AntiSamy是什么？

OWASP AntiSamy 项目可以有好几种定义。从技术角度看，它是一个可确保用户输入的 HTML/CSS 符合应用规范的 API。也可以这么说，它是个确保用户无法在 HTML 中提交恶意代码的 API，而这些恶意代码通常被输入到个人资料、评论等会被服务端存储的数据中。在 Web 应用程序中，“恶意代码”通常是指 Javascript。同时层叠样式表（CSS）在调用 Javascript 引擎的时候也会被认为是恶意代码。当然在很多情况下，一些“正常”的 HTML 和 CSS 也会被用于恶意的目的，所以我们也会对此予以处理。

冷静地说，AntiSamy 项目是违背现代安全机制的。因为出于安全考虑，安全机制和用户之间的沟通应该是单向的。而让潜在的攻击者知道验证的细节通常是不明智的，这样会让攻击者学习和探查到系统运行机制以找到薄弱环节。这种类型的信息泄露可能会造成意想不到的后果。如果一个登录系统，告诉访问者“用户名不存在”，这就泄露出一个事实：这个用户名在系统中确实不存在。攻击者可以使用一个字典，一个电话本，或者二者结合，在远程得到一个有效的用户列表。利用这些信息，攻击者可以使用暴力穷举破解进行攻击，或者引发大规模的账户锁定从而造成拒绝服务攻击。这是很常见的攻击形式。

但是，这种策略在当前的情况下并不实用。普通的互联网用户基本上都不怎么擅长写 HTML 和 CSS，那么他们从哪儿获取自己需要的 HTML 代码呢？通常他们只是从其它 Web 页面上复制一些内容。简单地拒绝输入，而不提供任何信息，这会让用户感到烦恼和沮丧。愤怒的用户会转去别的社交网站活动。

OWASP 的授权许可政策（详见 OWASP 网站的会员 FAQ 页面）允许 OWASP 项目在任何被批准的开源协议下发布。基于此，AntiSamy 项目遵照 BSD 开源协议进行发布。

作者是谁？

AntiSamy 项目最早由 Arshan Dabirsiaghi (arshan.dabirsiaghi @ gmail.com)在 Jason Li (li.jason.c @gmail.com)的帮助下创立，他们两位都是 Aspect Security (http://www.aspectsecurity.com/)成员.

AntiSamy的Java, .Net和其它语言版本之间有什么区别？

这个页面（http://www.owasp.org/index.php/AntiSamy_Version_Differences）上你可以看到不同版本之间的大致区别。由于这是一个没有资金来源的开源项目，所以不能期待所有版本的实现与预期功能都完全一致。如果有某个版本有一些功能缺失，请告诉我们。我们会尽力满足，或者写一个补丁。

我该如何开始？

在集成 AntiSamy 的过程中一共有 4 步，每一步的详述在下一个章节。主要层次的描述如下：：

1. 从 Google Code 下载 AntiSamy
2. 从标准策略文件中选择一个接近您所需的功能选项

    - antisamy-slashdot.xml
    - antisamy-ebay.xml
    - antisamy-myspace.xml
    - antisamy-anythinggoes.xml

3. 根据站点的需求定制策略文件
4. 在代码中调用 AntiSamy 的 API

第一步 – 下载AntiSamy

下面的介绍主要是针对 AntiSamy 的 Java 版本，如果你要看.Net 版本，请参考 AntiSamy .Net。

你可以根据需要下载相应的 AntiSamy 包。如果你想查看代码或者扩展它的功能，那么请下载源代码包。如果要集成 AntiSamy，那你可以下载 lib 包或者或者在你的架构中引用 Maven。。如果你想用 Maven，这里有一个引用 AntiSamy 的 POM 示例。如果你需要 jar 文件，那么下载 antisamy-X.X.X.jar（在 1.2 版本之前有个很容易混淆的 jar 叫做 antisamy-standalone-X.X.X.jar），它只包含了 AntiSamy 自身的库，适合成熟的企业级应用环境，而不用担心由于引进 AntiSamy 而引起 classpath 冲突。

对 1.2 之前的版本还可以选择下载 antisamy-standalone-X.X.X.jar，它不仅包含了 AntiSamy 代码，而且也包含了所有引用的支持库。这种方式适用于这样的项目，它不会用到与 AntiSamy 所引用的相同支持库的项目，因而避免了 classpath 和版本冲突。

为了方便起见，我们在下载页面中的 antisamy-required-libs.zip 中包含了运行 AntiSamy 需要的支持库。

你可以在 Google Code 下载 AntiSamy。

第二步 – 选择一个基准策略文件

一般情况下，你可以在预定义的策略文件中找到一个与你站点需求大致匹配的 AntiSamy 策略文件。这些策略各自代表了一个典型的应用场景，来允许用户提交 HTML(可能还有 CSS)内容。让我们具体的看一下这些策略文件：

1) antisamy-slashdot.xml

Slashdot (http://www.slashdot.org/)是一个提供技术新闻的网站，它允许用户用有限的 HTML 格式的内容匿名回帖。Slashdot 不仅仅是目前同类中最酷的网站之一，而且同时也曾是最容易被成功攻击的网站之一。更不幸的是，导致大部分用户遭受攻击的原由是臭名昭着的 goatse.cx 图片(请你不要刻意去看)。Slashdot 的安全策略非常严格：用户只能提交下列的 html 标签：<b>, <u>, <i>, <a>, <blockquote>，并且还不支持 CSS.

因此我们创建了这样的策略文件来实现类似的功能。它允许所有文本格式的标签来直接修饰字体、颜色或者强调作用。

2) antisamy-ebay.xml

众所周知，eBay (http://www.ebay.com/)是当下最流行的在线拍卖网站之一。它是一个面向公众的站点，因此它允许任何人发布一系列富 HTML 的内容。我们对 eBay 成为一些复杂 XSS 攻击的目标，并对攻击者充满吸引力丝毫不感到奇怪。由于 eBay 允许输入的内容列表包含了比 Slashdot 更多的富文本内容，所以它的受攻击面也要大得多。下面的标签看起来是 eBay 允许的（eBay 没有公开标签的验证规则）: <a>,...

3) antisamy-myspace.xml

MySpace (http://www.myspace.com/)是最流行的一个社交网站之一。用户允许提交几乎所有的他们想用的 HTML 和 CSS，只要不包含 JavaScript。MySpace 现在用一个黑名单来验证用户输入的 HTML，这就是为什么它曾受到 Samy 蠕虫攻击(http://namb.la/)的原因。Samy 蠕虫攻击利用了一个本应该列入黑名单的单词(eval)来进行组合碎片攻击的，其实这也是 AntiSamy 立项的原因。

4) antisamy-anythinggoes.xml

我也很难说出一个用这个策略文件的用例。如果你想允许所有有效的 HTML 和 CSS 元素输入（但能拒绝 JavaScript 或跟 CSS 相关的网络钓鱼攻击），你可以使用这个策略文件。其实即使 MySpace 也没有这么疯狂。然而，它确实提供了一个很好的参考，因为它包含了对于每个元素的基本规则，所以你在裁剪其它策略文件的时候可以把它作为一个知识库。

第三步 – 裁剪策略文件

一些小的组织可能使用默认的设置来部署 AntiSamy，而另外一些站点很可能需要严格的，业务驱动的规则来约束用户输入。在决定如何裁剪策略文件上应该考虑到受攻击面 – 它随着策略文件的大小增加而增加。

你还可以使用/修改一些“指令”，它们基本是一些高级的用户选项。http://www.owasp.org/index.php/AntiSamy_Directives 会告诉你有哪些命令以及支持它们的版本。

第四步 – 调用AntiSamy API

AntiSamy 非常容易使用。下面是一个根据策略文件调用 AntiSamy 的例子。

```
import org.owasp.validator.html.*;

Policy policy = Policy.getInstance(POLICY_FILE_LOCATION);

AntiSamy as = new AntiSamy();
CleanResults cr = as.scan(dirtyInput, policy);

MyUserDAO.storeUserProfile(cr.getCleanHTML()); // 一些自定义功能
```

有几种创建 Policy 实例的方式。`getInstance()`方法可以接受下面任意一种参数：

- 文件名字符串
- java.io.File 对象
- java.io.InputStream 对象

Policy 文件可以通过把文件名作为 `AntiSamy:scan()`的第二个参数来传递给 AnitySamy，如下所示：

```
AntiSamy as = new AntiSamy();
CleanResults cr = as.scan(dirtyInput, policyFilePath);
```

也可以通过创建一个 Policy 文件的 File 对象作为 `AntiSamy:scan()`的第二个参数来传递，如下所示：

```
AntiSamy as = new AntiSamy();
CleanResults cr = as.scan(dirtyInput, new File(policyFilePath));
```

第五步 – 分析CleanResults

CleanResults 对象提供了很多有用的信息：

`getErrorMessages()` –字符串错误信息列表

`getCleanHTML()` – 安全的 HTML 输出

`getCleanXMLDocumentFragment()` –从 `getCleanHTML()`得到的安全的 `XMLDocumentFragment` 片段

`getScanTime()` – 返回扫描时间(秒)

产品路线图

下面的章节描述了 AntiSamy 项目的在各种不同编程语言下的版本进展情况。

Grails

Daniel Bower 创建了 AntiSamy 的 Grails plugin。

.NET

.Net 版本的 AntiSamy 在 http://www.owasp.org/index.php/Category:OWASP_AntiSamy_Project_.NET 可以找到。这个项目是在 Summer of Code 2008 时创立的并且由 Jerry Hoff 开发。

这个版本目前处于停滞状态，现正在寻找一些优秀程序员进行开发以便满足新功能上的需求。如果它不能满足你的需求，不妨可以考虑一下微软的 AntiXSS 库。

Python

Python 版本现在正在由几个不同的组展开原型开发，如果有更多的消息，我们会在这里公布。如果你对这个项目感兴趣，请联系项目邮件列表。

PHP

虽然一开始就计划出 PHP 版本，但我们现在推荐你在 PHP 项目中使用 HTMLPurifier 来验证富文本输入。

有关AntiSamy的演讲

来自 OWASP & WASC AppSec U.S. 2007 Conference (San Jose, CA): AntiSamy - Picking a Fight with XSS (ppt) - by Arshan Dabirsiaghi – AntiSamy 项目主管

来自 OWASP AppSec Europe 2008 (Ghent, Belgium): The OWASP AntiSamy project (ppt) - by Jason Li - AntiSamy 项目贡献者

来自 OWASP AppSec India 2008 (Delhi, India): Validating Rich User Content (ppt) - by Jason Li - AntiSamy 项目贡献者

来自 Shmoocon 2009 (Washington, DC): AntiSamy - Picking a Fight with XSS (pptx) - by Arshan Dabirsiaghi - AntiSamy 项目主管

联系我们

你可以通过 邮件列表以及直接联系项目主管两种方式来获取项目信息。

OWASP AntySamy邮件列表

邮件列表的地址是 https://lists.owasp.org/mailman/listinfo/owasp-antisamy。这个邮件列表最开始不是公开的，并且在第一个版本发布的时候进行了清除。我们鼓励所有潜在和现有用户，以及无聊的攻击者加入这个列表讨论。我们欢迎可以对各种可能的攻击集思广益，讨论相关的正则表达式以及系统的集成。

直接给项目主管写信

如果有些内容不方便在邮件列表中提起，你可以直接联系 AntiSamy 的项目主管 Arshan Dabirsiaghi ， 邮箱是 arshan.dabirsiaghi@ aspectsecurity.com

问题跟踪

请访问 Google Code issue tracker.

赞助

最初的 Java 项目由 OWASP Spring Of Code 2007 发起。.Net 项目由 OWASP Summer of Code 2008 发起。

项目评估

本项目由 Jeff Williams 评定，他的评语在 http://spreadsheets.google.com/ccc?key=pAX6n7m2zaTW-JtGBqixbTw#gid=0 可以找到。

子项目

该项目有两个子项目：

- [×] OWASP AntiSamy Project .Java (0)
- [+] OWASP AntiSamy Project .NET (0)