

Using OWASP Security Controls

The Unofficial Cookbook

Omar Manzoor
Kevin Wall
Jeffrey Walton

Using OWASP Security Controls: The Unofficial Cookbook

by Omair Manzoor, Kevin Wall, and Jeffrey Walton

Publication date June 28, 2023

Table of Contents

1. Introduction	1
Source code	1
Contributing	1
Organization	1
Conventions	1
Applications	1
2. Input Validation	3
Input Validation	3
Documentation	3
3. Output Contexts	4
Output Contexts	4
4. Cross Site Scripting	7
Documentation	7
Strategy	7
HTML Body	8
HTML Attribute	8
HTML Body vs Attribute	8
JavaScript Block	9
innerHTML	9
URL	10
URI	10
5. SQL Injection	11
Documentation	11
Strategy	11
6. LDAP Injection	12
Documentation	12
Contexts	12
Strategy	12
Distinguished Name	14
Search Filter	15
encodeForLDAPSearchFilter	15
7. XML Injection	18
Documentation	18
Strategy	18
Index	20

List of Tables

3.1. Defense by Output Context	5
4.1. Safe Encoding of Dangerous HTML Characters	7
4.2. XSS Defense by Context	7
5.1. Safe Encoding of Dangerous SQL Characters	11
6.1. Encoding of LDAP Distinguished Name Characters	13
6.2. Encoding of LDAP Filter Characters	13
7.1. Encoding of XML Distinguished Name Characters	18
7.2. Encoding of XML Filter Characters	18

List of Examples

3.1. Incomplete Remediation	5
3.2. Complete Remediation	6
4.1. HTML Body, OWASP Encoder	8
4.2. HTML Body, ESAPI Encoder	8
4.3. HTML Attribute, OWASP Encoder	8
4.4. HTML Attribute, ESAPI Encoder	8
4.5. HTML Body vs Attribute, ESAPI Encoder	9
4.6. JavaScript Block, Pwned	9
4.7. JavaScript Block, OWASP Encoder	9
4.8. JavaScript Block, ESAPI Encoder	9
4.9. innerHTML, OWASP Encoder	10
4.10. innerHTML, ESAPI Encoder	10
4.11. URL, OWASP Encoder	10
4.12. URL, ESAPI Encoder	10
4.13. URI, OWASP Encoder	10
4.14. URI, ESAPI Encoder	10
6.1. LDAP Distinguished Name, ESAPI Encoder	14
6.2. Incorrect LDAP Distinguished Name, ESAPI Encoder	14
6.3. LDAP Search Filter, ESAPI Encoder	15
6.4. Incorrect LDAP Search Filter, ESAPI Encoder	15
6.5. LDAP Search Filter, Aggressive Encoding	16

Chapter 1. Introduction

This document is a guide to using OWASP security controls. The controls help developer teams write safer applications. OWASP provides two libraries to take from, and a team can use either library, depending on what is needed. The first library is the OWASP Java Encoder [<https://owasp.org/www-project-java-encoder/>], and the second library is the Enterprise Security API [<https://owasp.org/www-project-enterprise-security-api/>].

The OWASP Java Encoder is a collection of security controls for Java 1.5 and above. The project minimizes dependencies to make it easier to integrate the encoder. If the OWASP Encoder fits your needs then you should use it.

The Enterprise Security API (ESAPI) is a collection of security controls for Java 8 and above. The project is feature rich, has more controls and has more dependencies than the OWASP Encoder. The ESAPI project also supports more languages than the OWASP Encoder project. You should use ESAPI when you need rich control and language support.

Source code

The source code for the Java Encoder can be found on the OWASP GitHub at [owasp-java-encoder](https://github.com/OWASP/owasp-java-encoder) [<https://github.com/OWASP/owasp-java-encoder>]. The source code for ESAPI can be found on the ESAPI GitHub at [esapi-java-legacy](https://github.com/ESAPI/esapi-java-legacy) [<https://github.com/ESAPI/esapi-java-legacy>].

Contributing

This book is free software. If you see an opportunity for improvement, an error or an omission then please submit a pull request or open a bug report.

Organization

The book proceeds in eight parts. First, administrivia is discussed, like how to ... TODO: Add something for each chapter of this book.

Conventions

The book uses `Monospace` for program listings and code. Untrusted user input or attacker input is represented as `UNTRUSTED`, as in `encodeForHTML (UNTRUSTED)`, or as a block of text

```
String encoded = ESAPI().encoder().encodeForHTML (UNTRUSTED) ;
```

Applications

This book provides the tools you need to make an application better and safer. A webapp accepts user input, transforms or acts upon the data, and then provides an output. The compa-

ny's business logic is embodied in the transformation, and it is your responsibility to make the data and transformation safe from malicious users. Understanding the webapp, its dataflows, and the security controls available will help you build a better and safer application for the company and users.

Chapter 2. Input Validation

User input should be validated before your webapp consumes the data. This chapter of the book will show you how to use OWASP and ESAPI controls to assist in validating user data.

Input Validation

Validating user input to ensure it conforms to a particular specification is a cornerstone to many developer tasks. A developer will encounter many forms of data, including names, numbers, social security numbers, credit card numbers, and dates. ESAPI provides the `Validator` class to help with the task.

Documentation

Documentation for ESAPI `Validator` class is located at Interface Validator [<https://www.javadoc.io/static/org.owasp.esapi/esapi/2.5.2.0/index.html?org/owasp/esapi/Validator.html>].

Chapter 3. Output Contexts

The security controls used for vulnerable code depends on the output context. The controls used when building a web page will be different from the controls you use when storing data in a database. This chapter of the book will examine output contexts to understand how to remediate a potential vulnerability.

Output Contexts

The example web app below accepts a user input, and upon clicking `Submit`, the client sends the input to the server for further processing. The question is, *what controls or encoder should we use to handle the user's data?*



The answer to the question is, *it depends*. It depends on the output context. Or put another way, it depends on how the user's input is used in an output.

The image below shows the dataflow from the client, to the server, and then to various output contexts. The webapp processes user data five different ways. First, it returns the data to the user in a confirmation webpage. Second, it stores the data in a SQL database. Third, it writes the data to a plaintext log file. Fourth, it prints a copy of the user data. And finally, the webapp sends a confirmation email with the user's data.



Now that we know the output contexts, we can answer the question, *what controls or encoder should we use to handle the user's data?*

The table below shows defense by output context.

Table 3.1. Defense by Output Context

Context	Defense
HTML webpage	XSS encoding.
SQL database	Parameterized queries.
Application logging	Newline sanitization.
Line printer	None. Printing does not suffer injections.
Plaintext email	None. Plaintext email does not suffer injections.

As Defense by Output Context shows, injections are remediated using one of several methods, depending on the output context. A HTML webpage returned to the user will use XSS encoding. (XSS Defense by Context provides more details, depending further on context). A SQL query or insertion will use Prepared Statements or Parameterized Queries. Application logging will use Newline sanitization. And two output contexts — the printer and plaintext email — do not require any remediations because they do not suffer from injection attacks.

The code shown below in Example 3.1, “Incomplete Remediation” *will not* work as expected. It is an incomplete remediation because it only addresses injections in the HTML body of the webpage returned to the user.

Example 3.1. Incomplete Remediation

```
Encoder encoder = ESAPI.encoder();
String userText = encoder.encodeForHTML(
    request.getParameter("userText"));

// userText returned in a response webpage. Safe for use.
ConfirmUserText(userText);

// userText stored in SQL database. Not safe for use.
StoreUserText(userText);

// userText logged to log file. Not safe for use.
LogUserText(userText);

// userText sent to printer. Who cares?
PrintUserText(userText);

// userText sent in plaintext email. Who cares?
EmailUserText(userText);
```

The code shown below in Example 3.2, “Complete Remediation” *will* work as expected. It is a complete remediation because it addresses injections in all the output contexts, and not just the context of the web page returned to the user.

Example 3.2. Complete Remediation

```
String userText = request.getParameter("userText");

// userText returned in a response webpage. Safe for use.
Encoder encoder = ESAPI.encoder();
ConfirmUserText(encoder.encodeForHTML(userText));

// userText stored in SQL database. Safe for use.
StoreUserTextWithPreparedStatement(userText);

// userText logged to log file. Safe for use.
String scrubbedUserText = userText.replace('\n', '_')
                                   .replace('\r', '_')
                                   .replace("U+000D", "_")
                                   .replace("U+000A", "_");
LogUserText(scrubbedUserText);

// userText sent to printer. Who cares?
PrintUserText(userText);

// userText sent in plaintext email. Who cares?
EmailUserText(userText);
```

Chapter 4. Cross Site Scripting

Cross Site Scripting (XSS) is the scourge of web applications. The vulnerability is a client-side injection attack and occurs when the attacker inserts malicious code into a web page. The malicious code is often sent to the server and later used in a web page provided to subsequent users. This chapter of the book will show you how to use OWASP and ESAPI controls to protect against XSS vulnerabilities.

Documentation

Documentation for OWASP `Encoder` class is located at Class Encoder [<https://javadoc.io/static/org.owasp.encoder/encoder/1.2.3/index.html?org/owasp/encoder/Encoder.html>]. Documentation for ESAPI `Encoder` class is located at Interface Encoder [<https://www.javadoc.io/static/org.owasp.esapi/esapi/2.5.2.0/index.html?org/owasp/esapi/Encoder.html>].

Strategy

The strategy used to defend against XSS is output encoding. Output encoding is context dependent, and different encoder functions will be used depending on how the vulnerable parameter is used. The vulnerable parameter can be present in HTML, URLs, JavaScript, CSS or JSON.

Output encoding translates dangerous characters into a benign representation for display on a web page. The table below shows safe ways to encode dangerous characters.

Table 4.1. Safe Encoding of Dangerous HTML Characters

Character	Decimal	Hexadecimal	HTML Entity	Unicode
" (double quote)	"	"	"	\u0022
' (single quote)	'	'	'	\u0027
& (ampersand)	&	&	&	\u0026
< (less than)	<	<	<	\u003c
> (greater than)	>	>	>	\u003e

The table below shows XSS defense by context.

Table 4.2. XSS Defense by Context

Data Type	Context	Defense
String	HTML Body	HTML Entity Encoding
String	HTML Attribute	Minimal Attribute Encoding
String	GET Parameter	URL Encoding
String	URL	URL Validation
String	CSS	Structural Validation; CSS Hex Encoding

Data Type	Context	Defense
HTML	HTML Body	HTML Validation; HTML Sanitizer
N/A	DOM	OWASP Cheatsheet
JavaScript	All	Sandboxing
JSON	Client Parsing	JSON Validation

HTML Body

The primary defense against XSS in HTML body is HTML escaping or Entity encoding. Use the encoding method when the attacker controlled text is displayed to the user in a `<body>` tag.

Examples of encoding using various encoders are shown below.

Example 4.1. HTML Body, OWASP Encoder

```
<body><b><%= Encode.forHTML(UNTRUSTED) %>" /></b></body>
```

Example 4.2. HTML Body, ESAPI Encoder

```
<body><b><%= ESAPI.encoder().encodeForHTML(UNTRUSTED) %>" /></b></body>
```

HTML Attribute

The primary defense against XSS in HTML attribute is HTML escaping or Entity encoding. Use the encoding method when the attacker controlled text is part of an attribute, like in an `<input>` or `<image>` tag.

Examples of encoding using various encoders are shown below.

Example 4.3. HTML Attribute, OWASP Encoder

```
<input type="text" name="data"  
      value="<%= Encode.forHTMLAttribute(UNTRUSTED) %>" />  
  
<input type="text" name="data"  
      value="<%= Encode.forHTMLUnquotedAttribute(UNTRUSTED) %> />
```

Example 4.4. HTML Attribute, ESAPI Encoder

```
<input type="text" name="data"  
      value="<%= ESAPI.encoder().encodeForHTMLAttribute(UNTRUSTED) %>" />
```

HTML Body vs Attribute

The main difference between `encodeForHTML` and `encodeForHTMLAttribute` is, `encodeForHTML` does not encode the space character, while `encodeForHTMLAttribute` encodes the space character as ` `.

An example of HTML Body vs HTML Attribute using the ESAPI encoder is shown below. Notice how the space is handled between the words *Hello* and *World*.

Example 4.5. HTML Body vs Attribute, ESAPI Encoder

```
$ESAPI.encoder().encodeForHTML("<h1>Hello World!</h1>");  
&lt;h1&gt;Hello World!&lt;/h1&gt;  
  
$ESAPI.encoder().encodeForHTMLAttribute("<h1>Hello World!</h1>");  
&lt;h1&gt;Hello&#x20;World!&lt;/h1&gt;
```

JavaScript Block

The primary defense against XSS in JavaScript is avoid doing so. Allowing attacker controlled data directly inside a script is dangerous.

You should avoid code similar to the following example because it is nearly impossible to guard against XSS. DOM-based XSS is nearly impossible to prevent because the controls run client-side, not server-side. Most controls can be bypassed `onerror`.

Example 4.6. JavaScript Block, Pwned

```
<script type="text/javascript">  
    window.setInterval('<%= encode(UNTRUSTED and PWNEED) %>');  
</script>
```

Encoding can be used when the attacker controlled text is part of JavaScript. This remediation includes all DOM-based JavaScript event handler attributes such as `onfocus`, `onclick`, and `onload`.

Examples of encoding using various encoders are shown below.

Example 4.7. JavaScript Block, OWASP Encoder

```
<script type="text/javascript">  
    var msg = "<%= Encode.forJavaScriptBlock(UNTRUSTED) %>";  
</script>
```

Example 4.8. JavaScript Block, ESAPI Encoder

```
<script type="text/javascript">  
    var msg = "<%= ESAPI.encoder().encodeForJavaScript(UNTRUSTED) %>";  
</script>
```

innerHTML

The primary defense against XSS in HTML body is HTML escaping or Entity encoding. Use the encoding method when the attacker controlled text is displayed to the user in a `<body>` tag.

Examples of encoding using various encoders are shown below.

Example 4.9. innerHTML, OWASP Encoder

```
document.getElementById('message').innerHTML =  
    "<%= Encode.forHTML(UNTRUSTED) %>"
```

Example 4.10. innerHTML, ESAPI Encoder

```
document.getElementById('message').innerHTML =  
    "<%= ESAPI.encoder().encodeforHTML(UNTRUSTED) %>"
```

URL

The primary defense against XSS in a URL is URL escaping or percent encoding. Examples of encoding using various encoders are shown below.

Example 4.11. URL, OWASP Encoder

```
Foo bar baz
```

Example 4.12. URL, ESAPI Encoder

```
Foo bar baz
```

URI

The primary defense against XSS in a URI is URL escaping or percent encoding. Examples of encoding using various encoders are shown below.

Example 4.13. URI, OWASP Encoder

```
Foo bar baz
```

Example 4.14. URI, ESAPI Encoder

```
Foo bar baz
```

Chapter 5. SQL Injection

SQL Injection is the scourge of database programming. The vulnerability is a server-side injection attack and occurs when the attacker inserts malicious code into SQL query. This chapter of the book will show you how to use OWASP and ESAPI controls to protect against SQL vulnerabilities.

Documentation

Documentation for Java `PreparedStatement` class is located at Interface `PreparedStatement` [<https://docs.oracle.com/javase/8/docs/api/index.html?java/sql/PreparedStatement.html>].

Strategy

The primary defense against SQL Injections are Parameterized Queries or Prepared Statements. Parameterized Queries and Prepared Statements keep code and data separate to ensure attacker supplied data is not interpreted as code by the underlying database engine.

A secondary defense against SQL Injections is encoding dangerous characters so the attacker supplied data is not interpreted as code by the underlying database engine. Encoding is an anti-pattern and should not be used. Instead, you should use Parameterized Queries or Prepared Statements.

Encoding translates dangerous characters into a benign representations for storage in a database. The table below shows safe ways to encode dangerous characters.

Table 5.1. Safe Encoding of Dangerous SQL Characters

Character	Encoding	Comment
' (single quote)	"	Single quote escaped with a single quote

Chapter 6. LDAP Injection

LDAP Injection is the scourge of directory programming. The vulnerability is a server-side injection attack and occurs when the attacker inserts malicious code into LDAP query. This chapter of the book will show you how to use OWASP and ESAPI controls to protect against LDAP vulnerabilities.

Documentation

Documentation for ESAPI `Encoder` class is located at Interface Encoder [<https://www.javadoc.io/static/org.owasp.esapi/esapi/2.5.2.0/index.html?org/owasp/esapi/Encoder.html>].

Documentation for the two RFCs used in this chapter are located at RFC 4514, *Lightweight Directory Access Protocol (LDAP): String Representation of Distinguished Names* [<https://tools.ietf.org/html/rfc4514>] and RFC 4515, *Lightweight Directory Access Protocol (LDAP): String Representation of Search Filters* [<https://tools.ietf.org/html/rfc4515>].

An additional IETF document, RFC 4512, *Lightweight Directory Access Protocol (LDAP): Directory Information Models* [<https://tools.ietf.org/html/rfc4512>], is used to express the grammars provided in RFC 4514 and RFC 4515.

Contexts

When you program against the directory there are two different contexts you must be aware of. The contexts are LDAP Distinguished Names and LDAP Filter Strings. The context drives the encoder used, like `encodeForDN` and `encodeForLDAP`.

The first context is *LDAP Distinguished Names*, and it is used for LDAP distinguished names (DN). Encoding is specified in RFC 4514, *Lightweight Directory Access Protocol (LDAP): String Representation of Distinguished Names*. RFC 4514 uses the characters from ranges 0x01-0x21, 0x23-0x2A, 0x2D-0x3A, 0x3D, 0x3F-0x5B, 0x5D-0x7F, and hex encode characters outside the ranges.

The second context is *LDAP Filter Strings*, and they are used when searching the directory. Encoding is specified in RFC 4515, *Lightweight Directory Access Protocol (LDAP): String Representation of Search Filters*. RFC 4515 uses the characters from ranges 0x01-0x27, 0x2B-0x5B and 0x5D-0x7F, and hex encode characters outside the ranges.

It is possible to use a more restricted encoder for LDAP search filter strings. The restricted encoder could encode filter operators like `&`, `|`, `:`, `!` and `~`. An example of a more restricted filter encoder is shown below in the section called “`encodeForLDAPSearchFilter`”.

Strategy

The primary defense against LDAP Injections is encoding problematic characters. The characters to encode are context dependent as explained in the section called “Contexts”.

For distinguished names, the characters are provided in the grammar shown in RFC 4514, Section 3 [<https://tools.ietf.org/html/rfc4514#section-3>]. The table below shows the characters and how to encode them for the directory. Some characters fall under the "special" production rule, and they can be escaped with a slash. The remaining characters need hexadecimal encoding. ESAPI will use the special encoding for a character when available.

LUTF1, SUTF1 and TUTF1 are the subset of UTF-8 characters specified by RFC 4514 in leading (LUTF1), middle (SUTF1) and trailing (TUTF1) character positions. Distinctions are made between leading, middle stream and trailing characters like leading and trailing spaces. Leading and trailing spaces need to be escaped to preserve them in the distinguished name.

Table 6.1. Encoding of LDAP Distinguished Name Characters

Character	Encoding	Special	Comment
NUL	\00	None	Not in LUTF1, SUTF1 or TUTF1
SPACE	\20	\SPACE	Not in LUTF1 or TUTF1
"	\22	\"	Not in LUTF1, SUTF1 or TUTF1
#	\23	\#	Not in LUTF1
+	\2B	\+	Not in LUTF1, SUTF1 or TUTF1
,	\2C	\,	Not in LUTF1, SUTF1 or TUTF1
/	\2F	None	Required by Microsoft AD
;	\3B	\;	Not in LUTF1, SUTF1 or TUTF1
<	\3C	\<	Not in LUTF1, SUTF1 or TUTF1
=	\3D	\=	Optional, "special" production rule ^a
>	\3E	\>	Not in LUTF1, SUTF1 or TUTF1
ESC	\5C	\\	Not in LUTF1, SUTF1 or TUTF1
0x80-0xFFFF	\hh	None	Hexadecimal encoded, 2 digits ^b

^aThe EQUAL character is allowed in LUTF1, SUTF1 and TUTF1. However, the grammar has a production rule that allows EQUAL to have "special" encoding, too.

^bThe character is first converted to a UTF-8 multibyte sequence, then each byte is hexadecimal encoded.

For search filter strings, the characters are provided in the grammar shown in RFC 4515, Section 3 [<https://tools.ietf.org/html/rfc4515#section-3>]. The table below shows the characters and how to encode them for the directory. UTF1SUBSET is the subset of UTF-8 characters specified by RFC 4515.

Table 6.2. Encoding of LDAP Filter Characters

Character	Encoding	Comment
NUL	\00	Not in UTF1SUBSET
!	\21	Used in Filter grammar
&	\26	Used in Filter grammar
(\28	Not in UTF1SUBSET
)	\29	Not in UTF1SUBSET

Character	Encoding	Comment
*	\2A	Not in UTF1SUBSET
/	\2F	Required by Microsoft AD
:	\3A	Used in Filter grammar
<	\3C	Used in Filter grammar
=	\3D	Used in Filter grammar
>	\3E	Used in Filter grammar
ESC	\5C	Not in UTF1SUBSET
	\7C	Used in Filter grammar
~	\7E	Used in Filter grammar
0x80-0xFFFF	\hh	Hexadecimal encoded, 2 digits ^a

^aThe character is first converted to a UTF-8 multibyte sequence, then each byte is hexadecimal encoded.

Distinguished Name

The example below shows how to encode three relative distinguished names using ESAPI's `encodeForDN`. Recall the valid character ranges from RFC 4514 are 0x01-0x21, 0x23-0x2A, 0x2D-0x3A, 0x3D, 0x3F-0x5B, 0x5D-0x7F, and characters outside the ranges are hex encoded.

Note the code below encodes `cn`, `dc2` and `dc1` because the attacker controls the values. The code below does not encode `distinguishedName` because it is controlled by the program, and characters like `"`, `#`, `+`, `;`, `<` and `>` are control characters required by the processor.

Example 6.1. LDAP Distinguished Name, ESAPI Encoder

```
Encoder encoder = ESAPI.encoder();
String cn = encoder.encodeForDN(request.getParameter("cn"));
String dc2 = encoder.encodeForDN(request.getParameter("dc2"));
String dc1 = encoder.encodeForDN(request.getParameter("dc1"));
String distinguishedName = "CN=" + cn + ",DC=" + dc2 + ",DC=" + dc1;
```

Running the code with Domain Component (DC) set to `net`, Domain Component (DC) set to `example` and Common Name (CN) set to `James "Jim" Smith, III` will result in an encoded distinguished name of `CN=James \"Jim\" Smith\\, III,DC=example,DC=net`. Notice the quotes and comma in James Smith's name are escaped with a backslash.

Be certain to use `encodeForDN` on attacker controlled data, and not the entire distinguished name string. If you use `encodeForDN` on the entire distinguished name string, then the control characters will be encoded and modifications or updates on the object will fail. For example, the following code is incorrect, and `distinguishedName` will be ill-formed due to encoding control characters.

Example 6.2. Incorrect LDAP Distinguished Name, ESAPI Encoder

```
Encoder encoder = ESAPI.encoder();
```

```
String cn = request.getParameter("cn");
String dc2 = request.getParameter("dc2");
String dc1 = request.getParameter("dc1");
String dn = "CN=" + cn + ",DC=" + dc2 + ",DC=" + dc1;
String distinguishedName = encoder.encodeForDN(dn);
```

Search Filter

The example below shows how to encode a search filter using ESAPI's `encodeForLDAP`. Recall the valid character ranges from RFC 4515 are 0x01-0x27, 0x2B-0x5B and 0x5D-0x7F, and characters outside the ranges are hex encoded.

Note the code below encodes `cn` because the attacker controls the value. The code below does not encode `searchFilter` because it is controlled by the program, and characters like `(,), =` and `&` are control characters required by the filter.

Example 6.3. LDAP Search Filter, ESAPI Encoder

```
Encoder encoder = ESAPI.encoder();
String cn = encoder.encodeForLDAP(request.getParameter("cn"));
String searchFilter = "(&(objectClass=user)(cn=" + cn + "))";
```

Be certain to use `encodeForLDAP` on attacker controlled data, and not the entire filter string. If you use `encodeForLDAP` on the entire filter string, then the control characters will be encoded and the filter will fail. For example, the following code is incorrect, and `searchFilter` will be ill-formed due to encoding control characters.

Example 6.4. Incorrect LDAP Search Filter, ESAPI Encoder

```
Encoder encoder = ESAPI.encoder();
String cn = request.getParameter("cn");
String filter = "(&(objectClass=user)(cn=" + cn + "))";
String searchFilter = encoder.encodeForLDAP(filter);
```

encodeForLDAPSearchFilter

`encodeForLDAP` showed you how to safely perform a directory search. However, `encodeForLDAP` does not encode filter operators like `&`, `|`, `:`, `!` and `~`. Instead, the filter relies on a bullet proof LDAP parser to parse the input properly.

It is possible to use a more restricted encoder for LDAP search filter strings that can tolerate buggy LDAP parsers. The restricted encoder would encode filter operators like `&`, `|`, `:`, `!` and `~`, and it is shown in Example 6.5, “LDAP Search Filter, Aggressive Encoding”.

`encodeForLDAPSearchFilter` is aggressive in its encoding. The encoder is based on RFC 4515, but it also encodes control characters like `&`, `|`, `:`, `!` and `~`. The extra encoding will provide an additional layer of hardening in case of a buggy LDAP parser.

Example 6.5. LDAP Search Filter, Aggressive Encoding

```
String encodeForLDAPSearchFilter(String input, boolean encodeAsterisk)
{
    if( input == null ) {
        return null;
    }

    StringBuilder sb = new StringBuilder();

    for( int i = 0; i < input.length(); i++ ) {
        final char c = input.charAt(i);
        switch (c) {
            case 0x00: // NUL
                sb.append("\\00"); break;
            case '\\': // ESC
                sb.append("\\5c"); break;
            case '/':
                sb.append("\\2f"); break;
            case '*':
                if (encodeAsterisk) {
                    sb.append("\\2a");
                } else {
                    sb.append(c);
                }
                break;
            case '!':
                sb.append("\\21"); break;
            case '&':
                sb.append("\\26"); break;
            case '(':
                sb.append("\\28"); break;
            case ')':
                sb.append("\\29"); break;
            case ':':
                sb.append("\\3a"); break;
            case '<':
                sb.append("\\3c"); break;
            case '=':
                sb.append("\\3d"); break;
            case '>':
                sb.append("\\3e"); break;
            case '|':
                sb.append("\\7c"); break;
            case '~':
                sb.append("\\7e"); break;
            default:
                if( c >= 0x80 ) {
                    final byte[] u = String.valueOf(c).getBytes("UTF-8");
```

```
        for (byte b : u) {
            sb.append(String.format("\\%02x", b));
        }
    } else {
        sb.append(c);
    }
}
return sb.toString();
}
}
```

Chapter 7. XML Injection

XML Injection is the scourge of XML programming. The vulnerability is a server-side injection attack and occurs when the attacker inserts malicious code into XML document. This chapter of the book will show you how to use OWASP and ESAPI controls to protect against XML vulnerabilities.

Documentation

Documentation for ESAPI `Encoder` class is located at Interface Encoder [https://www.javadoc.io/static/org.owasp.esapi/esapi/2.5.2.0/index.html?org/owasp/esapi/Encoder.html].

Documentation for XML can be found at the W3C's *Extensible Markup Language (XML) 1.0 (Fifth Edition)* [https://www.w3.org/TR/REC-xml/].

Strategy

The primary defense against XML Injections is encoding problematic characters.

Table 7.1. Encoding of XML Distinguished Name Characters

Character	Encoding	Special	Comment
NUL	\00	None	Not in LUTF1, SUTF1 or TUTF1
SPACE	\20	\SPACE	Not in LUTF1 or TUTF1
"	\22	\"	Not in LUTF1, SUTF1 or TUTF1
#	\23	\#	Not in LUTF1
+	\2B	\+	Not in LUTF1, SUTF1 or TUTF1

For search filter strings, the characters are provided in the grammar shown in RFC 4515, Section 3 [https://tools.ietf.org/html/rfc4515#section-3]. The table below shows the characters and how to encode them for the directory. `UTF1SUBSET` is the subset of UTF-8 characters specified by RFC 4515.

Table 7.2. Encoding of XML Filter Characters

Character	Encoding	Comment
NUL	\00	Not in UTF1SUBSET
!	\21	Used in Filter grammar
&	\26	Used in Filter grammar
(\28	Not in UTF1SUBSET
)	\29	Not in UTF1SUBSET

Character	Encoding	Comment
*	\2A	Not in UTF1SUBSET
/	\2F	Required by Microsoft AD
:	\3A	Used in Filter grammar
<	\3C	Used in Filter grammar
=	\3D	Used in Filter grammar
>	\3E	Used in Filter grammar
ESC	\5C	Not in UTF1SUBSET
	\7C	Used in Filter grammar
~	\7E	Used in Filter grammar
0x80-0xFFFF	\hh	Hexadecimal encoded, 2 digits ^a

^aThe character is first converted to a UTF-8 multibyte sequence, then each byte is hexadecimal encoded.

Index

A

Administrivia, 1

C

Contributing, 1
Conventions, 1
Cross Site Scripting, 7
 Documentation, 7
 HTML Attribute, 8
 HTML Body, 8
 innerHTML, 9
 JavaScript, 9
 Strategy, 7
 URI, 10
 URL, 10

D

Design, 1

E

encodeURIComponent, 14
encodeURIComponent, 15
encodeURIComponentSearchFilter, 15

I

Input Validation, 3
 Documentation, 3
Introduction, 1

L

LDAP Injection, 12
 Context, 12
 Documentation, 12
 encodeURIComponent, 14
 encodeURIComponent, 15
 encodeURIComponentSearchFilter, 15
 LUTF1, 13
 Strategy, 12
 SUTF1, 13
 TUTF1, 13
 UTF1SUBSET, 13
LUTF1, 13

O

Output Contexts, 4, 4

Output Encoding, 7

S

Source code, 1
SQL Injection, 11
 Documentation, 11
 Strategy, 11
SUTF1, 13

T

TUTF1, 13

U

UTF1SUBSET, 13, 18

V

Validation, 3

X

XML Injection, 18
 Documentation, 18
 Strategy, 18
 UTF1SUBSET, 18
XSS, 7
 Documentation, 7
 HTML Attribute, 8
 HTML Body, 8
 innerHTML, 9
 JavaScript, 9
 Strategy, 7
 URI, 10
 URL, 10