com
cart.yourApp.com
```

then you'll need a Wildcard SSL Certificate. If you don't need that capability and only need to secure something like

```
yourApp.com
```

then a standard certificate will work just fine. The only deterrent to getting the Wildcard just in case you need it later is the cost.

Generating your Server Certificate

In order for the Certificate Authority to sign and generate your certificate you'll need to generate keys on your server and then upload those to the Certificate Authority.

This will require OpenSSL, if you don't have it on your server you'll need to install it. Installing applications across various server operating systems and distributions are out of the scope

of this book, hopefully if your at the point of needing to setup HTTPS you know your way around your server well. If you don't know your server operating system or distribution well it might be a good idea to hire someone to help you setup SSL certificates.

First create a directory to store your keys, people have differing opinions on the best place to store these but for our examples we'll stick with

```
/usr/bin/ssl/
```

Let's generate our private RSA key

```
$ openssl genrsa -out yourApp.key 1024
```

Then generate the CSR using the RSA key

```
$ openssl req -new -key yourApp.key -out yourApp.c\
sr
```

You'll now be asked several questions with smart defaults, the main one to pay attention to is "Common Name" which should match your domain name, eg "yourApp.com".

Now you have two new files

```
/usr/bin/ssl/yourApp.key
/usr/bin/ssl/yourApp.csr
```

Before you do anything else, make a backup copy of the .key file somewhere. Seriously, make two backup copies. If you lose the private key you'll need to buy a new certificate, and servers crash all the time.

Obtaining a SSL Certificate

The first step to getting up and running on HTTPS is to obtain a certificate. There are cheap/free certificates available from some certificate authorities but they won't come pre-installed on the popular web browsers so that makes them useless for external facing sites. If you're running an internal application then cheap alternatives and self signed certificates are valid options, for everyone else we'll need to purchase a certificate.

First off I recommending checking with your DNS provider to see if they offer any type of discounted or easy to setup certificates, for example DNSimple is the DNS provider I use and they offer subscription payments for certificates at a large discount.

If your DNS providers does not provide certificates Symantec/VeriSign is a well respected certificate authority.

Now go buy one.

You'll then need to walk through whatever process your chosen Certificate Authority has in place for setting up your certificate, usually you'll just upload your server certificate (yourApp.csr) and they will email you the signed certificate.

Your certificate authority will provide you with the sign certificate which we'll name yourAppSigned.crt. Copy this to your server, for this example I'll use the following path

```
/usr/bin/ssl/yourAppSigned.crt
```

Apache Setup

If you're using Apache follow these steps, if you're using a different web server skip this section and keep reading. Open your httpd.conf file in your favorite text editor. Note, some distros may use separate config files for https. For example, my laptop running OSX uses a httpd-ssl.conf file.

Add a VirtualHost similar to the following, it will likely closely match your existing VirtualHost for your HTTP site

```
<VirtualHost *:443>
  DocumentRoot "/path/to/your/app/htdocs"
  ServerName yourApp.com
  SSLEngine on
  SSLCertificateFile /usr/bin/ssl/yourAppSigned.crt
  SSLCertificateKeyFile /usr/bin/ssl/yourApp.key
</VirtualHost>
```

Restart Apache

```
$ apachectrl restart
```

or

```
$ service apache restart
```

will usually do the trick.

Try your site out with "https://yourApp.com", you should be good to go!

Nginx Setup

If you're using NGINX follow these steps, if you're using a different web server you'll need to research how to set this up with your server, sorry!

Open your Nginx virtual hosts file in your favorite text editor. Add a virtual host similar to the following, it should closely match your existing site setup

```
server {

  listen   443;

  server_name yourApp.com;
  location / {
    root   /path/to/your/app/htdocs;
    index  index.php;
  }

  ssl                                    on;
  ssl_certificate           /usr/bin/ssl/yourAppSigned.crt
  ssl_certificate_key   /usr/bin/ssl/yourApp.key

}
```

Restart Nginx

```
$ sudo /etc/init.d/nginx restart
```

or

```
$ service nginx restart
```

will usually handle it.

Try your site out with "https://yourApp.com", it should be ready!

Additional Resources

For Apache the best source is the docs

```
http://httpd.apache.org/docs/current/ssl/ssl_howto\
.html
```

For NGINX the WIKI is a great starting place

```
http://wiki.nginx.org/HttpSslModule
```

For anything else just replace "yourWebServerName" in the text below with the name of the software your using to serve web pages, then paste the full URL into your web browser

```
http://lmgtfy.com/?q=yourWebServerName+SSL+certifi\
cate+setup
```

Paths

Base Path

You should ensure that users are on the HTTPS version of your site whenever it is needed. This can be done in Apache/Nginx configs using redirects. Another simpler option is to set the base path of your application to use your HTTPS URL, eg "https://yourApp.com" and force a redirect using the base path if a user comes in on HTTP.

A lot of times you will want to allow HTTP on certain pages and require HTTPS on others, this is where your web server configs and proper routing in your code come in.

Relative Paths

One more thing to mention that isn't necessarily security related but will make your life a lot easier when using both HTTP and HTTPS on one site. URLs for assets, eg CSS or JS, can begin with double forward slashes instead of http:// or https:// to reference the current protocol. For example, on your home page you might have

```
<link type="text/css" rel="stylesheet" href="//ass\
ets/main.css" />
```

navigating to https://yourApp.com would cause this to load

```
https://yourApp.com/assets/main.css
```

whereas navigating to http://yourApp.com would load

```
http://yourApp.com/assets/main.css
```

That's just a little trick to make your life a little easier, because I care.

Done

You are done, good job. Pat yourself on the back, mix your favorite drink, and take a well deserved nap.

Chapter 3 - Password Encryption and Storage for Everyone

You should know how this works by now. Chris is a junior developer working for Marvel Comics[14] web team. It's an abnormally hot summer in Burbank. He has just been tasked with building the login functionality for the new web/tablet comic portal his team is building. His "team" really means Chris and the other developer. Chris might have forgotten to wear deodorant today, why is it so hot.

Chris plans out how the login system will work. It'll have the normal things you would expect, login/logout/forgot password/etc… In regards to passwords he'll need to store the user's password, compare it on login, and then email it back to the user if they forget it. Minutes pass. As he thinks through each part of the login process he starts to worry about the security implications of having users' passwords available to read by anyone who has, or gains, access to the database. He knows he should encrypt the passwords but what about decrypting for login? Or when a user forgets their password?

After researching for an excruciatingly boring 45 minutes Chris decides that he needs to use PHP's built in `mcrypt_encrypt()` and `mcrypt_decrypt()` methods. Chris is pretty stoked, secure encrypted passwords and all the dirty work on

[14]This is fiction built from truth. Please don't sue me Marvel.

encrypting and decrypting is done by PHP. He's going to be done with this project in half the time quoted.

That my friends (we're friends right?), is why when you reset your password on Marvel.com you get your plain-text password sent to you in an email. AN EMAIL. WITH YOUR PASSWORD IN IT. This is why you should go change your Marvel.com password to something you have never used anywhere else right now. Go ahead, I'll wait. I'll just be sobbing in the corner thinking about all of the people that will be exploited by this soon enough.

The moral of the story, don't store passwords. Store one way hashes of users' passwords. Now let's walk through how to do this right.

The Small Print

I am not a cryptographic expert. This is my personal advice based on experience. These are opinionated best practices and are not meant to be used as directions for securely storing nuclear launch codes. Your most important tweets will be kept safe though.

What is a Hash?

First we need to cover the basics. Hashing is not encryption, passwords should be one way hashed. This means that it is impossible to decrypt, hence the one way part of that. There is never a need to display a password back to a user/admin/-mother/anyone ever. Once a password is entered it becomes something totally different, a hash that can be recreated only by the original password being given as input.

Popular Attacks

Before discussing any further let's delve into the popular attacks against hashing algorithms.

Lookup Tables

A lookup table is simply a table of hashes where the password is known. It can be as simple as this

```
password | hash
------------------------
pass1    | bidfb2enkjnf
pass2    | psdfnojn3nod
etc...
```

This is then compared against the passwords hashes in your database to determine the password that was uses. This attack is useless if you are using random salts but it's easier than robbing a train if the hashes aren't salted.

Rainbow Tables

A rainbow table is technically sophisticated compared to a lookup table yet very similar. It is basically a less memory intensive way of achieving a lookup table through mathematic means. You can think of them interchangeably as we mention them here. Rainbow tables are also thwarted through the use of random salts so are less relevant with modern hashing algorithms.

Rainbow Tables are a very complex exploit that is really out of the scope of this book to explain. If you want to learn more you can read the original paper published by Martin Hellman[15] that introduces the concept. The Wikipedia article on Rainbow Tables[16] is also a good source for an easier to digest explanation.

[15] http://www-ee.stanford.edu/~hellman/publications/36.pdf

[16] http://en.wikipedia.org/wiki/Rainbow_table

Collision Attacks

A collision attack is the main security flaw found in most hashing algorithms. There are two types of collision attacks.

A classical collision attack is when two different values generate the same hash. A simple example is

```
string1 = 'abc123'
string2 = 'bcd234'

hash(string1) === hash(string2)
```

A chosen-prefix collision attack is when different prefixes are used to cause two separate values to generate the same hash. Here is an example of this

```
prefix1 = 'zxy'
prefix2 = 'abc'

string1 = 'abc123'
string2 = 'bcd234'

hash(prefix1 + string1) === hash(prefix2 + string2)
```

A Pinch of Salt

That's not enough though. The next problems to look out for are lookup tables and rainbow tables exploits. Both of these exploits basically keep a list of popular passwords and their resulting hashes. Rainbow Tables are a little more complex but that's basically what's going on. How do we combat this? Salts. RANDOM SALTS.

A salt is something that is appended to the password hash to make it unique. Salt is also something that is added to the rim of a margarita to make it delicious. Ah margaritas. Anyway. So you take a random string (salt) and combine it with the plain-text password string to give you a unique value. This means that even with a lookup table of known password hashes an attacker can't match up your user's password hash with the DB password hashes since a random salt has been used. Given two identical passwords the resulting hashes will be unique. A random salt is one of the most important pieces of your password security.

Random isn't always Random

You're salt needs to be random to be effective. Random !== random though. You don't necessarily need truly random numbers but `rand()` won't cut it either.

The problem with using PHP's built in `rand()` or `mt_rand()` functions is that they are seeded with data that can be manipulated and determined, they don't provide enough "randomness". The main ingredient to producing a truly random number is to include enough entropy into the source. Entropy is the amount of truly random information collected by the

system that is generating your random number. Most non-cryptographic random number generators, like `rand()` and `mt_rand()`, use algorithms to produce their numbers without enough outside sources of data to make them truly unique. This means that the data `rand()` produces can be manipulated and guessed by an attacker. After observing enough of the output of `rand()` an attacker an reliably determine the future output. In fact after returning as little as 624 values from `rand()`[17] it is possible to calculate all future values.

You have been warned, you SHOULD NOT use `rand()` for your salts. By buying this book you are hereby contractually obligated to not whine when I personally come beat you in the face with a 1st generation iPad if you use `rand()` for your salts. That thing was solid as a brick.

Calling your server's `/dev/random` is your best bet for true random on most systems. The issue with using this for login is that it blocks while collecting entropy from the system. Collecting entropy means that it will collect environmental data from your system; like various HW data, keyboard typing, mouse movements, disk access, etc. in order to generate a buffer of random bits. This means that it can take a long time to return, especially on servers that aren't busy. I ran in to this recently actually, a developer on my team used `/dev/random` to generate an activation code. Everything seemed fine, it passed testing fine. Then after we deployed the project to it's own production server, requests ended up taking over 60 seconds. The server wasn't busy since it was only hosting this one project, causing `/dev/random` to block for a good bit of time while it collected entropy. What was the solution? `/dev/urandom`

[17] http://eprint.iacr.org/2005/165.pdf

`/dev/urandom` isn't considered true random but it is cryptographically secure. This means that it might not be a truly random number but it is regarded as secure enough for use in salts. `/dev/urandom` will return a very good pseudo-random number immediately with no blocking. `/dev/urandom` uses the existing entropy pool to generate a pseudo-random number that is secure enough for the majority of authentication systems. If you're writing the login page for nuclear launch codes it might be best to make the user wait on `/dev/random` but for that social picture sharing site `/dev/urandom` is good enough.

Hashing Algorithms

Let's discuss the popular hashing algorithms.

MD5

I see the MD5 hashing algorithm used incorrectly more often than anything else. Usually because it is supported by most databases by default. MD5 has been mathematically proven for some time now to be insecure. The issue with MD5 is that it is trivially easy to produce collisions on modern hardware.

One of the most notable examples, in 2005 researchers were able to generate collisions in MD5 checksums using a laptop[18]. The significance of that is that it doesn't take a $200k beast of a server to break MD5, just any old laptop and that was in 2005. In 2005 people, that was like 100 internet years ago. No more MD5 for password hashing please. Non-secure hashes to verify data contents, sure. Just not for secure hashes that an attacker would be interested in breaking.

MD5 is not completely broken since it is still mostly secure when used with a proper salt. That doesn't mean that you shouldn't move on to a more future prove solution though.

SHA-1

Ah good old SHA-1, trusty and secure for years. Those are IRL years too, in internet years that's decades. In 2005 (2005 was a bad year for security) researchers from Shandong University released a research paper[19] proving that SHA-1 collisions

[18] http://cryptography.hyperlink.cz/md5/MD5_collisions.pdf

[19] http://link.springer.com/chapter/10.1007/11535218_2#page-1

could be reliably generated with less than 2^{69} hash operations. Collisions at around 2^{80} hashing operations are considered safe cryptographically. FYI 2^{80} is about $1.20892x10^{24}$., so "cryptographically secure" means pretty darn secure. Since then Moore's law[20] has ensured that SHA-1 is even less secure and should be avoided in any application that needs true security.

As I explained above for MD5, when used with a random salt SHA1 is still algorithmically secure.

SHA-256 / SHA-512

The SHA-2 standard was introduced as a successor for SHA1 in 2001. It's accepted was accelerated a bit when SHA-1 was proven to be insecure in 2005. SHA256 and SHA512 are basically the same; SHA-256 uses 32-bit words, SHA-512 uses 64-bit words. They also have a different number of rounds. The core algorithm is practically identical though.

SHA-2 is currently considered cryptographically secure with no known vulnerabilities when used with a sufficient number of rounds (>64).

It has not seen the same amount of scrutiny as Blowfish though, the cypher that BCrypt uses internally. The Blowfish cypher which has been around since 1991 and is still considered secure, yet using it with a weak key is a known weakness. Being based on a cypher gives BCrypt a additional layer of cost that makes it superior to a standard hashing algorithm, in other words BCrypt is slower by design. Slow is a good thing!

Even though I recommend BCrypt, SHA-256 or SHA-512 are currently valid and secure options for secure hashing when

[20] http://en.wikipedia.org/wiki/Moore's_law

used as part of a derivation algorithm, like PBKDF2 or the algorithm implemented with PHP's `crypt()` function.

BCrypt

BCrypt is seen by a lot of people as a newcomer and isn't as widely known. BCrypt was released in 1999 so it's not exactly new here. BCrypt is a key derivation function based on the Blowfish cypher, it is iterative so it protects against brute forcing due to the cost associated with generating a hash.

There are currently no published exploits of BCrypt despite the fact that it has seen considerable attention from cryptographic researchers. It has also been around for a good length of time so as of this writing (2014), BCrypt is considered cryptographically secure.

BCrypt does have a limitation of 72 characters for the plaintext password being encrypted. This is usually taken into consideration by either stripping the excess characters or by simply validating 72 as the max length.

BCrypt will be our choice for passwords in the following examples.

SCrypt

SCrypt is the new kid on the block. Released in 2012, it is a memory intensive key derivation function. Theoretically SCrypt is more secure than BCrypt due to the high cost inherent in the algorithm but since it is so new I don't personally recommend it's use at this time.

New is a bad thing in the cryptographic world, it means that SCrypt hasn't received the same level of attacks and scrutiny

as older algorithms. There have been a few exploits reported recently that don't mean SCrypt isn't secure, but do take away from the security advantage it theoretically has over BCrypt.

One big thing that SCrypt has going for it is that a few popular crypto-currencies are using it for their mining operations, most notably Litecoin and Dogecoin. This means that it will likely receive a large amount of attention sooner than its predecessors.

Storage

This section will be short, WAKE UP. In whatever system you store your passwords hashes, whether it's in a relational database, key store, lock box, sock drawer, or file system; use either an unlimited length text field or I recommend a using a varchar(255). You're hashing algorithm will produce a maximum length string so you don't have to worry about an attack overloading your database. Different hashing algorithms will produce different fixed length strings so you could set your field length based on your algorithms. I instead prefer to use a larger than needed field length constraint to handle future hashing possibilities rather than try to save a few bytes.

Using BCrypt your hash will always have a maximum character length of 60 characters. So in theory you could get away with a varchar(60) field in your database but this doesn't account for future changes. It's better to future proof your passwords now than to try to save a few bytes in your database. So just leave it as a text field or a varchar(255), you won't regret it.

Validation

The only validation that's needed on a password field is minimum length. You should allow any character, whitespace, phrases, etc so your users can construct as complex of a password as they want. Pass-phrases should be endorsed, "correct horse battery staple" is a much better password than "myNewPassword"[21]. Your only worry is that the password is not complex enough, hence the minimum length. For the love of all that is good, like cat gifs, don't do stupid things like using JavaScript to restrict copy-paste. If the user wants to use a password management tool do all you can to make they easy for them. If you do stupid things you'll make your users, and the cats very sad. OK, that rant is over now.

The only caveat to this is that with BCrypt only the first 72 characters of the password will be used so technically you could limit to a maximum of 72 characters and not loose any data. That does put a limit on your users though and is not future proof for your next hashing algorithm. If your user has a 74 character pass-phrase memorized it's best to let them use that and only use the first 72 characters than to make them think up a new pass-phrase. Some sources recommend hashing the passwords using a standard hashing algorithm (SHA-256, SHA-512, etc) and then BCrypting the resulting hash to account for this length issue. That is a perfectly valid option. I'm not going to recommend it here simply because with a valid salt plus 72 characters of the password you will have enough data to keep your hashes secure according to current research.

[21] http://xkcd.com/936/

Putting It All Together

Now that we've covered the basics let's write a little code, finally.

First I'm going to walk you through the traditional/deprecated way of doing this and then I'll walk you through the newer way that is available with PHP 5.5 and up. The reason for this is

1. I want you to understand what is going on behind the scene instead of just seeing wrapper functions
2. There is a decent chance that you're on an older version of PHP so I don't want you to be left behind

The scenario here is that we are registering a new user. We will need to generate a random salt, generate their password hash, then store both of these to an imaginary database for authentication in the future.

< PHP 5.5

If you are using a PHP version less than 5.3.7 you need to upgrade to at least 5.3.7 to have a decent level of security. There is really no other sound recommendation I feel comfortable giving. If you are on an older version there are many bug fixes that are patched with upgrades. Specifically to our case, there was a BCrypt vulnerability patched in 5.3.7. We will be using the $2y$ prefix is our examples, this is the "always to specification" prefix. Meaning it has been updated with the vulnerability fix and is the most up to date logic.

First off we're going to generate a unique random salt. There are a few ways to do this in PHP depending on what extensions are compiled on your system. You can most likely just do this

```
1  //generate the binary random salt
2  $saltLength = 22;
3  $binarySalt = mcrypt_create_iv(
4    $saltLength,
5    MCRYPT_DEV_URANDOM
6  );
7
8  //convert the binary salt into a safe string
9  $salt = substr(
10   strtr(
11     base64_encode($binarySalt),
12     '+',
13     '.'
14   ),
15   0,
16   $saltLength
17 );
```

We are calling `mcrypt_create_iv()` with the MCRYPT_DEV_URANDOM flag to tell it to buffer from /dev/urandom. Wrapping it in `bin2hex()` makes use get a hexadecimal string back instead of the straight binary data.

If MCrypt isn't available on your system you can always fall back to reading directly from /dev/urandom. Unless your on Windows, in which case just install Linux before it's too late. Go ahead, I'll wait. You can thank me later; I'll take check, cash, or card for the thank you gift.

```
//generate the binary random salt
$saltLength = 22;
$binarySalt = file_get_contents(
  '/dev/urandom',
  false,
  null,
  0,
  $saltLength
);

//convert the binary salt into a safe string
$salt = substr(
  strtr(
    base64_encode($binarySalt),
    '+', '.'
  ),
  0,
  $saltLength
);
```

Now that we have the salt let's hash the password as well

```
//generate the binary random salt
$saltLength = 22;
$binarySalt = mcrypt_create_iv(
  $saltLength,
  MCRYPT_DEV_URANDOM
);

//convert the binary salt into a safe string
$salt = substr(
  strtr(
```

```
11      base64_encode($binarySalt),
12      '+',
13      '.'
14    ),
15    0,
16    $saltLength
17  );
18
19  //set the cost of the bcrypt hashing
20  //remember to experiment on your server to find th\
21  e right value
22  $cost = 10;
23
24  //now we'll combine the algorithm code ($2y$) with\
25   the cost and our salt
26  $bcryptSalt = '$2y$' . $cost . '$' . $salt;
27
28  //hash it, hash it good
29  $passwordHash = crypt(
30    $_POST['password'],
31    $bcryptSalt
32  );
33
34  //verify a secure hash was returned
35  //this could be an error code or insecure hash
36  if (strlen($passwordHash) === 60) {
37
38    //this next part is just for demonstration
39    $db = new ImaginaryDatabase;
40    $db->user()->create(array(
41      'password' => $passwordHash
42    ));
```

```
43
44  }
45  else {
46
47    //error handling
48
49  }
```

So we hashed our password and saved it to the database. Notice how we didn't save the salt? That's because `crypt()` will store the selected algorithm, hashed password, and salt all within the returned hash.

Well we have the user signed up now, let's say they leave our site (how dare they!) and come back later, they need a way to login. Login is nice and simple, we just need to check their plaintext password against the hash stored in our database.

```
1  //we are going to start with a default state
2  //of failure, always assume failure first
3  $valid = FALSE;
4
5  //grab the hash from our imaginary database
6  $user = $db->user()->where(array(
7    'email' => $_POST['email']
8  ))->row();
9
10 //now check to see if the login password matches
11 $pass = $_POST['password'];
12 if (crypt($pass, $user->pass) === TRUE) {
13   $valid = TRUE;
14 }
15
16 //other checks, error handling, etc...
17
18 if ($valid === TRUE) {
19   //valid auth stuff
20 }
```

If you pass a previously generated hash as the second parameter to the `crypt()` function it will use that algorithm and salt to generate a new hash that you can then compare against your previously stored password hash.

>= PHP 5.5

New password hashing functions were introduced in PHP 5.5[22] to greatly simplify the process of handling passwords. The purpose of this is to hide a lot of the complexities and make PHP users secure using the default functions built into PHP without relying on all developers to know what they are doing. In other words they are trying to steal my book content, jerks.

In all seriousness though, use the PHP password functions when possible. They provide built in, up-to-date hashing with additional safety checks that we aren't even covering here, like protection against Timing Attacks[23].

We'll walk through the same exact steps here using PHP 5.5 syntax. We'll be using the password_hash function to generate our hashed password, this function will automatically create a random salt so we can skip that step completely. Let's just jump straight into it

```
1  //FYI - the default cost is 10, it can be customiz\
2  ed though
3
4  //hash it, hash it good
5  $passwordHash = password_hash($_POST['password']);
6
7
8  //this next part is just for demonstration
9  $db = new ImaginaryDatabase;
10 $db->user()->create(array(
```

[22] http://www.php.net/manual/en/book.password.php

[23] http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.65.9811

```
    'password' => $passwordHash
));
```

and then to verify the user's password at login you will do

```
//we are going to start with a default state of fa\
ilure
//always assume failure first
$valid = FALSE;

//grab the password hash from our imaginary databa\
se
$user = $db->user()->where([
  'email' => $_POST['email']
])->row();

//now check to see if the login password matches
$pass = $_POST['password'];
if (password_verify($pass, $user->pass) === TRUE) {
  $valid = TRUE;
}

//other checks, error handling, etc...

if ($valid === TRUE) {
  //valid auth stuff
}
```

As you can see, the new password functions make your life a lot easier in PHP 5.5. Another advantage to this functionality being baked into PHP is that you can rely on it remaining up to date in the future, instead of having to stay up to date on

the latest password algorithms yourself. Even though I know this has been riveting, leave it to the pros like the PHP core team when you can, that way you have more time to make cool stuff.

When you implement your authentication scheme make sure to reference the PHP docs to ensure your code is up to date. Also, be sure to take advantage of the other password functions available like `password_get_info()` and `password_needs_rehash()`.

Are you on PHP 5.3 still? Missing out on all the fun these users get using that new fangled 5.5? Not to worry, Anthony Ferrara has you covered. His password_compat library implements the new password functions in PHP 5.3.7 and up just for you. You can download it here on Anthony's Github[24] I **HIGHLY** recommend using this when possible instead of writing the code yourself, this has been tested so you can rely on it.

[24] https://github.com/ircmaxell/password_compat/

Brute Force Protection

You can have the best password hashes out there, like totally secure dude, but that doesn't do you any good if someone just hammers away at your login page until they find the correct password. Brute Forcing is the process of someone using software to repeatedly try different passwords until they get in.

Securing yourself from this type of attack is pretty easy, sadly I see so many sites that are vulnerable to this. So how do you secure yourself from it? Just make it take longer than is feasible for someone to find a password this way.

So someone tries to login, fail. They try again, fail. One more, fail. Now you make them wait 60 seconds before trying again. Done. That's really the simplest, yet still super effective method of preventing brute force attacks. If an attacker can only try three passwords per minute then it's going to take them forever to find the user's password so they move on. Your users are safe. You don't have to limit it at three exactly, three or five or ten are pretty popular numbers; personally I wouldn't go over ten per minute.

Next up you can expand on that a bit. Block based on IP. Only allow X number of login attempts from a certain IP across all users instead of just each individual user. Just be sure to include good error messages and reasonable access reinstatement times, corporate users often share the same public IP address so you don't want to punish thousands of valid users due to one bad user if you can help it.

Also note, the same logic applies to your API authentication. Don't leave an opening to attackers to simply take a different path to get the same result. If you secure something on

the frontend of your application and then expose that same functionality through an API make sure to secure it just as well there.

Upgrading Legacy Systems

Now to the elephpant in the room. Get it? ElePHPant. Yea, it's been a long night, sorry. So how to you upgrade your existing system that has MD5 passwords with no salt?

I'm going to give you two options.

- Path 1 - As each user logs in we will silently upgrade their hash to use BCrypt. They won't even know the difference. Soon enough you will have a database of well secured passwords.
- Path 2 - Use BCrypt to hash the existing MD5 hashes that are in the database. New passwords will be hashed with MD5 first and then BCrypt.

Upgrade Path 1

Path 1 is the traditional advice for migrating to new authentication schemes and is by far the best option in the majority of circumstances. To implement this you would do something similar to this

```
1  $valid = FALSE;
2
3  //grab the password hash from our imaginary databa\
4  se
5  $user = $db->user()->where([
6    'email' => $_POST['email']
7  ])->row();
8
9  if ($user->pass_hash_algo === 'bcrypt') {
```

```
10    //they have previously logged in and
11    //upgraded their hash so they're good
12    //proceed with verifying login as usual
13  }
14  else {
15    //this is modified from our old hash check
16    $oldHash = md5($_POST['password']);
17
18    if ($oldHash === $user->pass) {
19      //generate the new hash with the plaintext
20      //password they just gave us
21      $newHash = password_hash($_POST['password']);
22
23      //save the new hash and flag to our database
24      $user->pass = $newHash;
25      $user->pass_hash_upgraded = 'bcrypt';
26      $user->save();
27
28      $valid = TRUE;
29    }
30
31    //other checks, error handling, etc...
32  }
```

With this modified login system in place your users will upgrade to the new hash automatically as they login. This upgrade path can easily be modified throughout the years as the recommended hashing algorithms change. In a few years SCrypt could be the accepted standard and it would be trivial to upgrade users to that.

Upgrade Path 2

Path 2 isn't as clean of a solution but is immediately secure. Where Path 1 retains the insecure hashes until each user logs in next this will secure your data right now so you don't have to lose any sleep tonight. With this path you will immediately re-hash all current password hashes using BCrypt. Your new hashes are just BCrypt hashes of MD5 hashes of the passwords. This complicates the process since you will need to use both BCrypt and MD5 basically forever, this could become cumbersome years from now when you're SCrypting BCrypts of SHA1'd MD5 hashes.

This is a two step process, first write a script to do basically the following and run it against your current database. Please tell me you're doing this with a database dump and not on the prod server. Right? Cause that would make me sad... Also you should add some type of status to each record while you do this. Basically just use this as an example, no copy-pasting mmkay.

```
1  $users = $db->users()->result();
2
3  foreach ($users as $user) {
4
5    $newHash = password_hash($user->pass);
6
7    $user->pass = $newHash;
8    $user->save();
9
10 }
```

Now that your database has been updated change your registration method to do this

```
//hash it with MD5 then BCrypt
$passHash = password_hash(md5($_POST['password']));

//this next part is just for demonstration
$db = new ImaginaryDatabase;
$db->user()->create([
  'password' => $passHash
]);
```

The login method will need to check against an MD5 hashed password instead of the plaintext as well

```
//grab the hash from our imaginary database
$user = $db->user()-where([
  'email' => $_POST['email']
])->row();

//md5 hash the password before we bcrypt it
$md5Pass = md5($_POST['password']);

//now check to see if the login password matches
if (password_verify($md5Pass, $user->pass) === TRU\
E) {
  $valid = TRUE;
}

//other checks, error handling, etc...
```

It's Over. We're Safe.

That's all folks. There's a ton more to cover as always but this is the need to know info. I'll probably end up writing additional chapters or a second volume to cover advanced topics in more detail. Stay tuned for the next chapter, we'll cover restricting URLs, safely handling files, and ensuring that no one sees data they shouldn't.

Resources

If all of this talk about cryptography has interested you I strongly suggest you check out Bruce Schneider's work. You should definitely read his book, Applied Cryptography[25], it is the bible of crypto.

I would also like to bring attention to the work that a fellow member of the PHP community has been doing, Anthony Ferrara[26] championed the password_hash functionality integrated into the PHP core starting with version 5.5. This was a huge step towards insuring we have secure PHP applications by default in the future.

[25] https://www.schneier.com/book-applied.html

[26] http://blog.ircmaxell.com/

Chapter 4 - Authentication, Access Control, and Safe File Handing

Erica works for a small, local manufacturing company as their programmer and general IT person. She was recently tasked with developing software for the office staff, to upgrade them to the digital age instead of using paper records.

The first step was to have staff members scan documents when they were finished with them. Instead of filing the paper forms, they would be stored electronically.

Development of the archiving system proceeded well. The system consisted of a basic database listing of all the documents, with various filters such as department, date, tags, etc. Nothing too complex; Erica wanted to keep it simple.

The documents displayed to users are filtered based on the user's department and their position (e.g., user, manager, director). Everything worked well for several months, then it happened. There was a breach of a secure document. The subsequent investigation determined a disgruntled employee gained access to a file meant only for senior management. The employee leaked the document to a competitor to secure a winning bid for a cushy position for themselves at said company.

So how did it happen? Erica has been working all weekend trying to figure it out, when boom, she finds it. It's so obvious, but she never thought to secure it.

In Erica's system, files are uploaded and named after the ID fields of their respective database records. An uploaded PDF might be named "13424.pdf", while the next file, a DOC for example, would be named "13425.doc", and so on. Erica had secured the document listing that each user sees, but had not considered securing the files themselves when opened or downloaded by the browser. There was no protection to stop a user from gaining access to sensitive documents by simply guessing filenames by incrementing the ID.

The stories are getting a little long, but if you're committed enough you can make any story work. I once told a woman I was Kevin Costner, and it worked because I believed it.

Authentication

The first step to properly handling sensitive data is authentication. You need to ensure that a user is who they say they are.

Here we will expand on the code samples from the last chapter to validate a password, then denote for later use that the user has a valid login.

```
1  class UserModel
2  {
3    //... other methods
4
5    function login($user, $password)
6    {
7      if (password_verify($_POST['password'], $user-\
8  >pass) === TRUE) {
9        $valid = TRUE;
10     }
11
12     //other checks, error handling, etc...
13
14     if ($valid === TRUE) {
15       //successfully logged in
16
17       //just example session class
18       //your own code or framework will likely
19       //have it's own session wrapper
20       Session::put('user', $user->id);
21       return TRUE;
22     }
23
```

```
24      //failed to login
25      Session::put('user', NULL);
26      return FALSE;
27    }
28  }
```

We've saved the user's successful login. A key point here is noting their user ID. Just because they're logged in doesn't mean they have access to do anything they want in your application.

Access Control

In addition to determining if the user has a valid login, we need to determine if they have access to the page/section/feature they are requesting.

Let's start with a basic use case. You have two types of users: "Muggle" and "Admin". Muggles are your basic users with very restricted access. Admins are the site administrators whom have access to do everything. An Admin can edit or delete users, manage posts, etc. It is very important that a normal user - a Muggle - cannot perform these actions.

Different frameworks have differing conventions for when and where to check for proper access. In Laravel, filters are preferred. Symfony has voters. In most systems you may simply use the constructors of your controllers. The following is a hypothetical example; your exact implementation will vary, and is up to you.

First we're going to add a method to our `UserModel` for querying the current user.

```
1  class UserModel
2  {
3    //... other methods
4
5    //... login method
6
7    function current()
8    {
9      $user = FALSE;
10
11     if (isset($_SESSION['user']) === TRUE) {
```

```
12        $user = DB::findById($_SESSION['user']);
13      }
14
15      return $user;
16    }
17
18    //let's imagine there is a method for retrieving
19    //the user's groups.
20  }
```

Now let's use these methods to verify that this user has appropriate access.

```
1   class AdminController
2   {
3     function __construct()
4     {
5       $userModel = new UserModel;
6       $user = $userModel->current();
7
8       //check to make sure the user is logged in and
9       //is a member of the admin group
10      if ($user === FALSE || in_array('admin', $user\
11  ->groups) === FALSE) {
12
13        //please add a legit error message
14        //with headers and all that fancy stuff
15        die('Not Authorized');
16      }
17    }
18
19    //other admin only methods
20  }
```

You will usually use some type of abstracted access control layer on top of your regular controllers/routes/etc. This layer should map your routes to the access level required to view that route. For example, `/admin/*` might only be accessible to users in the "admin" group. `POST` and `PUT` requests might only be accessible to "editors". `DELETE` requests only accessible by "admins".

The point is, each page should check the user's access levels and determine if they are authorized for the requested action(s). You don't want to assume that, if a user can see a form and `POST` it to the correct endpoint, the user is authorized to perform that action. This will help ensure protection against malicious users, as well as unexpected changes to your data by users who should not have been able access that data in the first place.

You never want to breeze through authentication security. Take it seriously! You will save yourself or other developers from unnecessary development time and headaches in the future.

Validating Redirects

In the flow of your application, often times you will `POST` a form to an endpoint, validate that data, perform some action, then redirect the user to the next step in the application flow.

If the page you are redirecting to contains sensitive data, someone can bypass your expected flow by simply sending a request straight to this final page, bypassing your validation step.

There are a few ways to handle this. The first way is to simply not redirect at all. Most of the time when you think you need a redirect, you could simply call the next method directly. So instead of:

```
1  if ($valid === TRUE && $dataSaved === TRUE) {
2    URL::redirect('/blog/1/edit');
3  }
```

You could call the edit method without a redirect:

```
1  if ($valid === TRUE && $dataSaved === TRUE) {
2    $this->_edit(1);
3  }
```

In this example, the `_edit()` method would be a protected or private method that can't be accessed from a URL without going through the previous step.

If you do in fact need a separate endpoint, which you may if there are multiple entry points to this destination, then you

need to verify that the proper steps have been executed. Passing any appropriate data along each request in an expiring session variable (commonly called flash data) is usually the safest way.

```
1  if ($valid === TRUE && $dataSaved === TRUE) {
2    Session::flash('blogEditValid', TRUE);
3    URL::redirect('/blog/1/edit');
4  }
```

In your edit endpoint, you would verify the passed-along data:

```
public function edit($id)
{
  if (Session::get('blogEditValid') !== TRUE) {
    //again, this should be a proper error
    //in a real world app
    die('Not Authorized');
  }
}
```

Obfuscation

Have you heard the term "security through obscurity"? It's rarely true, but in some cases it is helpful. Most applications will use an ID field in each table as a primary key. This ID is then used throughout the system to access data. It is passed through URLs, forms, and APIs to denote what piece of data is needed.

Sometimes, though, you don't want to expose the user to the actual row ID. Maybe you are launching a new product and don't want the user to know that they are only the 13th user. Maybe you have public data but don't want your site to be easily crawled by scraping bots.

In these cases, you can obfuscate the ID to something that isn't incremental (such as 1, 2, 3, ...) but can be translated to your ID field. Rather than doing this:

```
Route::get('/edit/{id}', function($id)
{
  //id = 4321234
  $post = $db->post()->where([
    'id' => $id
  ])->row();
});
```

You could do:

```
Route::get('/edit/{hash}', function($hash)
{
  //hash = BaPjae
  $id = HashId::decrypt($hash);

  $post = $db->post()->where([
    'id' => $id
  ])->row();
});
```

In this example, the `HashId::decrypt()` call is simply taking some preexisting security hash from your server and applying it against the passed hash to determine the ID. This can also be called in reverse to generate a hash:

```
$hash = HashId::encrypt(4321234);
echo $hash; //outputs BaPjae
```

It is rather trivial to write these hashing methods yourself, but I recommend using the tried and true HashIDs[27] libraries. These libraries are not only easy to use, but are well maintained and available across many different programming languages to ensure interoperability throughout your entire infrastructure.

There are cases where obfuscation can be argued as needed or it could even be required in your particular use case. For example, I've seen people use obscurity as an additional safeguard for HIPAA data. One use case was a specification that required keeping HIPAA sensitive data in a separate database, in this case it was a three tier server architecture.

[27] http://www.hashids.org/php/

Primary keys used by the web application were stored in one database (along with non-sensitive data), sensitive data was stored in another, and an intermediary database stored the relations. This was designed so that if any one of these databases were compromised the data obtained would be anonymous or non-sensitive.

I'm going to reiterate this point because it is extremely important: in most cases, obscurity doesn't protect you from any legitimate attacks. You shouldn't rely on it for security. It is simply a means of making things a little harder to find.

Safe File Handing

Circling back to our original story: if you have documents that are served to your users for viewing or downloading, you can't simply set access control on the `*.pdf` files. *Why not?* See, I knew you were going to ask that. It helps that I'm the narrator here.

What you need to do is store the file on your server where it can't be accessed from your web server. Here's one example of a recommended directory structure:

```
application/
composer.json
composer.lock
htdocs/
    .htaccess
    assets/
    templates/
    index.php
    robots.txt
uploads/
workers/
```

`htdocs/` is the location that Apache (or whatever web server you're using) would serve for your domain. `uploads/` is where you would store uploaded documents. You can also do this using your server's configuration. In Apache you could return a `404` to any request to the uploads directory if it were in the `htdocs` tree.

To implement this in your application you would have an endpoint for accessing the documents depending on their

type. So maybe you need to serve up monthly financial statements, but only to users in the "accounting" group.

Here we are defining the endpoint to access this. With this endpoint we will verify the user's access level, read the file, and finally output it to the browser with the appropriate headers.

```
Route::get(
  '/accounting/statements/{year}/{month}',
  function($year, $month) {

        //check user access
        $userModel = new UserModel;
        $user = $userModel->current();

        //check to make sure the user is logged in
        //and a member of accounting
        if ($user === FALSE || in_array('accounting', $us\
er->groups) === FALSE) {
          //please add a legit error message
          die('Not Authorized');
        }

        //the user has access, let's read the file
        $directory = __DIR__ . '../uploads/acct/stmnts/';
        $filename  = (int) substr($year, 0, 4)  .
                     (int) substr($month, 0, 2) .
                     '.pdf';

        //error handling for invalid files

        //open the file
```

```
26          $fileHandle = fopen($directory . $filename, 'r');
27
28          //read the file and close it
29          $fileContents = fread($fileHandle);
30          fclose($fileHandle);
31
32          //send the appropriate headers
33          header('Content-type: application/pdf');
34          header('Content-Disposition: inline; filename="' \
35  . $filename . '"');
36          header('Content-Length: ' . filesize($directory .\
37   $filename));
38          header('Expires: 0');
39          header('Cache-Control: must-revalidate');
40          header('Pragma: public');
41
42          //echo the file contents to the browser
43          echo $fileContents;
44
45  });
```

This prevents someone that shouldn't see this file from accessing it, since it is outside of the web server tree and it is being checked against the proper access controls.

Recap

We've covered basic authentication procedures, proper access control through never trusting without validation, and safe file handling procedures. I hope you find this helpful while developing your next awesome idea.

Keep reading, 'cause we're not done yet. Remember, we're done when I say we're done!

Chapter 5 - Safe Defaults, Cross Site Scripting, and Other Popular Hacks

No story this time. This chapter is a catch-all for other attacks you need to protect against, so there isn't an overarching narrative. Try to contain your disappointment.

Never Trust Yourself - Use Safe Defaults

One of the core concepts of a secure system is safe defaults. Whenever possible (and it's usually possible), you should define variables, properties, etc., early with a safe default.

A safe default usually means a NULL, empty, or FALSE state. When determining logic flow, the default should always be a failure. For example, in the earlier authentication examples I check if the password is correct. If it is, we proceed to the positive application logic. If it fails, the function executes the default logic for a non-positive result.

Let's look at a basic example with form validation.

```
Route::post('/signup', function()
{
  //check for a valid form
  if (Form::validate('signup') === TRUE) {
    //process the form
  }

  //the default logic is failure
  die('Invalid Form Data');
});
```

Never Trust Dynamic Typing. It's Not Your Friend.

Dynamic typing is a feature loved by newer programmers, since it seems to make development "easier". Dynamic typing means you don't have to be so picky about types; you just get close enough and it'll work. The problem with this is it doesn't always work the way you'd expect.

Let's look at a classic example with a native PHP function, `stripos()`. The scenario is that you're trying to figure out if the letter "i" exists in the phrase "I am the one who knocks".

```
$phrase = 'I am the one who knocks';

$letterExists = stripos($phrase, 'i');

if ($letterExists != FALSE) {
  echo 'we should be here';
  return TRUE;
}

echo "we shouldn't be here, yet we are";
return FALSE;
```

The `stripos()` function returns the index of the match in the haystack string. Since "i" is the first letter of the string, `strpos()` is returning `0`. Due to dynamic typing, `!= FALSE` evaluates the same as `!= 0`. Now let's fix this with an explicit check using `!==` instead.

```
$phrase = 'I am the one who knocks';

$letterExists = stripos($phrase, 'i');

if ($letterExists !== FALSE) {
  echo 'we should be here';
  return TRUE;
}

echo "we shouldn't be here, annnnnnd we aren't";
return FALSE;
```

Another lesser-known advantage to explicit checks is performance. Most of the time an explicit check will be faster,

since it doesn't need typecasting. If you're ever curious about performance differences in PHP, I recommend checking out PHP Bench[28].

[28] http://www.phpbench.com/

Cross Site Scripting

Cross Site Scripting (XSS) is the process of injecting malicious code into the target website. This can be done in several ways, but the end result is the user's browser runs unauthorized code as themselves, within their current session.

Non-Persistent XSS

This is the traditional type of XSS exploit. It involves injecting data into a site and then guiding users to the malicious content.

Say a page on your site takes `?page_num=2&per_page=50` as query string parameters. If you do not escape these parameters, an attacker could change their values to malicious code. This code could take the user to a delete page, run JavaScript in their browser, or any number of client side attacks.

After injecting their malicious code, the attacker somehow gets a user to visit the page. When the user arrives, the application will verify their valid user session and execute the malicious code. A user could even end up deleting their own account!

Persistent XSS

A persistent XSS exploit is an exploit stored permanently on the server. For example, a social sharing site like Facebook allows users to save messages and display them to other users. An attacker could store malicious code in a Facebook post. If Facebook was not properly escaping this data when displaying it back to other users, that code would be executed.

So anyone that sees the attacker's status would be running this malicious code.

Attack Entry Points

I just mentioned a couple examples of XSS exploit entry points, but those are not the only ones. Basically anywhere you take input from a user and display it back on a web page is an opportunity for an attacker to exploit your site.

Be sure to think about **all** places data enters your system, not just input fields. For example, maybe you allow users to upload images and then re-display the image XIFF data. Maybe you parse uploaded CSV files of various data exports from external programs. Anywhere that you re-display information, it needs to be protected.

How To Protect Yourself

The fix for this is not very difficult. First, never take data directly from a URL and echo it back to the browser. The same goes for data from other sources, like your database or uploaded files. To protect yourself, you simply need to escape data going into your database and escape data being displayed back to your users.

We've already discussed proper database escaping in Chapter 1, so we'll focus on displaying data. PHP makes this very simple with the built-in function `htmlentities()`. This function will properly handle the majority of data.

Let's see what this looks like implemented. Your view file used to look like this:

```
<h1>Title</h1>

Hello <?=$name?>,
```

Now with the protection applied, it looks like this:

```
<h1>Title</h1>

Hello <?=htmlentities($name)?>
```

If you want to stay DRY[29], you would probably abstract this away into a view helper library. Most frameworks already include this in their view or template libraries.

[29] http://en.wikipedia.org/wiki/Don't_repeat_yourself

Cross Site Request Forgery

Cross Site Request Forgery (CSRF) is basically the opposite of an XSS exploit. Where XSS takes advantage of the user by means of a trusted web site, CSRF takes advantage of the web site by means of a trusted user.

An example of this is an attacker sending out fake emails with a link to delete a blog post, an email, or whatever. The target user then clicks the link and arrives at a delete page. Since the user is an administrator with a valid session, your application goes ahead and deletes the record as requested. The user had no idea what the link was taking them to and now their account has been deleted without their consent. Not cool.

This doesn't have to be a text link either; it is often attached to an image or a button. This might sound like a small risk since most critical web site functions are behind forms that expect POSTed data. But this can just as easily be expanded upon to use a button or JavaScript to submit hidden forms.

How To Protect Against It

The first step is to ensure no data-altering actions are performed by GET requests. Anything that performs an action on data should require a POST, PUT, or DELETE request. If the user clicks a delete button, they should then be taken to a form used to confirm the action. If data-altering actions need to be performed over GET (maybe for a RESTful API), you can require a unique token in the query string. In the following examples we will be using POST data but the exact same concepts apply when dealing with GET requests; just set the token in the query string instead of in the POST parameters.

Now that you are submitting forms for your data manipulations, we will need to add CSRF tokens to our forms. Our CSRF token will be a standard Nonce (Number used Once). We will generate a random token, store it in the user's session, then add it as a hidden field to our form. Once the form is POSTed, we can check the CSRF token against the one in the session to validate the request.

First, we'll create a function to generate the token. This will usually be placed in a universally callable place, maybe as a route filter, voter, or a helper library.

```
1  //assuming the rest of the form class here
2
3  static function generateCsrf()
4  {
5    $token = mcrypt_create_iv(
6      16,
7      MCRYPT_DEV_URANDOM
8    );
9
10   Session::flash(
11     'csrfToken',
12     $token
13   );
14
15   return $token;
16 }
```

Note that we are using session flash data here. Flash data will be stored to the session, but can only be accessed on one request before it is destroyed. This concept is supported

in most session wrapping classes. This keeps our token from being valid on more than one request.

Next we'll call this within our route closure and pass the token to the view that is generating the form:

```
Route::get('/signup', function()
{
  $data['token'] = Form::generateCsrf();

  return View::render('signup.form', $data);
});
```

And now for our view, `signup/form.php`

```
<form method="POST" action="/signup">

  <label>
        First Name:
        <input type="text" name="first_name" />
  </label>

  <label>
        Last Name:
        <input type="text" name="last_name" />
  </label>

  <label>
        Email:
        <input type="text" name="email" />
  </label>

  <input type="hidden" name="token" value="<?=$tok\
```

```
en?>" />

  <input type="submit" name="submit" value="Signup\
" />

</form>
```

When this form is POSTed we can now validate the token:

```
Route::post('/signup', function()
{
  //this would probably be abstracted away into
  //a route filter or your form validation
  if ($_POST['token'] === Session::get('csrfToken'\
)) {
        //process the form
  }

  //like earlier, you should add a
  //legit error message here
  die('Invalid Form Data');
});
```

Now that this token checking is in place, if an attacker tricks a user into submitting a fake form, the request will fail. The user will not have a matching CSRF token in their session data for your website.

Multiple Form Submits

Another prolific issue in PHP applications is multiple form submissions. A user submits a form to perform some action â€" let's say transferring 20 Bitcoins from one wallet to another. The user clicks submit, but they don't notice a change immediately, so they click again. Now the user has inadvertently transferred 40 Bitcoins instead of 20. Luckily for us, the CSRF token logic we just discussed will handle this without any extra work. To prevent this in most situations, we just need to pass a unique token, validate it, then clear it once processed.

Race Conditions

Race conditions are not super common in PHP, but are very hard to debug once they occur. It is best to handle them before they happen. A race condition is when multiple things are happening at once, causing unexpected logic flow. The issue is when Block B executes before Block A, due to Block A taking longer to perform.

A basic example is two processes writing to the same file. Without a transactional locking mechanism in place, data is susceptible to corruption. If you expect Process A to be finished writing to a file before Process B starts, data could be overwritten, written on top of other data, or many other types of data corruption. With transactional locking in place, you ensure Process A is finished writing to the file before allowing Process B to write to the same file.

This example is very process-specific, but the main concept of preventing race conditions is to make logic transactional where it should be.

Another example is with database writes. To prevent race conditions in the database, use transactions to apply certain database changes only if all statements are successful.

The general concept is that if only one thing is supposed to happen at a time, you should check to ensure that each step has finished before proceeding to the next step.

Outdated Libraries / External Programs

Another quick item to bring to your attention is outdated libraries. The best way to ensure that your code stays safe is to keep all of your dependencies up to date. No matter how secure the code you personally write is, all it takes is a single security vulnerability in any library you use and your site can be exploited.

Another thing to keep in mind is external programs you use on your server. For example, PhpMyAdmin has had several security flaws throughout its lifetime that have left servers vulnerable. Outdated Wordpress installations are also well known in development circles as a back door for hackers. Any program that exposes critical functions on your server is a possible attack entry point.

Try to keep your external dependencies to a minimum, and always keep them up to date with their latest security releases.

Destructor

I had a lot of fun during the process of writing this. I also drank a lot of rum. I truly hope that you learning something and enjoyed reading this, we need to meet that knowledge to rum ratio.

Please get in contact[30] if you want to learn more about any of the subjects covered here or just argue over the meaning of life.

Thanks for reading.

[30] feedback@buildsecurephpapps.com

About the Author

Ben Edmunds[31] leads development teams to create cutting-edge web and mobile applications. He is an active leader, developer, and speaker in various development communities. He has been developing software professionally for over 10 years and in that time has worked on everything from robotics to government projects.

PHP Town Hall podcast co-host. Portland PHP Usergroup co-organizer. Open source advocate.

Security Audit / Consulting

Need help applying the items discussed in this book? Or would you like someone to go behind you to verify that your web application is secure?

I offer security auditing and consulting on a limited basis each year. If you're interested please get in touch, I can be reached via email at consulting@benedmunds.com[32]

[31] http://benedmunds.com

[32] mailto:consulting@benedmunds.com