(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109040097 A
(43)申请公布日 2018.12.18

(21)申请号 201810967299.6

(22)申请日 2018.08.23

(71)申请人 彩讯科技股份有限公司
地址 518057 广东省深圳市南山区科技南十二路18号长虹科技大厦4楼01-11单元

(72)发明人 杨良志 白琳 汪志新 丁德平 瞿勇金

(74)专利代理机构 北京品源专利代理有限公司 11332
代理人 孟金喆

(51)Int.Cl.
H04L 29/06(2006.01)

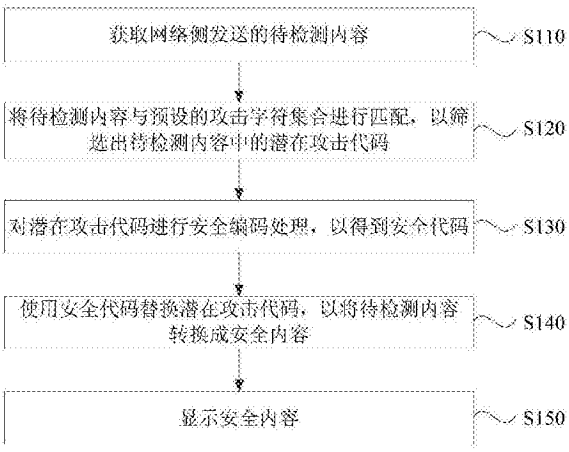
权利要求书2页 说明书12页 附图4页

(54)发明名称

一种跨站脚本攻击的防御方法、装置、设备和存储介质

(57)摘要

本发明公开了一种跨站脚本攻击的防御方法、装置、设备和存储介质。该方法包括:获取网络侧发送的待检测内容;将所述待检测内容与预设的攻击字符集合进行匹配,以筛选出所述待检测内容中的潜在攻击代码;对所述潜在攻击代码进行安全编码处理,以得到安全代码;使用所述安全代码替换所述潜在攻击代码,以将所述待检测内容转换成安全内容;显示所述安全内容。本发明解决了现有技术中无法对XSS攻击进行动态安全防御的技术问题,达到了可根据用户需求对预设的攻击字符集合进行配置,以确保用户所输入的待检测内容符合应用程序编程接口的应用规范,以有效防止XSS攻击的技术效果。



CN 109040097 A

1.一种跨站脚本攻击的防御方法,其特征在于,包括:

获取网络侧发送的待检测内容;

将所述待检测内容与预设的攻击字符集合进行匹配,以筛选出所述待检测内容中的潜在攻击代码;

对所述潜在攻击代码进行安全编码处理,以得到安全代码;

使用所述安全代码替换所述潜在攻击代码,以将所述待检测内容转换成安全内容;

显示所述安全内容。

2.根据权利要求1所述的跨站脚本攻击的防御方法,其特征在于,所述将所述待检测内容与预设的攻击字符集合进行匹配,以筛选出所述待检测内容中的潜在攻击代码包括:

将所述待检测内容与预设的攻击代码集合匹配,以筛选出所述待检测内容中的攻击代码;和/或,

将所述待检测内容与预设的变种代码集合匹配,以筛选出所述待检测内容中的变种代码。

3.根据权利要求2所述的跨站脚本攻击的防御方法,其特征在于,预设的攻击代码集合包括下述至少一项:攻击关键字集合、攻击关键字符集合以及攻击标签集合。

4.根据权利要求2所述的跨站脚本攻击的防御方法,其特征在于,预设的变种代码集合包括下述至少一项:原始标签结构变种代码集合、fromCharCode编码变种代码集合、大小写转换变种代码集合、JavaScript转码变种代码集合、标签事件触发变种代码集合、HTML转码变种代码集合、混合型变种代码集合、CSS文本变种代码集合以及CSS属性值变种代码集合。

5.根据权利要求1所述的跨站脚本攻击的防御方法,其特征在于,获取网络侧发送的待检测内容之后,还包括:

利用预设的安全规则表达式对所述待检测内容进行扫描,以确定所述待检测内容是否符合所述预设的安全规则表达式;

若所述待检测内容不符合所述预设的安全规则表达式,则执行将所述待检测内容与预设的攻击字符集合进行匹配,以筛选出所述待检测内容中的潜在攻击代码的操作。

6.根据权利要求1所述的跨站脚本攻击的防御方法,其特征在于,所述安全编码处理包括下述至少一项:HTML字符转义、URL转码验证、CSS Hex编码、JavaScript Hex编码、JavaScript特殊字符转义、JavaScript敏感字转义以及Unicode编码。

7.根据权利要求1所述的跨站脚本攻击的防御方法,其特征在于,获取网络侧发送的待检测内容包括:

获取网络侧发送的待检测内容以及发送者的账号信息;

将所述账号信息与预设的黑名单进行匹配,以筛选出包含在黑名单中的账号信息;

提取所述黑名单中的账号信息对应的待检测内容,以执行将所述待检测内容与预设的攻击字符集合进行匹配,以筛选出所述待检测内容中的潜在攻击代码的操作。

8.一种跨站脚本攻击的防御装置,其特征在于,包括:

获取模块,用于获取网络侧发送的待检测内容;

匹配筛选模块,用于将所述待检测内容与预设的攻击字符集合进行匹配,以筛选出所述待检测内容中的潜在攻击代码;

预处理模块,用于对所述潜在攻击代码进行安全编码处理,以得到安全代码;

转换模块，用于使用所述安全代码替换所述潜在攻击代码，以将所述待检测内容转换成安全内容；

显示模块，用于显示所述安全内容。

9.一种跨站脚本攻击的防御设备，其特征在于，包括：存储器以及一个或多个处理器；

所述存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-7中任一所述的跨站脚本攻击的防御方法。

10.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1-7中任一所述的跨站脚本攻击的防御方法。

一种跨站脚本攻击的防御方法、装置、设备和存储介质

技术领域

[0001] 本发明实施例涉及网络安全技术，尤其涉及一种跨站脚本攻击的防御方法、装置、设备和存储介质。

背景技术

[0002] 在网络发达的当今时代，脚本语言在网络应用中被大量使用。但随之而来的XSS(cross-site scripting，跨站脚本)攻击成为了目前互联网最为严重的安全问题之一。

[0003] 目前，为了解决XSS攻击，一般站在攻击者的角度来思考并提高程序的安全性，即提高开发人员的安全意识，杜绝一切输入源中可能存在的恶意代码。一般而言，开发人员采用一个安全过滤器对恶意代码进行全局拦截处理，以保证网络的安全。

[0004] 但在实际操作中，用户在web前端输入的内容是无规范的。当用户以不同的规范在web前端输入内容时，XSS攻击发生的位置也是不同的。因此，对XSS攻击的防御是一个动态的过程，而如何对XSS攻击进行动态防御是目前需要解决的一个难题。

发明内容

[0005] 本发明提供一种跨站脚本攻击的防御方法、装置、设备和存储介质，以解决现有技术中无法对XSS攻击进行动态防御的技术问题。

[0006] 第一方面，本发明实施例提供了一种跨站脚本攻击的防御方法，包括：

[0007] 获取网络侧发送的待检测内容；

[0008] 将所述待检测内容与预设的攻击字符集合进行匹配，以筛选出所述待检测内容中的潜在攻击代码；

[0009] 对所述潜在攻击代码进行安全编码处理，以得到安全代码；

[0010] 使用所述安全代码替换所述潜在攻击代码，以将所述待检测内容转换成安全内容；

[0011] 显示所述安全内容。

[0012] 第二方面，本发明实施例还提供了一种跨站脚本攻击的防御装置，包括：

[0013] 获取模块，用于获取网络侧发送的待检测内容；

[0014] 匹配筛选模块，用于将所述待检测内容与预设的攻击字符集合进行匹配，以筛选出所述待检测内容中的潜在攻击代码；

[0015] 预处理模块，用于对所述潜在攻击代码进行安全编码处理，以得到安全代码；

[0016] 转换模块，用于使用所述安全代码替换所述潜在攻击代码，以将所述待检测内容转换成安全内容；

[0017] 显示模块，用于显示所述安全内容。

[0018] 第三方面，本发明实施例还提供了一种跨站脚本攻击的防御设备，包括：存储器以及一个或多个处理器；

[0019] 所述存储器，用于存储一个或多个程序；

[0020] 当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如第一方面所述的跨站脚本攻击的防御方法。

[0021] 第四方面，本发明实施例还提供了一种一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如第一方面所述的跨站脚本攻击的防御方法。

[0022] 本发明通过获取网络侧发送的待检测内容，并将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码；对潜在攻击代码进行安全编码处理，以得到安全代码；使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容；显示安全内容的技术手段，解决了现有技术中无法对XSS攻击进行动态防御的技术问题，实现了可根据用户需求对预设的攻击字符集合进行配置，以确保用户所输入的待检测内容符合应用程序编程接口的应用规范，从而有效防止了XSS攻击的技术效果。

附图说明

[0023] 图1是本发明实施例一提供的一种跨站脚本攻击的防御方法的流程图；

[0024] 图2是本发明实施例二提供的一种跨站脚本攻击的防御方法的流程图；

[0025] 图3是本发明实施例三提供的一种跨站脚本攻击的防御方法的流程图；

[0026] 图4是本发明实施例四提供的一种跨站脚本攻击的防御装置的结构框图；

[0027] 图5是本发明实施例五提供的一种跨站脚本攻击的防御设备的硬件结构示意图。

具体实施方式

[0028] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释本发明，而非对本发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部结构。

[0029] 实施例一

[0030] 图1是本发明实施例一提供的一种跨站脚本攻击的防御方法的流程图，本实施例可适用于对跨站脚本攻击进行防御的情况，该方法可以由跨站脚本攻击的防御设备来执行，该跨站脚本攻击的防御设备可以通过软件和/或硬件的方式实现，该跨站脚本攻击的防御设备可以是两个或多个物理实体构成，也可以是一个物理实体构成。

[0031] 一般而言，XSS攻击包括：反射型(Non-persistent)XSS攻击、存储型(persistent)XSS攻击和DOM-base型XSS攻击。具体来说，反射型XSS攻击的恶意代码一般存在用于输入网址的地址栏中，在用户点击一个链接到目标网站的恶意链接时来实施攻击。存储型XSS攻击的恶意代码一般被保存到目标网站的服务器中，比如常见的场景是博客、论坛等社交网站，即在博客、论坛中的帖子、访问日志或留言评论中保存有恶意代码。当用户向服务器请求获取这些存储信息时，就获取到了该存储信息中所包含的恶意代码，也可理解存储型XSS攻击是用户在浏览带有恶意代码的“正常页面”时进行触发。DOM-based XSS攻击是一种基于DOM(Document Objeet Model，文档对象模型)的漏洞，不需要服务器解析响应的直接参与，而是当用户通过交互修改浏览器页面中的DOM并显示在浏览器上时，就有可能产生这种漏洞。

[0032] 具体来说，跨站脚本攻击的防御是一个动态的概念，在XSS攻击的防御过程中，由于用户的不同业务需求，不能将用户所输入的某些内容进行完全过滤，则可以在服务器进行Web安全编码，以将有害的数据转换为无害的数据。但由于用户输入的内容是没有规律性

的，导致发生XSS攻击的位置是不同的，则需要根据用户需求对存储有攻击代码的策略文件进行定制化配置，以便于用户实时调整策略文件，进而达到既能满足安全需要，即正确处理非法的和不安全的数据，又能满足业务需要，从而不会错误拦截或处理正常业务数据的目标。

[0033] 其中，跨站脚本攻击的防御，可以理解为将一段JS/CSS/HTML富文本经过多个检测链，而每个检测链按照一定的检测规则与拦截规则对富文本进行过滤匹配的过程。其中，检测链可以为预设的攻击代码集合，也可为预设的变种代码集合。相对应的，检测规则与拦截规则，可以理解为与预设的攻击代码集合，或预设的变种代码集合进行匹配的规则。在本实施例中，策略文件可以理解为预设的攻击代码集合。

[0034] 为了适应上述需求，本方案中提出了一种跨站脚本攻击的防御方法。

[0035] 具体地，如图1所示，本实施例提供的跨站脚本攻击的防御方法具体包括如下步骤：

[0036] S110、获取网络侧发送的待检测内容。

[0037] 其中，网络侧，可以理解为服务器端。具体来说，服务器端可以与多个客户端进行通信连接，在本实施例中，通过服务器端与其他客户端进行数据传输。其中，服务器端为在服务器的这一端。一般而言，客户端与服务器端相对应，为用户提供本地服务的程序，在因特网发展以后，经常使用的客户端包括网页浏览器、收寄电子邮件时的电子邮件客户端，以及即时通讯的客户端软件等。其中，待检测内容为用户在其他客户端所输入的web内容。具体的，待检测内容的形式可以根据用户的实际操作情况进行设置，示例性地，待检测内容可以为用户通过贴吧、博客、微博等自媒体平台发布的文章，以及收寄的邮件内容等，也可以为用户在自媒体平台上的留言评论等，对此并不进行限定。

[0038] S120、将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码。

[0039] 其中，预设的攻击字符集合，可以理解为带有潜在攻击代码的各个字符的集合。其中，潜在攻击代码，可以理解为带有攻击性的恶意代码。一般来说，在预设的攻击字符集合为不同类型的代码集合时，从待检测内容中所筛选出来的潜在攻击代码的类型也是不同的。在本实施例中，预设的攻击字符集合可以包括：预设的攻击代码集合和预设的变种代码集合。具体来说，在预设的攻击字符集合为不同代码集合时，该步骤包括下述不同的方案。具体是：

[0040] 方案一：将待检测内容与预设的攻击代码集合匹配，以筛选出待检测内容中的攻击代码。

[0041] 在本实施例中，预设的攻击代码集合包括下述至少一项：攻击关键字集合、攻击关键字符集合以及攻击标签集合。其中，攻击关键字集合可以理解为带有恶意代码的所有关键字的集合；攻击关键字符集合可以理解为带有恶意代码的所有关键字符的集合；攻击标签集合可以理解为带有恶意代码的所有标签的集合。

[0042] 具体来说，在获取到客户端发送的待检测内容时，将待检测内容中的各个代码与预设的攻击代码集合中对应的攻击关键字集合、攻击关键字符集合和/或攻击标签集合进行匹配，以筛选出待检测内容中与预设的攻击代码集合中相匹配的代码，并将相匹配的代码列为待检测内容中的潜在攻击代码，并将其筛选出来，以便于后续对潜在攻击代码进行

安全编码处理。

[0043] 示例性地，假设在某网站中不允许出现<和>这个关键字，同时待检测内容为用户在博客上所发布的一篇文章，并且在文章中出现了<和>这个关键字，则在对这篇文章进行发布时，检测到这篇文章中的<和>与预设的攻击代码中的过滤关键字相匹配，则将这篇文章中的<和>作为待检测内容中的潜在攻击代码，并将其筛选出来，以便于对攻击代码进行安全编码处理。

[0044] 方案二：将待检测内容与预设的变种代码集合匹配，以筛选出待检测内容中的变种代码。

[0045] 一般来说，在获取到网络侧所输入的待检测内容，将待检测内容与预设的攻击代码集合进行匹配，以筛选出待检测内容中的攻击代码，然后再对待检测内容与预设的变种代码集合匹配，以筛选出待检测内容中的变种代码。具体来说，攻击者在客户端上故意尝试输入各种错误的待检测内容，在检测到待检测内容中有恶意代码时，在网页输入页面上返回一个错误提示，此时攻击者接收到错误提示后，会根据错误提示的相关信息对待检测内容的代码进行变种修改，以在对待检测内容与预设的攻击代码集合进行匹配时，无法检测到待检测内容中的恶意代码，从而造成XSS攻击。在本实施例中，为了识别出变种代码，以避免引起XSS攻击，在筛选出待检测内容中的攻击代码之后，对待检测内容的各个代码与预设的变种代码集合进行匹配，以判断待检测内容中的各个代码是否发生变种修改，若检测到有待检测内容中有代码与预设的变种代码集合匹配，则将该代码列为待检测内容中的变种代码，并将其筛选出来，以便于对变种代码进行安全编码处理。

[0046] 在本实施例中，预设的变种代码集合包括：原始标签结构变种代码集合、fromCharCode编码变种代码集合、大小写转换变种代码集合、JavaScript转码变种代码集合、标签事件触发变种代码集合、HTML转码变种代码集合、混合型变种代码集合、CSS文本变种代码集合以及CSS属性值变种代码集合。

[0047] 其中，原始标签结构变种，可以理解为对标签的原始结构进行变种之后得到的代码。fromCharCode()是JavaScript中的一个函数，即fromCharCode()可接受一个指定的Unicode值，然后返回一个字符串，则fromCharCode编码变种可以理解为对所返回的字符串进行变种而得到另一个字符串。大小写转换变种，可以理解为将大写字母转换为小写字母，或者将小写字母转换为大写字母的变种。JavaScript转码变种，可以理解为在对JavaScript编码转换时而发生的变种。标签事件触发变种，可以理解为在触发标签事件时而发生的变种。HTML转码变种，可以理解为在实现HTML转码时而发生的变种。混合型变种，可以理解为既存在各种转码变种，也存在标签事件触发变种。CSS文本变种，可以理解为对CSS文本所对应的代码进行变种。CSS属性值变种，可以理解为对CSS属性值对应的代码进行变种。

[0048] 具体地，在对不同类型的变种代码进行匹配时，需采用不同类型的变种检测方式，示例性地，对待检测内容与预设的变种代码集合进行匹配的方式至少有以下几种：

[0049] (1)</script>闭合当前脚本，然后输入自定义内容进行变种检测，比如：<script>var b="</script><script>alert(document.cookie);//";</script>

[0050] (2)对破坏原始标签结构进行变种检测，示例性地，<IMG><SCRIPT>alert("text")</SCRIPT>

[0051] (3)对fromCharCode编码变种进行检测,示例性地,<IMG SRC＝/onerror＝"alert(String.fromCharCode(88,83,83))"></img>

[0052] (4)对大小写转换变种进行检测,比如:<IMG SRC＝JaVaScRiPt:alert(‘text’)>

[0053] (5)对JavaScript转码变种进行检测,比如:<IMG SRC＝&#0000106&40000097&#0000118&0000097&#0000115&0000099&0000114&>

[0054] (6)对标签事件触发变种进行检测,比如:<IMG SRC＝#onmouseover＝”alert(‘text’)”>

[0055] (7)对HTML转码变种进行检测,比如:<IMG SRC＝javascrip>

[0056] (8)混合型变种检测,比如:<IMG SRC＝”java&#xOD;ascript:alert(‘text’)”>

[0057] (9)对CSS文本变种进行检测,比如:<style>li{list-style-image:url("javascript:alert('css')");}</style>

[0058] (10)对CSS属性值变种进行检测,比如:<BOBY BACKGROUND＝”javascript:alert(‘XSS’)”>

[0059] 示例性地,以大小写转换变种为例,对将待检测内容与预设的变种代码集合匹配,以筛选出待检测内容的变种代码进行说明。假设一网站中不允许出现全为小写字母的一个字符串,但攻击者为了绕过对待检测内容的防御检测,在该字符串中加入一个大写字母,从而在对待检测内容进行防御检测时,认为该字符串不具有攻击性,即未发现该字符串中包含有恶意代码,从而用户在访问该网站时,攻击者可利用字符串中小写字母所包含的恶意代码对窃取cookies,或者读取用户未公开的资料,比如:邮件列表或者内容,系统的客户资料,以及联系人列表等等,从而导致了XSS攻击。若在从待检测内容中筛选出攻击代码之后,对待检测内容中的变种代码进行检测,从而可检测出该待检测内容中包含有大小写转换的变种代码,从而避免了XSS攻击,进而保证了用户信息的安全。

[0060] S130、对潜在攻击代码进行安全编码处理,以得到安全代码。

[0061] 在本实施例中,安全编码处理,可以理解为将潜在攻击代码转换为安全代码的处理操作。其中,安全编码处理可以为数据消毒。具体来说,从待检测内容中筛选出攻击代码和变种代码之后,对攻击代码和变种代码进行数据消毒,比如进行编码、转义或禁用操作。比如,对|,<,>,‘,“,&,#,javascript,expression这些常见字符进行编码与过滤,以使这些字符转换为符合API(Application Programming Interface,应用程序编程接口)的应用规范,即将这些常见字符转换为安全代码。其中,安全代码,可以理解为不具有攻击性的代码。一般来说,安全编码处理包括下述至少一项:HTML字符转义、URL转码验证、CSS Hex编码、JavaScript Hex编码、JavaScript特殊字符转义、JavaScript敏感字转义以及Unicode编码。示例性地,(1)HTML字符转义,比如:将“&”转义字符为&(2)URL转码验证,用于判断待检测内容中是否存在特殊的未转码字符,比如:&,space。(3)CSS Hex编码,即将待检测内容中的数据转换成符合Intel HEX数据格式的编码,以防止转码变种的攻击。(4)JavaScript Hex编码,即通过JS进行Hex编码。(5)JavaScript特殊字符转义,即对JavaScript特殊字符进行转义。(6)JavaScript敏感字转义,即将JavaScript敏感字进行转义。(7)Unicode编码,即按照Unicode规则进行编码。

[0062] S140、使用安全代码替换潜在攻击代码,以将待检测内容转换成安全内容。

[0063] 在本实施例中，将安全编码处理之后所得到的安全代码，依次替换待检测内容中所对应的潜在攻击代码，以使待检测内容转换成安全内容。具体来说，使用安全代码替换潜在攻击代码，也可理解为对潜在攻击代码进行转义、编码等操作，从而使潜在攻击代码转换为安全代码的过程。

[0064] S150、显示安全内容。

[0065] 在本实施例中，在将待检测内容转换成安全内容之后，输出安全编码处理之后得到的待检测内容，并在前端上显示安全内容，以供用户查看和使用。其中，在后端与前端进行交互时，需根据前端与后端提前所约定的json格式进行交互，则在后端对待检测内容进行安全编码处理，并得到安全内容时，后端需先将安全内容转换为json数组的格式，再返回到前端。其中，前端即网站前台部分，运行在PC(personal computer，个人计算机)端、移动端等浏览器上展现给用户浏览的网页。

[0066] 在此需要说明的是，在得到安全内容之后，以Email形式向管理员进行报警。具体来说，对待检测内容中所包含的攻击代码和变种代码进行整理，以及对攻击代码和变种代码进行安全编码处理之后所得到的安全代码进行整理，然后将攻击代码、变种代码和安全代码编辑至邮件内容中，并基于POP(Post Office Protocol，邮局协议)协议将邮件发送至管理员邮箱，以便于管理员根据攻击代码、变种代码以及安全代码对定制化的策略文件进行更新，进而保证了用户信息的安全。

[0067] 本实施例的技术方案，通过获取网络侧发送的待检测内容；将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码；对潜在攻击代码进行安全编码处理，以得到安全代码；使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容；显示安全内容的技术手段，解决了现有技术中无法对XSS攻击进行动态防御的技术问题，达到了可根据用户需求对预设的攻击字符集合进行配置，以确保用户所输入的待检测内容符合应用程序编程接口的应用规范，从而有效防止了XSS攻击的技术效果。

[0068] 实施例二

[0069] 图2是本发明实施例二提供的一种跨站脚本攻击的防御方法的流程图。本实施例是在上述实施例的基础上，对跨站脚本攻击的防御方法作进一步地具体化。如图2所示，本实施例中的跨站脚本攻击的防御方法具体包括如下步骤：

[0070] S210、获取网络侧发送的待检测内容以及发送者的账号信息。

[0071] 其中，发送者的账号信息可以理解为用户在客户端的不同网络平台上的账号信息，比如，用户在博客、微博、贴吧等自媒体平台上的账号信息。在本实施例中，可将发送者的账号信息列为黑名单或者白名单。其中，黑名单或白名单所针对的对象为发送者的账号信息或IP(Internet Protocol，网络之间互连的协议)地址的。在白名单中所包含的账号信息，可以理解为该账号信息所对应的发送者为可信任机构，其所投递的内容中不包含有攻击性的恶意代码。而黑名单中所包含的账号信息，可以理解为该账号信息所对应的发送者为不可信任机构，其所投递的内容中可能包含有攻击性的恶意代码。一般来说，在XSS攻击的防御过程中，遵循白名单放行，黑名单过滤的原则，即在黑名单中所包含的账号信息是不能出现的对象，而在白名单中所包含的账号信息是可以被接受的对象。

[0072] 一般而言，在获取网络侧发送的待检测内容以及发送者的账号信息之前，可根据用户需求对策略文件进行定制化配置。具体来说，在网络安全领域，由于用户不同的业务需

求，不可能将用户所输入内容进行完全过滤，用户可根据输入内容和所需要体现出的效果对策略文件进行配置。比如，网站A可以允许出现关键字符1，而网站B不允许出现关键字符1，则在对网站A中的策略文件进行配置时，可直接将关键字符1列到预设的攻击字符集合，以防御XSS攻击。

[0073] S220、将账号信息与预设的黑名单进行匹配，以筛选出包含在黑名单中的账号信息。

[0074] 其中，黑名单为不允许出现的包含有账号信息的列表。具体来说，在获取到发送者的账号信息时，将账号信息与预设的黑名单中所包含的账号信息进行匹配，若检测到该账号信息与预设的名单中所包含的账号信息一致，则将该账号信息筛选出来。

[0075] 在本实施例中，在账号信息与预设的黑名单中所包含的账号信息不匹配时，则表明该账号信息为白名单中所包含的账号信息。在发送者通过该账号信息所输入的待检测内容为不具有攻击性的恶意代码，可以直接将该待检测内容输出并显示在客户端，以供用户查看。

[0076] S230、提取黑名单中的账号信息对应的待检测内容。

[0077] 在此需要说明的是，在提取黑名单中的账号信息对应的待检测内容之后，有两种方案进行选择，分别为：在提取黑名单中的账号信息对应的待检测内容，可以直接执行步骤S240；当然，在提取黑名单中的账号信息对应的待检测内容，也可以执行将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码的操作。

[0078] S240、利用预设的安全规则表达式对待检测内容进行扫描，以确定待检测内容是否符合预设的安全规则表达式。

[0079] 其中，规则表达式又称为正则表达式，通常用来检索、替换那些符合某个模式/规则的文本。一般来说，正则表达式通常缩写成“regex”，单数有regexp、regex，复数有regexps、regexes、regexen。在本实施例中，安全规则表达式，可以理解为不包含有攻击性的恶意代码的规则表达式。在此说明的是，在待检测内容为不同形式时，安全规则表达式的形式也是不同的。

[0080] 例如：(1)在待检测内容为http链接时，安全规则表达式的规则如下：(\s)*((ht|f)tp(s?):)[\p{L}\p{N}]+[\p{L}\p{N}\p{Zs}\n\.\#\|@*\$%\+&;:\-_～,\?＝/!\(\)\042<>]*(\s)*

[0081] 其中，(\s)*表示匹配零次或多次空白字符；\p{L}表示匹配小写字母字符；\p{N}表示匹配任何类型的数字字符；\n表示匹配换行符；\.\#\|@*\$%\:表示匹配.#|@*$%+;&表示匹配&;:\-_～,\?＝/!\(\)表示匹配;:-_～?＝!()；\042表示匹配八进制的双引号，表示匹配双引号；<>表示匹配<>。

[0082] (2)在待检测内容为HTML标签时，安全规则表达式的规则如下：[a-zA-Z0-9\s,\-_]+

[0083] 其中，a-z，表示允许出现a-z之间的任何小写字母；A-Z，表示允许出现A-Z之间的任何大写字母；0-9表示允许出现0至9之间的任何数字；\s，表示允许出现非换行的所有空白字符；\-，表示允许出现-这个字符。

[0084] 示例性地，HTML标签可以包括：id，class，title，background，align，width等。

[0085] (3)在待检测内容为CSS标签时，安全规则表达式的规则如下：\[[a-zA-Z0-9\-_]+

((=|～=|\|=){1}[a-zA-Z0-9\-_]+){1}\]

[0086] 其中，=，表示匹配=这个字符；{1}表示匹配1次；+表示[a-zA-Z0-9\-_]中的字符可以出现一个或n个。

[0087] 示例性地，CSS标签可以包括：azimuth、background-image、background-position、border-top-color等。

[0088] S250、若待检测内容不符合预设的安全规则表达式，将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码。

[0089] 具体来说，在待检测内容不符合预设的安全规则表达式时，则说明待检测内容中包含有潜在攻击代码，则需对待检测内容进行过滤，以筛选出待检测内容中的潜在攻击代码。具体的过滤方式，可以将待检测内容与预设的攻击代码集合，以及预设的变种代码集合进行匹配，以分别筛选出待检测内容中的攻击代码和变种代码。其中，具体的匹配筛选过程参见上述实施例中的描述，在此不再赘述。

[0090] 进一步的，若待检测内容符合预设的安全规则表达式，则直接显示待检测内容。

[0091] S260、对潜在攻击代码进行安全编码处理，以得到安全代码。

[0092] S270、使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容。

[0093] S280、显示安全内容。

[0094] 本实施例的技术方案，在上述实施例的基础上，在获取到网络侧发送的待检测内容之后，利用预设的安全规则表达式对待检测内容进行扫描，以确定待检测内容是否符合预设的安全规则表达式，并在待检测内容不符合预设的安全规则表达式时，执行将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码的操作，从而实现了在对待检测内容作进一步过滤，进而提高了跨站脚本攻击的防御的处理速度。

[0095] 实施例三

[0096] 图3是本发明实施例三提供的一种跨站脚本攻击的防御方法的流程图。本实施例是在上述实施例的基础上，作为一个优选实施例对跨站脚本攻击的防御方法进行具体说明。如图3所示，本实施例中的跨站脚本攻击的防御方法的具体操作步骤如下：

[0097] 根据用户需求对策略文件进行配置，以生成XML(eXtensible Markup Language，可扩展标记语言)格式的策略文件。具体的，策略文件可包括以下三种内容：依据全局规则对全局变量进行配置；对黑白名单进行配置；对过滤关键字、过滤关键字符和过滤标签进行配置。其中，根据用户需求对策略文件进行配置，可以理解为用户想在Web前端所要输出的内容，以及所需要体现出的效果而对策略文件进行定制化配置。在此需要说明的是，可将某些过滤关键字、过滤关键字符和/或过滤标签设置为全局变量，从而设置为全局变量的过滤关键字、过滤关键字符和/或过滤标签在所有规则之上。比如，对某个过滤关键字进行识别，若该过滤关键字为全局变量，则可访问所有标签以识别提取出该过滤关键字；而若该过滤关键字为局部变量，则只能访问其中一个或指定的标签，以识别提取出该过滤关键字。其中，过滤关键字、过滤关键字符以及过滤标签分别与上述实施例中的攻击关键字集合、攻击关键字符集合以及攻击标签集合对应，在此不再赘述。即，过滤关键字为攻击关键字集合中的关键字；过滤关键字符为攻击关键字符集合中的关键字符；过滤标签为攻击标签集合中的标签。

[0098] 然后，通过DOM4J解析XML DOM树中的表单元素以解析得到策略文件中配置的黑白

名单、过滤关键字、过滤关键字符和过滤标签，并将黑白名单分别加入黑名单队列和白名单队列，将过滤关键字加入Hash-Map，以及将过滤标签加入过滤链表中。其中，表单元素可以理解为策略文件中的配置项，策略文件为一种XML格式的文件，只有对策略文件进行解析才能得到配置项。解析就是将策略文件加载到程序代码中，以通过对程序代码进行初始化才能得到黑白名单、过滤关键字、过滤关键字符以及过滤标签。在此说明的是，将过滤及检测规则配置到策略文件中的好处，是为了在对过滤及检测规则进行修改时，只对策略文件进行修改，而不需要更改程序代码，然后将配置完成的策略文件加载到程序代码中即可，从而简化了配置文件的修改步骤，提高了用户的使用体验。

[0099] 在获取到用户所输入的待检测内容之后，对待检测内容进行解析，并对输入待检测内容的发送者的账号信息进行检测分析，判断该账号信息是否在预设的黑名单中，若该账号信息在预设的白名单中且不在预设的黑名单中，则该账号信息发送的任何待检测内容都不进行过滤检测，直接将待检测内容输出至前端。而若该账号信息在预设的黑名单中，需对该账号信息所发送的待检测内容进行过滤检测。

[0100] 具体的，将待检测内容经过扫描器，即利用预设的安全规则表达式对待检测内容进行扫描，以确定待检测内容是否符合预设的安全规则表达式，若待检测内容不符合预设的安全规则表达式，则通过过滤及检测规则对待检测内容进行过滤检测，以对过滤关键字、过滤关键字符以及过滤标签进行过滤检测，以筛选出待检测内容中的攻击代码；然后在对待检测内容进行变种代码检测，即将待检测内容与预设的变种代码集合进行匹配，以筛选出待检测内容中的变种代码，并对变种代码进行分类，比如为JS变种、CSS变种或Dom Based变种。其中，过滤及检测规则，可以理解为预设的攻击代码集合。

[0101] 在筛选出待检测内容中的攻击代码和变种代码之后，将待检测内容经过XSS过滤链，以对攻击代码和变种代码进行安全编码处理，其中，安全编码处理依次为HTML字符转义，URL转码验证，CSS Hex编码，JavaScript Hex编码，JavaScript特殊字符转义，JavaScript敏感字转义以及Unicode编码，然后将安全编码处理之后得到的待检测内容返回json数组的格式，并在前端进行输出，以供用户进行查看。

[0102] 本实施例的技术方案，通过对策略文件进行定制化配置，以确保用户输入的待检测内容符合应用程序编程接口的应用规范，有效防止了XSS攻击，进而保证了用户信息的安全。

[0103] 实施例四

[0104] 图4是本发明实施例四提供的一种跨站脚本攻击的防御装置的结构框图，该装置可以由硬件/软件实现。如图4所示，该防御装置包括：获取模块310、匹配筛选模块320、预处理模块330、转换模块340和显示模块350。

[0105] 其中，获取模块310，用于获取网络侧发送的待检测内容；

[0106] 匹配筛选模块320，用于将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码；

[0107] 预处理模块330，用于对潜在攻击代码进行安全编码处理，以得到安全代码；

[0108] 转换模块340，用于使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容；

[0109] 显示模块350，用于显示安全内容。

[0110] 本实施例的技术方案，通过获取网络侧发送的待检测内容；将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码；对潜在攻击代码进行安全编码处理，以得到安全代码；使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容；显示安全内容的技术手段，解决了现有技术中无法对XSS攻击进行动态防御的技术问题，达到了可根据用户需求对预设的攻击字符集合进行配置，以确保用户所输入的待检测内容符合应用程序编程接口的应用规范，从而有效防止了XSS攻击的技术效果。

[0111] 在上述实施例的基础上，匹配筛选模块具体用于：

[0112] 将待检测内容与预设的攻击代码集合匹配，以筛选出待检测内容中的攻击代码；

[0113] 和/或，将待检测内容与预设的变种代码集合匹配，以筛选出待检测内容中的变种代码。

[0114] 在上述实施例的基础上，预设的攻击代码集合包括：攻击关键字集合、攻击关键字符集合以及攻击标签集合。

[0115] 在上述实施例的基础上，预设的变种代码集合包括：原始标签结构变种代码集合、fromCharCode编码变种代码集合、大小写转换变种代码集合、javascript转码变种代码集合、标签事件触发变种代码集合、HTML转码变种代码集合、混合型变种代码集合、CSS文本变种代码集合以及CSS属性值变种代码集合。

[0116] 在上述实施例的基础上，跨站脚本攻击的防御装置，还包括：

[0117] 确定模块，用于利用预设的安全规则表达式对待检测内容进行扫描，以确定待检测内容是否符合预设的安全规则表达式；

[0118] 执行模块，用于若待检测内容不符合预设的安全规则表达式，则执行将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码的操作。

[0119] 在上述实施例的基础上，安全编码处理包括下述至少一项：HTML字符转义、URL转码验证、CSS Hex编码、JavaScript Hex编码、JavaScript特殊字符转义、JavaScript敏感字转义以及Unicode编码。

[0120] 在上述实施例的基础上，获取模块包括：

[0121] 获取单元，用于获取网络侧发送的待检测内容以及发送者的账号信息；

[0122] 匹配筛选单元，用于将账号信息与预设的黑名单进行匹配，以筛选出包含在黑名单中的账号信息；

[0123] 提取执行单元，用于提取黑名单中的账号信息对应的待检测内容，以执行将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码的操作。

[0124] 上述跨站脚本攻击的防御装置可执行本发明任意实施例所提供的跨站脚本攻击的防御方法，具备执行方法相应的功能模块和有益效果。

[0125] 实施例五

[0126] 图5是本发明实施例五提供的一种跨站脚本攻击的防御设备的硬件结构示意图。本发明实施例五中的跨站脚本攻击的防御设备以计算机设备为例进行说明。如图5所示，本发明实施例五提供的计算机设备，包括：处理器410和存储器420、输入装置430和输出装置440。该计算机设备中的处理器410可以是一个或多个，图5中以一个处理器410为例，所述计算机设备中的处理器410、存储器420、输入装置430和输出装置440可以通过总线或其他方式连接，图5中以通过总线连接为例。

[0127] 该计算机设备中的存储器420作为一种计算机可读存储介质，可用于存储一个或多个程序，所述程序可以是软件程序、计算机可执行程序以及模块，如本发明实施例一、二或三所提供跨站脚本攻击的防御方法对应的程序指令/模块(例如，图4所示的跨站脚本攻击的防御装置中的模块，包括：获取模块310、匹配筛选模块320、预处理模块330、转换模块340和显示模块350)。处理器410通过运行存储在存储器420中的软件程序、指令以及模块，从而执行计算机设备的各种功能应用以及数据处理，即实现上述方法实施例中的跨站脚本攻击的防御方法。

[0128] 存储器420可包括存储程序区和存储数据区，其中，存储程序区可存储操作系统、至少一个功能所需的应用程序；存储数据区可存储根据设备的使用所创建的数据等。此外，存储器420可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中，存储器420可进一步包括相对于处理器410远程设置的存储器，这些远程存储器可以通过网络连接至设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0129] 输入装置430可用于接收用户输入的数字或字符信息，以产生与终端设备的用户设置以及功能控制有关的键信号输入。输出装置440可包括显示屏等显示设备。

[0130] 上述跨站脚本攻击的防御设备可执行本发明任意实施例所提供的跨站脚本攻击的防御方法，且具备相应的功能和有益效果。

[0131] 实施例六

[0132] 本发明实施例六还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现本发明实施例提供的跨站脚本攻击的防御方法，该方法包括：

[0133] 获取网络侧发送的待检测内容；

[0134] 将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码；

[0135] 对潜在攻击代码进行安全编码处理，以得到安全代码；

[0136] 使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容；

[0137] 显示安全内容。

[0138] 本发明实施例的计算机存储介质，可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是--但不限于--电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0139] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于

由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0140] 计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括——但不限于无线、电线、光缆、RF等等，或者上述的任意合适的组合。

[0141] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0142] 注意，上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解，本发明不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此，虽然通过以上实施例对本发明进行了较为详细的说明，但是本发明不仅仅限于以上实施例，在不脱离本发明构思的情况下，还可以包括更多其他等效实施例，而本发明的范围由所附的权利要求范围决定。

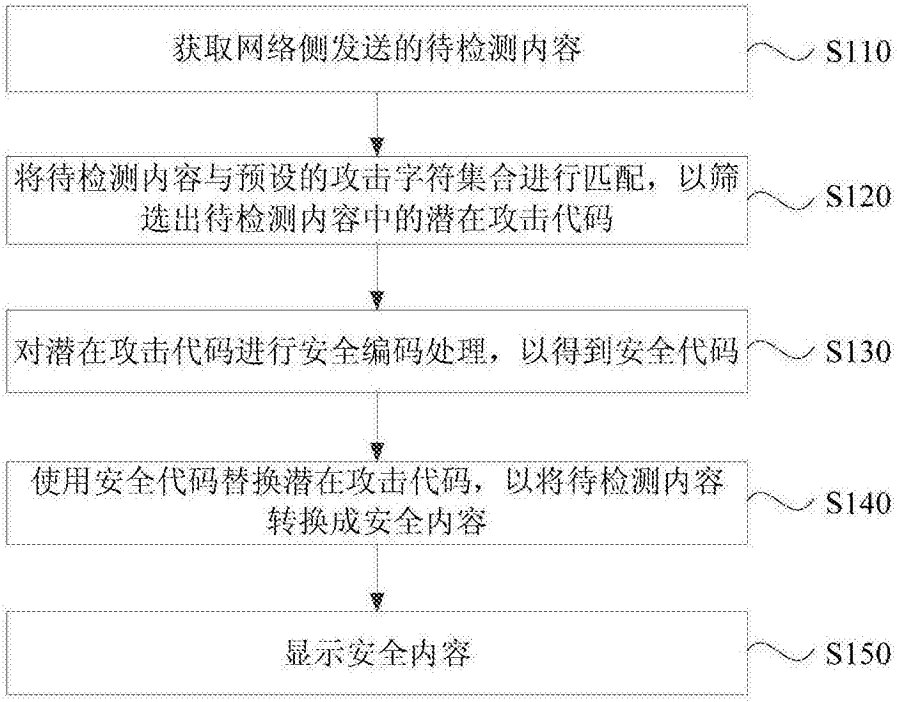

获取网络侧发送的待检测内容
S110
将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码
S120
对潜在攻击代码进行安全编码处理，以得到安全代码
S130
使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容
S140
显示安全内容
S150

图1

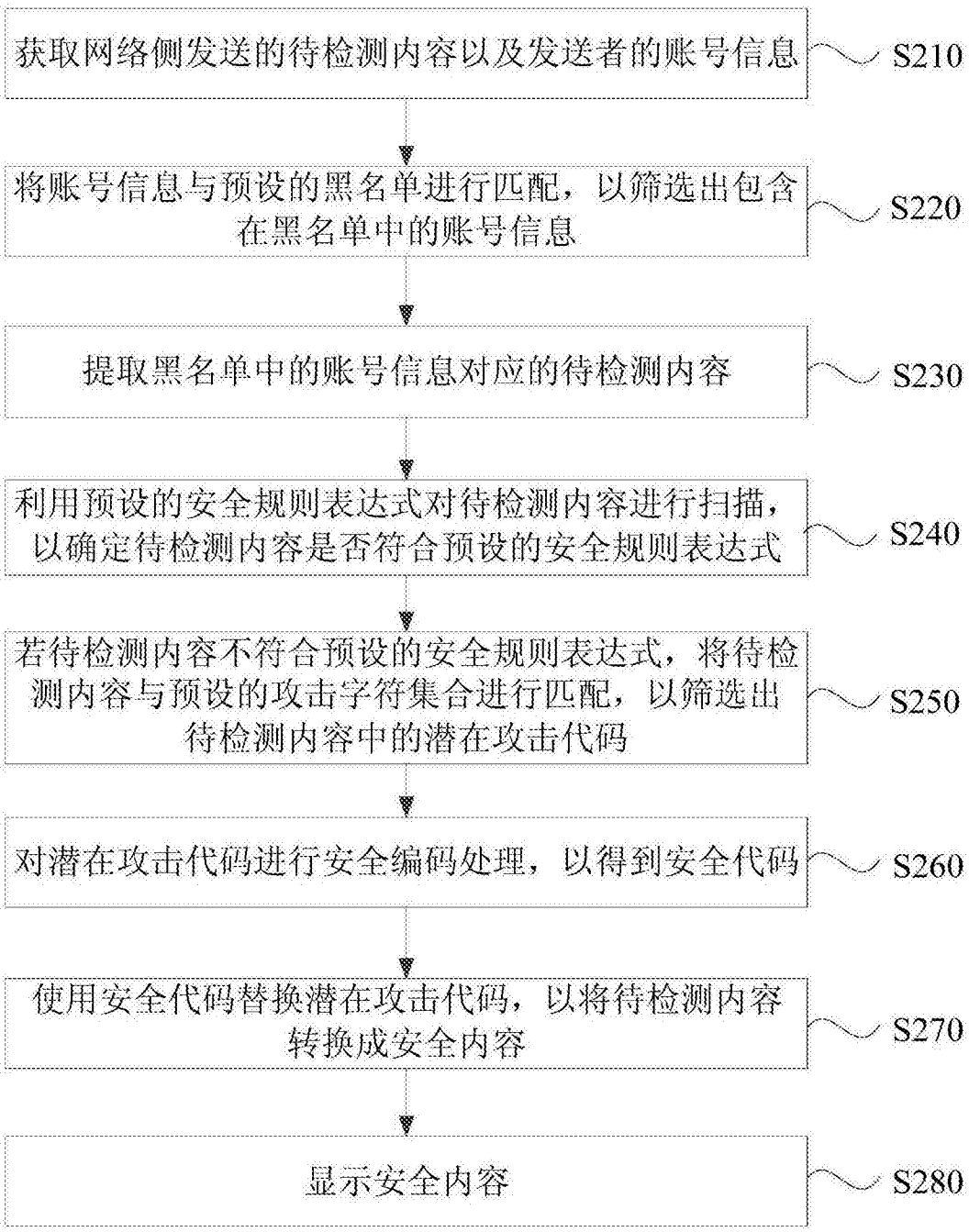

获取网络侧发送的待检测内容以及发送者的账号信息
S210
将账号信息与预设的黑名单进行匹配，以筛选出包含在黑名单中的账号信息
S220
提取黑名单中的账号信息对应的待检测内容
S230
利用预设的安全规则表达式对待检测内容进行扫描，以确定待检测内容是否符合预设的安全规则表达式
S240
若待检测内容不符合预设的安全规则表达式，将待检测内容与预设的攻击字符集合进行匹配，以筛选出待检测内容中的潜在攻击代码
S250
对潜在攻击代码进行安全编码处理，以得到安全代码
S260
使用安全代码替换潜在攻击代码，以将待检测内容转换成安全内容
S270
显示安全内容
S280

图2

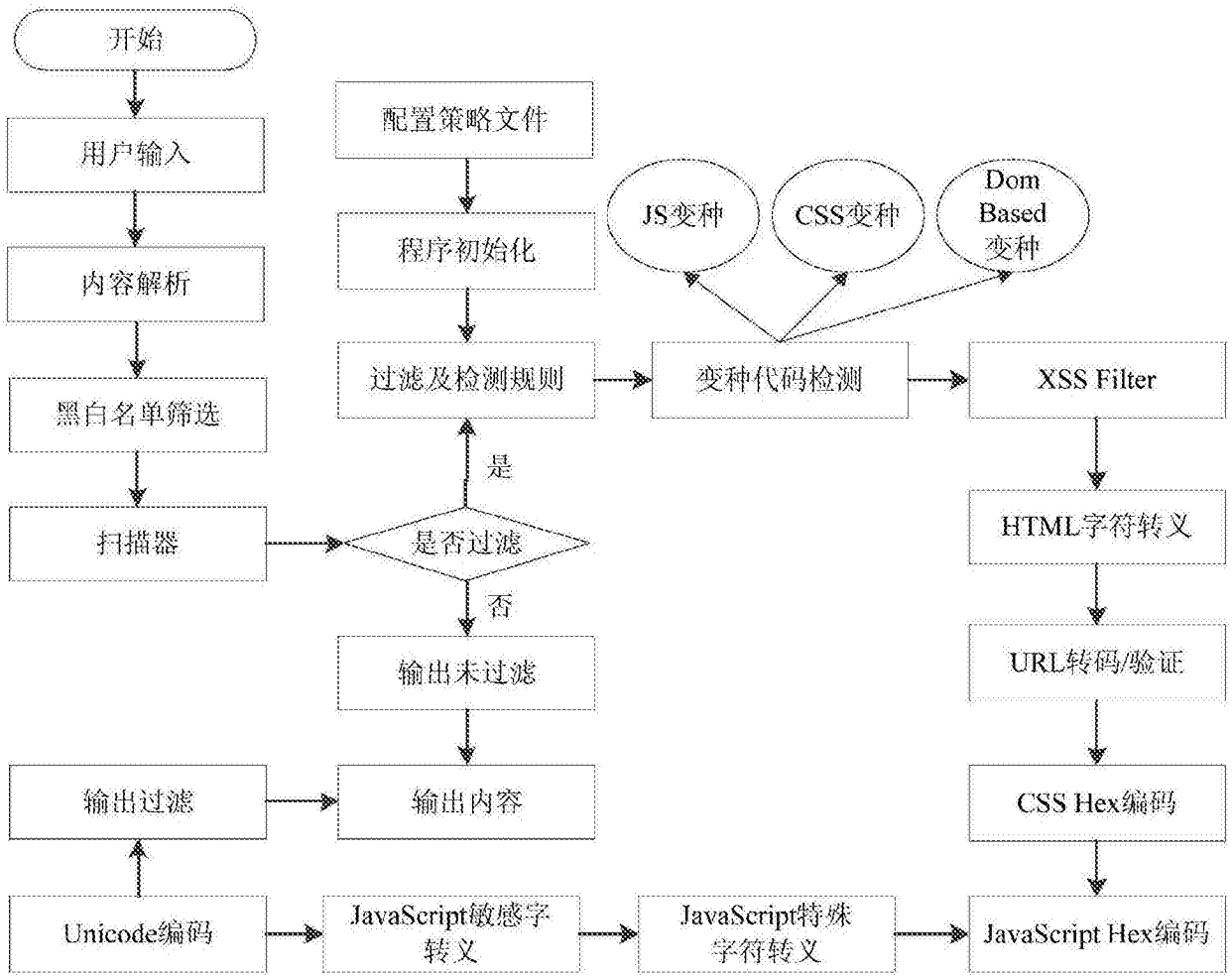

开始
用户输入
内容解析
黑白名单筛选
扫描器
配置策略文件
程序初始化
过滤及检测规则
是否过滤
是
否
输出未过滤
输出过滤
输出内容
Unicode编码
JavaScript敏感字转义
JavaScript特殊字符转义
JS变种
CSS变种
Dom Based变种
变种代码检测
XSS Filter
HTML字符转义
URL转码/验证
CSS Hex编码
JavaScript Hex编码

图3

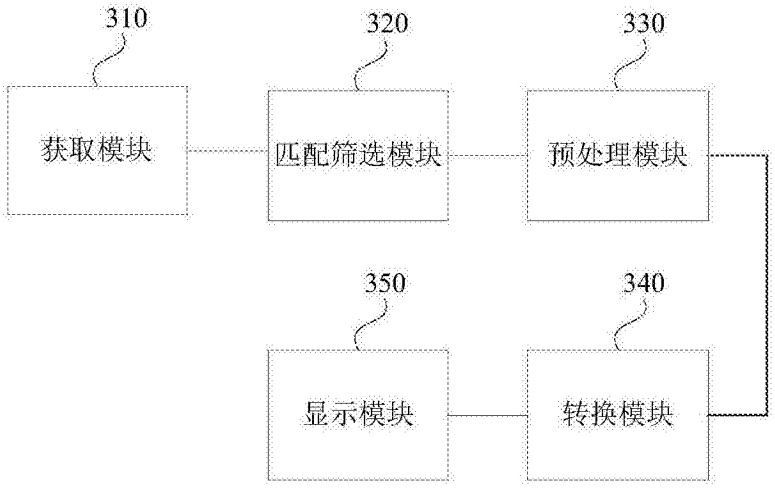

310
320
330
获取模块
匹配筛选模块
预处理模块
350
340
显示模块
转换模块

图4

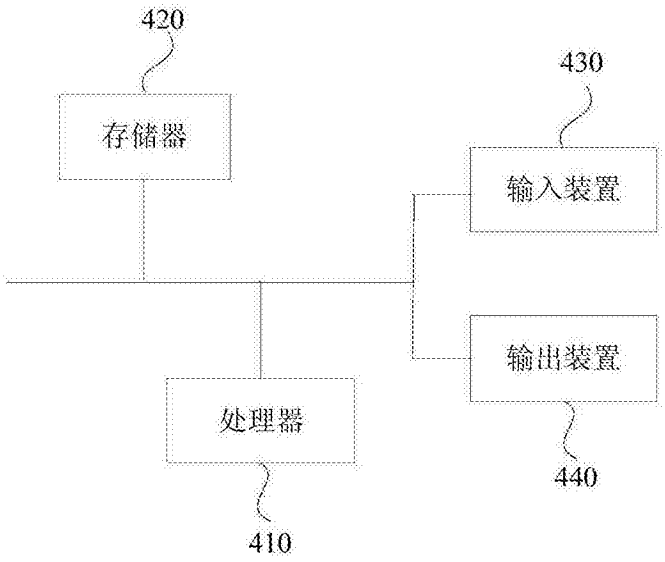


图5