УДК 004.056.53

Р.Х. Хамдамов, К.Ф. Керимов

МАТЕМАТИЧЕСКИЙ АЛГОРИТМ ОБНАРУЖЕНИЯ XSS-АТАК НА ВЕБ-ПРИЛОЖЕНИЯ

Ключевые слова: атаки межсайтового скриптинга (XSS), уязвимость, сигнатуры, информационная безопасность.

Введение

Математическое моделирование и идентификация информационных объектов крайне важны при решении задач распознавания образов. Одной из таких задач является обнаружение атак на веб-приложения. Исследования, посвященные выявлению таких атак, начались сравнительно недавно. Но тем не менее в этом направлении проводится много исследований.

В данной статье предлагается математическое моделирование и способ идентификации атак межсайтового скриптинга (XSS) с помощью ограниченной снизу функции, которая зависит от входной строки. Для построения этой функции применяются специальные знаки и ключевые слова, которые часто встречаются при построении XSS-атак.

С помощью предлагаемого метода можно обнаруживать XSS-атаки, используя специальный символ или ключевое слово. Экспериментально можно показать, что предлагаемый метод обнаружения с использованием набора многочисленных символов и слов позволяет более точно определить угрозу безопасности в виде XSS-атак.

Атаки межсайтового скриптинга

XSS-атаки на сайт проводятся путем введения кода JavaScript, VB Script, ActiveX или HTML. Цель нападающих — сбор индивидуальной информации, изменение настроек пользователя, а также показ ложной рекламы для пользователей. Атака на веб-приложение выполняется с помощью HTTP-запроса, в котором необходимо менять значение полей и данные в Cookie. Задачей данной работы является разработка алгоритма обнаружения XSS-атак. Для ее решения сосредоточим внимание на символах, которые зачастую включаются в строку XSS-атак.

Связанные работы по обнаружению атаки XSS

XSS-атаки на веб-приложения выполняются путем введения вредоносного кода в HTTP-запрос. Разработано множество методов, предотвращающих атаки с точки зрения защиты как на стороне сервера, так и клиента [1]. Тем не менее эти технические приемы не всегда могут защитить от неизвестных атак. Поэтому очень важна разработка новых способов для защиты веб-приложений от неизвестных атак.

В [2] предложен математический метод выявления уязвимостей вида XSS, основанный на удельном весе каждой сигнатуры символов атаки.

В этой статье предлагается алгоритм обнаружения атак на веб-приложения с помощью информации, касающейся частоты и коэффициента важности символов атаки. Предлагаемый алгоритм не требует сложных математических методов, таких как оптимизация и др. Так как метод, основанный на обработке имеющихся видов реальных уязвимостей вида XSS, носит общий характер, предложенный алгоритм можно использовать в качестве онлайн-платформы для обнаружения атак межсайтового скриптинга.

Обнаружение XSS-атак

В данной работе определим, является ли вводимая строка XSS-атакой, используя набор специальных символов. Для этого необходимо моделировать информационный объект, т.е. запрос, состоящий из последовательности специальных символов (табл. 1) и ключевых слов (табл. 2). Эти таблицы составлены путем анализа более 500 реальных угроз информационной безопасности вида XSS [3–8].

Для определения XSS-атаки вводим ее характеристики с помощью специальных символов и ключевых слов. Напомним, что для построения XSS-атак очень часто используются специальные символы и ключевые слова, представленные в табл. 1, 2.

Таблица 1

Переменная	Специальные символы	Коэффициент важности	Переменная	Специальные символы	Коэффициент важности
A_1	“	0,561	A_8	:	0,174
A_2	>	0,331	A_9	.	0,997
A_3	/	0,511	A_{10}	(	0,532
A_4	<	0,331	A_{11}	)	0,532
A_5	Пробел	0,485	A_{12}	–	0,144
A_6	=	0,609	…	…	…
A_7	‘	0,318	A_{32}	$	0,003

Таблица 2

Переменная	Ключевые слова	Коэффициент важности	Переменная	Ключевые слова	Коэффициент важности
A_{33}	FScommand	0,003	A_{40}	Form	0,020
A_{34}	<script>	0,074	A_{41}	xlink:href	0,003
A_{35}	</script>	0,164	A_{42}	seekSegmentTime	0,003
A_{36}	on\w*	0,003	A_{43}	FSCommand	0,003
A_{37}	style	0,097	A_{44}	Applet	0,003
A_{38}	xmlns:xdp	0,003	…	…	…
A_{39}	Formaction	0,010	A_{78}	Svg	0,017

Допустим, наблюдается некоторая входная строка L. Пусть $x_1, x_2, ..., x_{32}$ — частота появления в L специальных символов (табл. 1) и $x_{33}, x_{34}, ..., x_{78}$ — частота появления ключевых слов (табл. 2), x_{79} — частота появления всех остальных знаков и чисел 0, 1, 2, ..., 9 в строке L. С точки зрения определения XSS-атак обычные символы $a, b, ..., z$ и числа 0, 1, …, 9 не имеют особого значения. Поэтому в данной работе будем считать, что частота появления всех этих символов и чисел в наблюдаемой строке L равна единице, т.е. $x_{79} = 1$. Таким образом, любую строку L можно определить с помощью определенных характеристик следующим образом: $L = (x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}, x_{79})$ как элемент некоторого фазового пространства X.

Из L видно, что любой ее элемент из построенного пространства X лежит на гиперплоскости $\Gamma = \{L = (x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}, x_{79}) : x_{79} = 1\}$. Из данного уравнения гиперплоскости можно предположить, что чем больше частота появления специальных символов и ключевых слов во входной строке, тем очевиднее, что входная строка несет угрозу в виде XSS-атаки. Поэтому естественно пред-

положить, что функция определения атаки должна быть возрастающей по переменным $x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}$ и убывающей — по переменной x_{79}. Исходя из этих соображений для определения XSS-атак предлагается следующая функция, которая является возрастающей по переменным $x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}$ и убывающей — по x_{79}.

$$f(L) = f(x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}, x_{79}) = \frac{\sum_{i=1}^{78} x_i}{\sum_{i=1}^{78} x_i + x_{79}}.$$

Так как в данной работе полагается, что частота появления всех остальных знаков и чисел 0, 1, 2, ..., 9 в строке L равна единице, из последнего равенства получим

$$f(L) = f(x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}, x_{79}) = \frac{\sum_{i=1}^{78} x_i}{\sum_{i=1}^{78} x_i + 1}. \quad (1)$$

Таким образом, если входная строка L является XSS-атакой, то она, по крайней мере, должна содержать один специальный символ из табл. 1 или одно ключевое слово из табл. 2. Поэтому $\sum_{i=1}^{78} x_i \geq 1$, и так как функция $f(L)$ является возрастающей по каждой из переменных x_i, ее минимум при $\sum_{i=1}^{78} x_i \geq 1$ достигается в точке L_0, для которой $\sum_{i=1}^{78} x_i = 1$.

Таким образом, если L — произвольная строка и $f(L) \geq 1/2$, то L, возможно, является XSS-атакой. В этом случае для ее построения используются, как минимум, либо один специальный символ и одно ключевое слово, либо одно ключевое слово. Если же $f(L) < 1/2$, то тогда входная строка, возможно, является нормальной. Поэтому функцию (1) можно использовать для распознавания XSS-атак и нормальных строк, построенных с помощью специальных символов и ключевых слов.

Далее уточним алгоритм, используя при построении функции распознавания коэффициенты важности специальных символов. Для этого построим функцию

$$f_1(L) = f_1(x_1, x_2, ..., x_{32}, x_{33}, ..., x_{78}, x_{79}) = \frac{\sum_{i=1}^{78} k_i x_i}{\sum_{i=1}^{78} k_i x_i + 1}, \quad (2)$$

где $k_i, 0 < k_i < 1,$ — коэффициенты важности специальных символов из табл. 1. Коэффициенты важности вычислены посредством исследования 299 реальных XSS-атак. Используя значения коэффициентов важности и вид функции (2), легко определить минимум новой функции $f_1(L)$ при условии $\sum_{i=1}^{78} x_i \geq 0{,}003$. Минимум этой функции достигается в точке L_0, для координат которой выполняется равенство $\sum_{i=1}^{78} x_i = 0{,}003$.

Таким образом, из новой функции $f_1(L)$ имеем, что если L — произвольная строка и $f_1(L) \geq 0{,}003$, то L, возможно, является XSS-атакой, и в этом случае для ее построения используются, как минимум, либо один специальный символ и одно ключевое слово, либо одно ключевое слово. Если $f(L) < 0{,}003$, то входная строка является нормальной. Поэтому функцию (2) можно применять для распознавания XSS-атак и нормальных строк, построенных с помощью специальных символов и ключевых слов.

Заключение

В данной работе предложена модель информационного объекта, на которой построен алгоритм обнаружения XSS-атак на веб-приложения посредством выявления частоты появления специальных символов при создании входящих запросов и определения их коэффициентов важности. При этом следует отметить, что коэффициенты важности специальных символов (табл. 1) более существенны, в отличие от коэффициентов важности ключевых слов (табл. 2), которые близки к нулю. Поэтому если при построении функций (1), (2) не учесть некоторые параметры с нулевыми коэффициентами, то распознавание входящих данных, возможно, будет происходить с некоторой малой погрешностью.

Использование предложенного метода не требует сложных математических расчетов. Посредством статистической обработки реальных входящих данных вычислены коэффициенты важности всех специальных символов и ключевых слов, задействованных при построении XSS-атак. Применяя полученный алгоритм распознавания для модели, учитывающей коэффициенты важности специальных символов, получим новый алгоритм распознавания XSS-атак. Таким образом, данный метод можно использовать в качестве онлайн-инструмента для обнаружения XSS-атак.

Р.Х. Хамдамов, К.Ф. Керімов

МАТЕМАТИЧНИЙ АЛГОРИТМ ВИЯВЛЕННЯ XSS-АТАК НА ВЕБ-ДОДАТКИ

Останнім часом атаки на веб-додатки, такі як SQL-ін'єкції та міжсайтовий скриптинг (XSS), мають тенденцію до збільшення. У статті запропоновано новий алгоритм виявлення XSS-атак, побудований на аналізі частоти появи спеціальних символів, а також математичне моделювання та спосіб ідентифікації XSS-атак за допомогою обмеженої знизу функції, яка залежить від вхідного рядка. Для побудови цієї функції використано спеціальні символи та ключові слова, які часто зустрічаються у побудові XSS-атак. Математичне моделювання та ідентифікація інформаційних об'єктів важливі при вирішенні задач розпізнавання образів. Однією з таких задач є виявлення атак на веб-додатки. Дослідження з виявлення та вивчення атак на веб-додатки розпочалися порівняно не так давно. Але тим не менше в цьому напрямку існує багато досліджень. У запропонованому методі можна виявляти XSS-атаки, використовуючи один спеціальний символ або одне ключове слово. Проте експериментально можна показати, що даний метод виявлення з використанням набору спеціальних символів і ключових слів дозволяє більш точно визначити загрозу безпеці у вигляді XSS-атак. Метою даної роботи є розробка алгоритму виявлення XSS-атак за допомогою символів, які часто використовують у побудові вхідного рядка XSS-атак.

Ключові слова: атаки міжсайтового скриптингу (XSS), вразливість, сигнатури, інформаційна безпека.

R.Kh. Khamdamov, K.F. Kerimov

MATHEMATICAL ALGORITHM FOR DETECTING XSS ATTACKS ON WEB APPLICATIONS

Recently, attacks on web applications, such as SQL injection and cross-site scripting (XSS), have tended to increase. In this article, we proposed a new algorithm for detecting XSS attacks on a web application based on the analysis of the frequency of occurrence of special characters. The paper proposes mathematical modeling and a method for identifying XSS attacks using a function bounded below that depends on the input string. To build this function, special characters and keywords were used, which are often found in the construction of XSS attacks. Mathematical modeling and identification of information objects plays an important role in solving the problems of pattern recognition. One such task is to detect attacks or normal requests to web applications. Research devoted to the study of the detection of attacks or normal requests to web applications began relatively recently. Nevertheless, there is a lot of research in this direction. In this paper, we propose mathematical modeling and a method for identifying XSS attacks using a function bounded below that depends on the input string. To build this feature, we used special characters and keywords that are often found in building XSS attacks. In the proposed method, it is possible to detect XSS attacks using one special character or one keyword. Nevertheless, it can be experimentally shown that the proposed detection method using a set of numerous characters and words allows us to determine more accurately the vulnerability of the type of XSS attacks. The aim of this work is to develop an algorithm for detecting XSS attacks. To achieve this, we focused on the characters that are often included in the XSS attack string.

Keywords: cross-site scripting (XSS), vulnerability, signatures, information security.

1. Khamdamov R.Kh., Kerimov K.F., Ibrahimov J.O. Method of developing a web-application firewall. *Journal of Automation and Information Sciences*. 2019. **51**, N6. P. 65–74. DOI: 10.1615/JAutomatInfScien.v51.i6.60 New York, USA 6,2019.
2. Sonoda M., Matsuda T., Koizumi D., Hirasawa S. On automatic detection of SQL injection attacks by the feature extraction of the single character. *Proceeding in 2011 International Conference on Security of Information and Networks*, *ACM.* 2011. P. 81–86.
3. Щеглов А.Ю. Защита компьютерной информации от несанкционированного доступа. М. : Наука и техника, 2002. 46 с.
4. Ржавский К.В. Информационная безопасность: практическая защита информационных технологий и телекоммуникационных систем: учебное пособие. Волгоград : ВолГУ, 2002. 50 с.
5. Низамутдинов М.К. Тактика защиты и нападения на ИТ-приложения. Санкт-Петербург : БХВ-Петербург, 2005. С. 30–60.
6. Домарев В.В. Безопасность информационных технологий. Методология создания систем защиты. М. : ДиаСофт, 2002. 56 с.
7. Cross-site-scripting (XSS). Tutorial [online]. http://www.veracode.com/security/xss.
8. Opanasenko V.N., Kryvyi S.L. Synthesis of adaptive logical networks on the basis of Zhegalkin polynomials. *Cybernetics and Systems Analysis*. 2015. **51**, N 6. P. 969–977. DOI: 10.1007/s10559-015-9790-1.

Получено 11.06.2020
После доработки 27.08.2020