第 36 卷 第 6 期 计 算 机 工 程 2010 年 3 月
Vol.36 No.6 Computer Engineering March 2010
·安全技术· 文章编号：1000—3428(2010)06—0152—03 文献标识码：A 中图分类号：TP393.08

Web 跨站脚本漏洞检测工具的设计与实现

陈建青，张玉清
(中国科学院研究生院国家计算机网络入侵防范中心，北京 100049)

摘 要：分析跨站脚本漏洞的形成原因，提出从攻击作用位置角度对跨站脚本漏洞进行分类的方法，在此基础上完善跨站脚本漏洞检测模型，实现动态的漏洞检测工具，弥补现有工具的缺陷，检测结果更为完整。实验证明，该工具能有效检测 Web 应用程序中的跨站脚本漏洞，较同类工具更具优越性。

关键词：Web 应用；跨站脚本；漏洞

Design and Realization of Web Cross-site Scripting Vulnerability Detection Tool

CHEN Jian-qing, ZHANG Yu-qing
(National Computer Network Intrusion Protection Center, Graduate University of Chinese Academy of Sciences, Beijing 100049)

【Abstract】 This paper analyzes Cross-Site Scripting(XSS) vulnerability, proposes an XSS vulnerability classification method, optimizes the XSS detecting model, and accomplishes a dynamic auto-detecting tool. It remedies the shortage of the original tool, and carries out a better result. Experiments show its feasibility and advantages compared with similar products.

【Key words】 Web application; Cross-Site Scripting(XSS); vulnerability

1 概述

近年来，Web 应用程序漏洞已经成为互联网上最严重的安全隐患之一。

从表 1 可以看出，在近年来发现的漏洞中，相当一部分属于 Web 应用程序漏洞；而跨站脚本(Cross-Site Scripting, XSS)漏洞又是 Web 应用程序漏洞中数量突出的一类漏洞。一般来说，该类漏洞允许攻击者将危险脚本引入到可信站点之中，从而令用户访问这些站点时在不知情的状态下执行了危险脚本。攻击者继而能够获得这些用户的机密信息，结合其他手段，甚至可以进一步控制用户系统。因此，对跨站脚本漏洞进行深入研究是必要的。

表 1 NVD[1]数据库漏洞比例年表

年份	漏洞总数	Web 漏洞数量	XSS 漏洞数量	XSS 漏洞占 Web 漏洞比例/(%)
2005	4 927	1 856	848	45.69
2006	6 600	3 633	1 290	35.51
2007	6 244	2 546	872	34.25
2008	5 532	2 442	772	31.61

目前，国际上许多组织都在进行跨站脚本漏洞检测的研究工作并推出了相应工具，如 Paros Proxy[2], XSS-Me[3]等。然而，现有工具在漏洞扫描的完整性、准确率等方面依然存在问题。本文通过对跨站脚本漏洞机理的深入分析，提出了跨站脚本漏洞的另一种分类方法，并基于此分类设计并实现了一套挖掘跨站脚本漏洞的工具 WAVFinder(Web Applications Vulnerabilities Finder)。在现实网络环境下，WAVFinder 能够以较高的效率更全面地挖掘跨站脚本漏洞，具有很高的实用价值。

2 跨站脚本漏洞机理和分类

跨站脚本是一种应用层的攻击。如果 Web 应用程序没有对用户输入进行验证便直接将其解析成 HTML 代码，并在网页中显示相应内容，就会引入跨站脚本漏洞。当用户访问到上述包含恶意 HTML 代码的页面时，攻击代码便会执行，从而威胁到用户的安全。具体表现可能有会话劫持和 cookie 毒药、突破内外网安全策略、屏蔽和伪造页面信息、拒绝服务攻击、辅助攻击等[4]。

为了更好地研究跨站脚本漏洞，一般通过对漏洞进行分类来实现。

2.1 依据形成原因的分类

现有的分类方法一般根据漏洞的不同形成原因，将跨站脚本漏洞分为以下 3 类：

(1)反射型跨站脚本漏洞(Reflected XSS)

反射型跨站脚本漏洞出现在 Web 客户端使用 Server 端脚本生成页面为用户提供数据的情况下。如果用户数据被直接包含在页面中，便可能造成客户端代码注入到动态页面中。反射型跨站脚本漏洞是目前利用最为广泛的跨站脚本漏洞，针对它的典型攻击一般通过构造畸形 URL 实现。

(2)存储型跨站脚本漏洞(Stored XSS)

存储型跨站脚本漏洞出现在由 Web 应用程序提供数据的情况下。攻击者设法将包含攻击代码的数据保存于服务器上，之后，只要用户浏览到包含这些数据的网页，便会受到攻击。

基金项目：国家自然科学基金资助项目(60573048, 60773135, 90718007)；国家“863”计划基金资助项目(2007AA01Z427, 2007AA01Z450)

作者简介：陈建青(1983－)，女，硕士研究生，主研方向：网络与信息系统安全；张玉清，教授、博士生导师

收稿日期：2009-11-06 **E-mail：**chenjq@nipc.org.cn

相对反射型跨站脚本漏洞来说，由于攻击代码直接来自 Web 应用程序服务器，存储型跨站脚本漏洞波及范围会更广，带来的影响也更为严重。

(3)基于 DOM 的跨站脚本漏洞[5](DOM Based XSS)

基于 DOM 的跨站脚本漏洞又叫作本地跨站脚本的漏洞，此类型的漏洞存在于页面中客户端脚本自身。当页面中的 JavaScript 代码访问了 URL 请求参数，并未经编码便直接使用相应参数信息在自身所在的页面中输出某些 HTML，就有可能出现此类型的跨站脚本漏洞。

上述分类方法的主要依据在于漏洞的形成原因，但对漏洞利用的细节缺乏深入分析。为了解决这一问题，提出一种新的分类方法：依据攻击的作用位置为跨站脚本漏洞分类。

2.2 依据作用位置的分类

在对上述各类跨站脚本漏洞进行研究的过程中发现，根据用户输入被提交到的位置不同，跨站脚本漏洞可以有 2 种不同的作用方式，相应攻击代码的模式也有所区别。现阶段被普遍认可的通用跨站脚本漏洞检测工具只能够适用于部分漏洞，对另一些漏洞则完全没有效果。因此，应当从另一角度对漏洞进行分类，在进行漏洞检测时对不同类型的漏洞采取不同检测方法，从而保证结果的完整性。

根据攻击向量的作用位置不同，可将跨站脚本漏洞细分为“闭合标签后”(I 类)和“标签属性中”(II 类)2 种。

(1)“闭合标签后”型跨站脚本漏洞(I 类)

此类跨站脚本漏洞一般基于<script></script>标签实现，目前发现的绝大多数跨站脚本漏洞均为 I 类跨站脚本漏洞。对这种漏洞的利用一般分为 2 个步骤：1)依据实际情况将用户输入被提交到的位置之前不完整的 HTML 标签闭合；2)引入站外的一段包含恶意代码的脚本。

代表性的攻击代码为

```
<script>alert('xss')</script>
```

(2)“标签属性中”型跨站脚本漏洞(II 类)

HTML 中嵌入 JavaScript 脚本有 2 种方式：(1)利用<script></script>标签；(2)在某些支持 JavaScript 的属性(如 src 等)中，直接通过“javascript:”伪协议引入并执行 JavaScript 语句。II 类跨站脚本漏洞便是基于后一种方式产生的。它的典型攻击代码为

```
javascript:alert('xss')
```

II 类跨站脚本漏洞多出现在允许用户指定图片或超链接地址的情况下，如论坛等网络应用的用户头像等处。由于 II 类跨站脚本漏洞要求程序将用户输入提交到特定几种属性中，它的存在范围相对较窄；同时，如果允许闭合其所在位置前的标签，它又可以转换为 I 类跨站脚本漏洞。因此，此类跨站脚本漏洞往往没有被特别处理。然而，II 类跨站脚本漏洞作为跨站脚本漏洞的一个分支，应当对其进行深入研究。

首先，II 类跨站脚本漏洞的威胁程度不容忽视。由于此种漏洞中仍然可以执行脚本，因此可直接编写 JavaScript 程序在网页中执行；同时，通过进行 DOM 操作，也可以引入站外的 JavaScript 文件。可见，它能够造成与 I 类漏洞程度相当的威胁。

其次，II 类跨站脚本漏洞可以回避部分输入验证。由于常见的 I 类跨站脚本漏洞中必然存在 HTML 标签，因此，程序员在编写 Web 应用程序时，普遍采用“转义输出尖括号”的方法来避免漏洞的出现。然而，在针对 II 类漏洞的攻击代码中可以回避尖括号，从而绕过某些验证机制。这种手段对 II 类跨站脚本漏洞是无效的。

由此可见，II 类漏洞具有实际的应用意义，对它进行检测是有必要的。

3 跨站脚本漏洞检测原理

依据分析角度不同，现有的针对上述漏洞的挖掘方法主要分为静态分析和动态分析 2 类。

静态分析的对象为文件源代码。对于源码中包含输入参数的代码段，如果用户输入直接与程序输出相关联，就可能存在缺乏过滤的现象，从而产生跨站脚本漏洞。然而，由于静态分析的方法需要直接接触网站源码，并不适用第三方漏洞检测工具所面临的现实网络环境，因此，采用动态分析的方法进行漏洞挖掘。

动态分析方法通过构造包含特定攻击代码的输入尝试实现跨站脚本攻击，通过分析服务器返回的响应信息，可以判断应用程序中是否存在漏洞。由于此方法无需获得源代码，更适用于在客户端进行检测，但它可能存在漏报的现象，因此只能用于检测应用程序中现有的漏洞，而不能作为判定应用程序安全的标准。

动态分析方法通过分析服务器响应信息从而判断其中是否存在漏洞。因此，该方法对反射型跨站脚本漏洞的检测有较好的效果；而对于存储型跨站脚本漏洞来说，用户输入被显示的位置和服务器的响应页面可能不一致，检测起来有一定困难。

具体而言，采用动态分析方法的挖掘工具首先向服务器提交包含攻击代码的畸形输入；然后，挖掘工具尝试捕获服务器响应信息。如果响应信息中存在特定数据，则说明该 Web 应用程序中可能存在跨站脚本漏洞。图 1 说明了对于单个网页的主要漏洞挖掘流程。

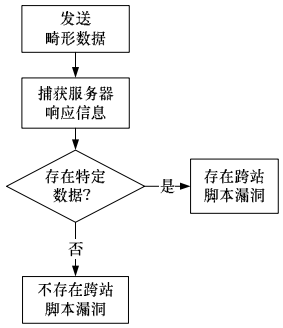

图 1 单个网页跨站脚本漏洞挖掘流程

响应信息中可能存在的特定数据一般是和畸形数据相对应的一段脚本语言代码。一般情况下，只需要检测服务器响应信息中是否存在上述字段，就可以大致判断是否存在跨站脚本攻击漏洞。然而，如果攻击向量的选择不当，或没有充分判断服务器对攻击向量各种可能的响应情况，则有可能出现一定程度的漏报和误报。因此，为保证工具的运行效果，应对攻击向量和其作用结果进行深入分析。

WAVFinder 对于单个网页的挖掘思路也正如上所述。进一步的，WAVFinder 通过内嵌的网络爬虫模块批量地获取一个网站内的所有网页，从而实现了漏洞挖掘的自动化。

4 WAVFinder 的设计与实现

通过分析跨站脚本漏洞检测的原理，本文建立了基于动态分析的反射型跨站脚本漏洞检测模型，并自主开发了工具——WAVFinder。该工具可检测任意网站中存在的 I 类和 II 类反射型跨站脚本漏洞。

该检测模型的设计基于“渗透测试”思路。它从一个攻击者可能存在的位置进行，通过模拟一次真实的攻击事件，分析服务器的业务处理，从而对跨站脚本漏洞进行检测。

模型的结构如图 2 所示，它主要由网络爬虫模块和漏洞检测模块组成。

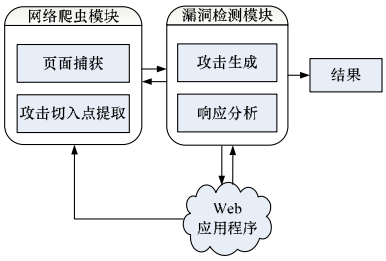

图 2　跨站脚本漏洞检测模型结构

网络爬虫模块实现页面捕获和攻击切入点提取 2 个方面的功能。

获取页面内容是检测的前提。如果将网站的逻辑结构抽象作有向图，那么在给定入口网址(一般使用 Web 应用程序的 root 位置)之后，Spider 爬取网页内容的行为可以视为由指定位置开始、沿页面内链接方向进行图的遍历。为了降低不必要的计算，只需对待处理的链接进行模糊匹配。

在捕获到页面以后，网络爬虫模块还需要负责提取攻击切入点用以测试。用户向 Web 应用程序服务器提交数据的方式主要有 POST 和 GET 这 2 种，它们分别对应 Form 表单和包含参数的 URL。Form 表单的攻击位置在于允许用户提交字符串的域，因此就完整表单，应当过滤非字符串输入的干扰内容。对于含参数的 URL，应进行排除站外链接、提取参数名称等处理。

网页检测模块实现了构造发送攻击包和判断漏洞存在性 2 个功能，通过构造包含畸形数据的攻击包、向服务器发送请求、分析服务器响应等几个步骤，确定检测对象中是否存在跨站脚本漏洞。判断 I 类跨站脚本漏洞的存在性较为简单，只需验证攻击字符串是否存在于服务器响应中即可；而对 II 类漏洞来说，还需要进一步判断攻击向量前后是否存在特定串。

为了协调用户不同安全级别的需要，在消耗时间与结果完整性之间取得平衡，WAVFinder 提供了包含最常见攻击向量的“典型扫描”和全面覆盖各种攻击向量的“全面扫描”2 种功能。

5　问题与改进

衡量跨站脚本漏洞检测工具效果的主要指标有漏报率、误报率和扫描速度等几个因素，它们分别反映了工具的准确率和效率。主要从检测 II 类跨站脚本漏洞和全面考虑字符编码 2 个角度进行改进工作，从而降低工具的漏报率。

实验证明，Paros Proxy, XSS-Me 等跨站脚本漏洞扫描工具均未对 II 类跨站脚本漏洞进行正确处理，这会直接导致漏报的发生。鉴于 II 类跨站脚本漏洞仍然有一定存在范围，且它与 I 类跨站脚本漏洞能够造成同样程度的威胁，WAVFinder 增加了 II 类跨站脚本漏洞的检测功能。

为了正确挖掘出这些跨站脚本漏洞，应当首先提取可能存在脚本执行风险的属性集合，继而判断用户输入是否被提交至上述属性中。

存在 II 类跨站脚本漏洞风险的属性主要有 HREF, SRC/DYNSRC/LOWSRC, ONLOAD, ONCLICK, BACKGROUND 等，其中又以 SRC 最为常见。在 WAVFinder 的具体实现的网页检测模块中，使用了如下的检测正则表达式：

```
String searchExpr = ATTR + "=\\s*?([\\\"']?)" + attackString;
```

其中，ATTR 为上述属性集中的成员；attackString 为攻击字符串的正则表达式形式。

通过上述检查，即可确定是否存在 II 类跨站脚本漏洞。

数据在 Internet 上传播时往往经过编码，如果 Web 应用程序的设计者没有进行正确处理特定的编码格式，便有可能引入跨站脚本漏洞。WAVFinder 在设计时全面的考虑不编码、ASCII 码、HTML 编码等多种编码形式，通过相应的攻击方式，可绕过网站对用户的输入验证，达到更好的检测效果。

此外，由于网站的自有数据可能给检测工作带来干扰因素，WAVFinder 在生成畸形检测数据时向其中加入了随机字符串。这在一定程度上起到了降低漏洞漏报率的作用。

6　测试及结果分析

为检测 WAVFinder 的有效性，从国内外各大网站中选取了若干样本进行了安全扫描，并成功地在 12 个网站中发现了 67 个跨站脚本漏洞。结合其中 3 个漏洞较为集中的网站的检测结果，经过人工验证，WAVFinder 对其中跨站脚本漏洞的判定准确率超过 70%。

在测试过程中，WAVFinder 从某论坛系统中发现了 II 类跨站脚本漏洞。该论坛程序对用户输入的图片地址未能加以妥善处理，通过“预览帖子”功能可能造成脚本执行的影响。当用户预览“[img]JavaScript:alert(666)[/img]”的内容时，便会弹出如图 3 所示的窗口。通过对页面源代码片段(图 4)的分析可以判断，该漏洞为 II 类跨站脚本漏洞。

图 3　某论坛系统“预览帖子”功能中的跨站脚本漏洞

```
<td>
  <span class="bold">1</span>
  <br><br>
  <img src="JavaScript:alert(666)" border="0" onloa
  onMouseOver="if(this.alt) this.style.cursor='
  <br>
</td>
```

图 4　JavaScript 注入点

Paros 和 XSS-Me 分别是代理服务器模式和浏览器插件模式 Web 应用程序漏洞挖掘工具的代表性产品。在政府网站和门户网站中各选取了一个站点作为代表，结合存在 II 类跨站脚本漏洞的样本，将 WAVFinder 与同类工具的扫描结果进行了对比。其结果如表 2 所示。

表 2　XSS 漏洞测试数据比较

工具	某政府网站		某门户网站		II 类漏洞识别
	数量	耗时/s	数量	耗时/s	
WAVFinder	5	49	8	129	成功
Paros Proxy 3.2.13	3	152	1	4	失败
XSS-Me 0.4.0	程序错误		程序错误		失败

由于对跨站脚本漏洞种类认识不够全面，Paros Proxy 和 XSS-Me 均无法识别 II 类跨站脚本漏洞。此外，Paros Proxy 不支持多线程检测，因此，它对某政府网站进行扫描所消耗的时间较长。WAVFinder 与这些产品相比，体现出工作效率高、适用范围较广等特点。

(下转第 157 页)