УДК 004.738.52

Д.Л. Жусов

МЕТОДИКА ФИЛЬТРАЦИИ ПОТОКА ЗАПРОСОВ К WEB-СЕРВЕРУ

На основе разработанной модели предложена методика фильтрации потока запросов к web-серверу. Обоснованно применены как детерминированный, так и интеллектуальный подходы к фильтрации. Произведена оценка эффективности предлагаемых решений.

Введение. В настоящее время для повышения эффективности механизмов государственного управления на основе создания общей информационно-технологической инфраструктуры в деятельность органов государственной власти (ОГВ) активно внедряются современные информационные технологии [1, 2]. В системе информационного обеспечения органов государственной власти (ОГВ) важное место занимают web-серверы, предназначенные для предоставления информации и обеспечения межведомственного взаимодействия и взаимодействия ОГВ с населением и организациями. При этом на официальных web-сайтах информация открыта для пользователей сети Интернет и не содержит конфиденциальных сведений.

Сегодня web-серверы функционируют на основе динамического метода формирования web-страниц (ДМФС), что на практике приводит к множеству реализаций специально организованных компьютерных атак, направленных на подмену информации на web-страницах (атакам подмены контента) [3]. В связи с этим актуальной является задача обеспечения безопасного функционирования web-серверов в условиях компьютерных атак подмены контента и практической реализации соответствующих технических решений.

Анализ источников [4, 5] показал, что для реализации процедуры обнаружения компьютерных атак наиболее часто используется сигнатурный метод. Однако его эффективность напрямую связана с обязательным поддержанием в актуальном состоянии базы данных сигнатур атак путем ее постоянного обновления. В то же время процесс фильтрации запросов к web-серверам с ДМФС существенно усложняется, в связи со сложностью обнаружения сигнатурным методом модифицированных нарушителем комбинаций значений параметров доступа, передаваемых серверу для обработки запроса. Это может привести к изменению содержимого страниц (подмене контента). Успеху реализации компьютерной атаки в этом случае способствует невозможность полного описания вариантов компьютерной атаки сигнатурным методом.

В связи с этим для повышения защищенности web-серверов с динамически формируемыми страницами от компьютерных атак подмены контента необходима методика фильтрации потока запросов, учитывающая возможные изменения в структуре запросов.

Цель работы – разработать методику фильтрации потока запросов к web-серверу, учитывающую особенности его функционирования в условиях компьютерных атак подмены контента, провести оценку эффективности решений.

Постановка задачи. Таким образом, задачу можно сформулировать следующим образом: для web-сервера, имеющего в своем составе различные динамически формируемые информационные ресурсы $web = \{resource_i\}$, необходимо обеспечить защиту от компьютерных атак подмены контента информационных страниц посредством фильтрации потока запросов $z(\overrightarrow{Per})$, характеризующихся различными значениями переменных доступа:

$$\Phi : z(\overrightarrow{Per}) - z(\overrightarrow{Per}_{\text{безоп}}) = \Delta z . \qquad (1)$$

При этом в зависимости от уровня допустимого риска к обработке запросов функционирование системы фильтрации должно обеспечивать требуемый уровень:

$$\begin{cases} \Phi \to \max : z(\overrightarrow{Per}) \in \Psi_0, \Delta z(\overrightarrow{Per}) = 0 \\ \Phi = \Phi^{\text{тр}} : z(\overrightarrow{Per}) \notin \Psi_0, \Delta z(\overrightarrow{Per}) = \\ = \varepsilon : \begin{cases} \mu_{\text{сх}}(z, z_{\text{КА}}) < \mu_{\text{сх}}(z, z_{\text{безоп}}) \\ \mu_{\text{сх}}(z, z_{\text{безоп}}) \ge H. \end{cases} \end{cases} \qquad (2)$$

Обозначения:

ε – степень опасности запросов к web-серверу;

$\mu_{\text{сх}}$ – степень сходства.

Структура запроса к web-серверу, определенная в [6, 7], представлена выражением 3:

$$z =< Method, Id, Per, zgl > \quad (3)$$

где $Method$ – метод доступа к информационному ресурсу;

Id – идентификатор информационного web-ресурса;

$\overrightarrow{Per} = \{x_1\alpha_1, \ldots x_j\alpha_j\}$ – переменные доступа к web-ресурсу и их значения;

zgl – заголовок HTTP-протокола.

При постановке задачи приняты следующие ограничения и допущения:

– компьютерные атаки проводятся нарушителями посредством url-запросов к web-серверу;

– конфиденциальная информация на сервере отсутствует;

– пользователи не ограничены в формировании значений переменных доступа запросов.

Предпосылки проведения исследования. Результат анализа существующих подходов к фильтрации запросов к web-серверу показал, что возможен подход, заключающийся в формировании множества безопасных значений параметров доступа запросов и последующей идентификации их модификаций в запросах, поступающих на web-сервер для обработки. Предложенная в [8, 9] модель обнаружения компьютерных атак на web-сервер отличается использованием заранее заданных безопасных запросов. Однако формирование безопасных значений параметров осуществляется администратором безопасности сервера, что является элементом субъективизма в процессе обработки клиентских запросов. Данная модель позволяет разработать алгоритм идентификации в потоке данных следующих типов HTTP-запросов, приводящих к успешной реализации компьютерной атаки:

– синтаксис которых не соответствует требованиям RFC 2616 и RFC 1945;

– сформированных на основе методов, которые не поддерживаются прикладным ПО web-сервера;

– направленных к несуществующим информационным ресурсам web-сервера;

– содержащих параметры, количество символов в которых превышает заданные ограничения;

– сформированных на основе версии протокола HTTP, которая не поддерживается прикладным ПО web-сервера;

– содержащих запрещенные типы заголовков.

В вышеуказанной модели не учитывается тот факт, что запросы, имеющие одинаковое число переменных доступа и количество символов в значениях переменных доступа, на практике могут быть принципиально различными. Например, запрос http://www.mail.ru?par1= =http://www.yandex.ru/ index.php по этим признакам идентичен запросу http://www.mail.ru?par1= =http://www.attack.ru/atack.php, в то же время результаты их выполнения будут различными. Для реализации компьютерной атаки нарушитель, анализируя web-ресурс http://www.mail.ru, находит разрешенную администратором ссылку на сайт http://www.yandex.ru/index.php, после чего загружает на сервер http://www.attack.ru специально созданный скрипт attack.php и формирует URL-запрос к серверу http://www.mail.ru с модифицированными значениями переменной доступа par1. В результате выполнения данного запроса реализуется компьютерная атака типа php-including.

В связи с этим для повышения защищенности web-серверов с ДМФС необходима модель фильтрации потока запросов, учитывающая описанные недостатки.

Результаты исследования

Функциональная модель подсистемы фильтрации потока запросов к web-серверу. Для предотвращения поступления на обработку web-серверу небезопасных значений была разработана функциональная модель подсистемы фильтрации потока запросов к web-серверу (рисунок 1).

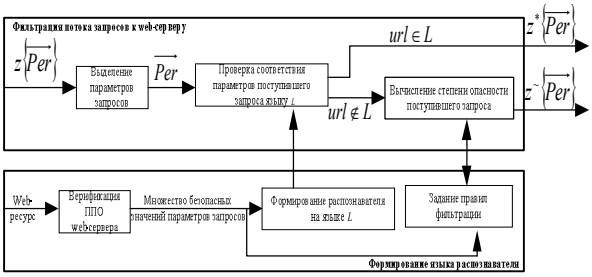

Рисунок 1 – Функциональная модель подсистемы фильтрации потока запросов к web-серверу

Значения параметров доступа каждого поступившего запроса проверяются на соответствие безопасным значениям. В случае их корректности запрос обрабатывается сервером. В противном случае осуществляется анализ запроса на степень его близости к сигнатурам компьютерных атак и безопасным запросам. По результатам анализа принимается решение об отнесении его к одному из классов.

Математическая модель фильтрации потока запросов к web-серверу. Для решения задачи проверки соответствия параметров поступившего запроса на предварительном этапе формируется язык автомата-распознавателя на

основе процедуры верификации ППО [10], суть которой заключается в построении множества безопасных значений параметров доступа запросов.

Для решения обработки данных верификации ППО по различным методикам используем математический аппарат теории распознавания образов [11]. С его помощью оценим похожесть выделенных безопасных запросов с набором контролируемых параметров доступа

$\tilde{S} z_{r_{i-1}+v} = (\alpha_1,...,\alpha_q)$, на запрос, полученный в ходе верификации $\tilde{S} z' = (\beta_1,...,\beta_q)$. Данные запросы считаются похожими, если выполняется не менее чем δ неравенств $|\alpha_j - \beta_j| \le \varepsilon_j, j = 1,...,q$. На основании анализа оценок запроса

$$\begin{cases} \Gamma(z', \Omega_1) = \sum\limits_{S_A} \Gamma(z', \Omega_1) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \Gamma(z', \Omega_m) = \sum\limits_{S_A} \Gamma(z', \Omega_m) \end{cases} \quad (4)$$

принимается решение либо об отнесении запроса z' к одному из классов Ω_i, $i = 1,...,m$, либо об отказе от его распознавания, что соответствует следующему выражению:

$$\Gamma(z', \Omega_i) > \Gamma(z', \Omega_m). \quad (5)$$

Сформированный на множестве безопасных запросов конечный автомат позволяет осуществлять верификацию выделенных из поступающего запроса параметров доступа и их значений на предмет соответствия языку автомата:

$$Avt = < S, Z, Y, s_0, \delta, \alpha\alpha, S_{fyn} >. \quad (6)$$

Выбор конечного автомата как средства распознавания варианта компьютерной атаки обусловлен его свойством находиться в одном из возможных состояний в зависимости от поступающего к нему на обработку запроса. Граф состояний и переходов процесса проверки соответствия параметров поступившего запроса языку безопасных запросов L представлен на рисунке 2.

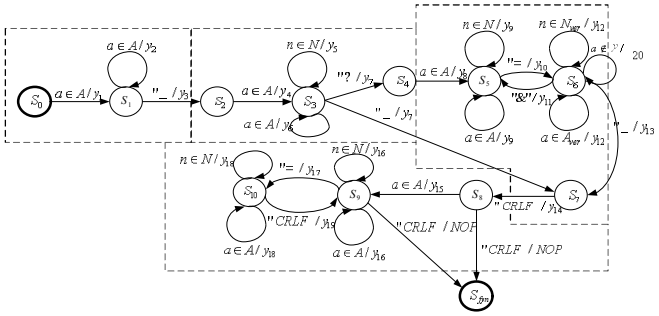

Рисунок 2 – Граф состояний и переходов процесса проверки соответствия параметров поступившего запроса языку безопасных запросов L

Результат работы автомата-распознавателя сводится к достижению им одного из двух конечных состояний. В случае анализа запроса, состоящего из безопасных значений параметров доступа, автомат переходит в состояние S_{fyn}, в противном случае принимается решение об анализе автоматом небезопасного запроса.

$$T(Avt) = \{X \mid \delta(S, Z) \text{ находится в } S_{fyn}\}. \quad (7)$$

В то же время, учитывая, что пользователи не ограничены в формировании значений переменных доступа запросов, и с целью сохранения функциональности web-ресурса необходимо разработать процедуру оценки степени опасности запроса, распознанного конечным автоматом как небезопасный.

С этой целью введем соответствующие процедуры, основанные на математическом аппарате теории нечетких множеств [12].

Пусть задано исходное множество безопасных запросов к определенному ресурсу web-сервера $Z = \{z_i, i = \overline{1,m}\}$ и множество значений его переменных доступа $K = \{k_i, i = \overline{1,n}\}$, по которым должны оцениваться запросы. При этом рассматриваются однотипные запросы, когда каждому запросу $z \in Z$ присущи все переменные доступа из рассматриваемого множества $k \in K$. Для каждого из значений переменных доступа $k \in K$ известно нечеткое отношение предпочтения Ψ_k на множестве корректных запросов Z, т.е. известна функция принадлежности $\mu_\Psi(z_i, z_j, k)$, значение которой понимается как степень предпочтительности запроса z_i запросу z_j по переменной доступа k.

Элементы множества K различны по важности, т.е. на множестве значений переменных доступа заданы отношения, характеризующиеся величинами $\lambda(k_1, k_2)$, понимаемые как степень, с которой значение k_1 считается не менее важным значения k_2.

Задача заключается в том, что по имеющейся информации о значениях переменных доступа поступившего запроса осуществить оценку степени его опасности на основе исходного множества безопасных запросов Z.

Решение поставленной задачи предлагается искать, опираясь на понятие нечеткого множества недоминируемых альтернатив, функция принадлежности которого определяется следующим выражением [13]:

$$\mu_{\Psi}^{нд}(z) = \inf_{z_i \in Z}\left[1 - \max\left\{0, \mu_{\Psi}(z_i, z) - \mu_{\Psi}(z, z_i)\right\}\right] = \\ = 1 - \sup_{z_i \in Z} \max\left\{0, \mu_{\Psi}(z_i, z) - \mu_{\Psi}(z, z_i)\right\}, \quad (8)$$

где $\mu_{\Psi}(z, z_i)$ - функция принадлежности нечеткого отношения предпочтения (н.о.п) на множестве Z.

Значение $\mu_{\Psi}^{нд}(z)$ представляет собой степень, с которой запрос z не доминируется ни одним из запросов множества Z. Другими словами, это степень, с которой запрос z можно считать безопасным с точки зрения близости значений его переменных доступа к запросу из безопасного множества.

Для нахождения множества недоминируемых запросов по множеству значений переменных доступа K рассмотрим нечеткое множество (Z, μ_{Θ_1}), где μ_{Θ_1} - функция принадлежности пересечения Θ_1 исходных отношений Ψ_k. Подмножество недоминируемых запросов во множестве (Z, μ_{Θ_1}) совпадает с множеством недоминируемых запросов для исходного набора функций $\mu_{\Psi}(z_i, z_j, k)$ [13]. При этом пересечению отношений Ψ_k, заданных на множестве значений переменных доступа K, соответствует функция принадлежности μ_{Θ_1} [13]:

$$\mu_{\Theta_1}(z, z_i) = \min\left\{\mu_{\Psi}(z, z_i, k_1), \mu_{\Psi}(z, z_i, k_n)\right\}. \quad (9)$$

Путем подстановки (9) в (8), получаем выражение для функции принадлежности нечеткого подмножества недоминируемых запросов на множестве (Z, μ_{Θ_1}):

$$\mu_{\Theta_1}^{iд}(z) = 1 - \sup_{z_i \in Z} \max\left\{0, \mu_{\Theta_1}(z_i, z) - \mu_{\Theta_1}(z, z_i)\right\}. \quad (10)$$

Однако выражение (9) не учитывает различия переменных доступа по степени их важности, задаваемых величинами $\lambda(k_1, k_2)$. Поэтому необходимо ввести свертку отношений вида [13]:

$$\mu_{\Theta_2}(z, z_i) = \sum_{j=1}^{n} \delta_{\lambda j} \cdot \mu_{\Psi}(z, z_i, k_j), \sum_{j=1}^{n} \delta_{\lambda j}, \quad (11)$$

$$\delta_{\lambda j} \geq 0, \quad j = \overline{1, n},$$

где коэффициенты $\delta_{\lambda j}$ имеют смысл коэффициентов важности и могут быть определены на основе заданного множества отношений предпочтения по формуле [12]:

$$\delta_{\lambda j} = \lim_{s \to \infty} \delta_{\lambda j}^{(s)}, \quad (12)$$

$$\delta_{\lambda j}^{(s)} = \frac{\sum_{i=1}^{n} \lambda^{(s)}(k_j, k_i)}{\sum_{j=1}^{n}\sum_{i=1}^{n} \lambda^{(s)}(k_j, k_i)},$$

где $\delta_{\lambda j}^{(s)}$ – относительная сила s-го порядка признака k_j; $\lambda^{(s)}(k_j, k_i)$ – ji-й элемент матрицы λ, возведенный в степень s.

В соответствии с (8) функция принадлежности нечеткого подмножества недоминируемых запросов на множестве (Z, μ_{Θ_2}) определяется выражением:

$$\mu_{\Theta_2}^{нд}(z) = 1 - \sup_{z_i \in Z} \max\left\{0, \mu_{\Theta_2}(z_i, z) - \mu_{\Theta_2}(z, z_i)\right\}. \quad (13)$$

Результирующее множество недоминируемых запросов находим как пересечение множеств $\mu_{\Theta_1}^{нд}$ и $\mu_{\Theta_2}^{нд}$ [13], функция принадлежности которого $\mu^{нд}(z)$ определяется выражением:

$$\mu^{нд}(z) = \min\left\{\mu_{\Theta_1}^{нд}(z), \mu_{\Theta_2}^{нд}(z)\right\}. \quad (14)$$

Для принятия введем соответствующие критерии, если допускается выбор максимально безопасного запроса:

$$Z_i^{нд} = \left\{z \mid z \in Z, \mu_i^{нд}(z) = \sup_{z_i \in Z} \mu^{нд}(z_i)\right\} \quad (15)$$

и если допускается выбор запроса со степенью безопасности не ниже H:

$$Z_i^{нд} = \left\{z \mid z \in Z, \mu_i^{нд}(z) > H\right\}. \quad (16)$$

Методика фильтрации потока запросов к web-серверу. Результаты моделирования были положены в основу методики фильтрации потока запросов к web-серверу (рисунок 3).

Перед началом функционирования системы фильтрации предварительно выполняются этапы 1-5 методики. На этих этапах задаются требования к обработке запросов, осуществляется верификация прикладного ПО сервера различными процедурами, по итогам которой формируется безопасное множество запросов и диапазоны разрешенных размеров значений переменных доступа. На их основе формируется автомат-распознаватель структуры запросов и задается соответствие расстояния Левенштайна и степени опасности запроса.

Рисунок 3 – Методика фильтрации потока запросов к web-серверу

На этапе функционирования сервера осуществляется фильтрация потока запросов к нему (этап 6), основу которых составляет соответствующий алгоритм, позволяющий повысить защищенность web-сервера от компьютерных атак подмены контента за счет:

– проверки корректности структуры запросов пользователей;

– соотнесения поступающих запросов с множеством безопасных запросов и множеством компьютерных атак.

По итогам его работы принимается решение об обработке запроса сервером. С целью оценки качества функционирования подсистемы фильтрации на этапе 8 осуществляется контроль целостности информационных ресурсов.

Экспериментальные исследования показали, что при применении разработанных решений вероятность ложной тревоги не превышает 5 %, а пропуска компьютерной атаки, содержащейся в базе данных, не превышает 2 %.

Таким образом, в результате исследований на основе анализа существующих методов обнаружения компьютерных атак была разработана методика фильтрации потока запросов к web-серверу, позволяющая обеспечить защиту от компьютерных атак подмены контента информационных web-ресурсов; проведена оценка эффективности разработанных решений.

Заключение. Решение задачи обнаружения компьютерных атак подмены контента на web-серверы с динамически формируемыми страницами позволяет повысить защищенность их функционирования. В расчетах, основанных на применении математического аппарата теорий распознавания образов и нечетких множеств, учитывается выполнение требований как по функциональности предоставления информации пользователям, так и по ее защищенности.

Научная новизна работы заключается в представлении процедуры фильтрации во взаимосвязи с требованиями по ее уровню, а также с синтаксисом языка и структурой HTTP-запросов; в инвариантности предлагаемой модели к программно-аппаратному обеспечению web-сервера.

Практическая значимость определяется возможностью внедрения разработанной методики в состав действующих программно-аппаратных комплексов.

Библиографический список

1. *Никитов В.А.* и др. Информационное обеспечение государственного управления. – М.: Славянский диалог, 2000. – 415 с.

2. Концепция использования информационных технологий в деятельности федеральных органов государственной власти до 2010 года// распоряжение Правительства РФ от 27 сентября 2004 г. № 1244-р

3. Постановление Правительства РФ № 98: "Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления" от 12.02.2003 г.

4. *Лукацкий А.В.* Обнаружение атак. – СПб.: БХВ-Петербург, 2001. – 624 с.

5. *Spafford H. Eugene, Kumar Sandep* A pattern matching model for misuse intrusion detection. COAST Project, Purdue University, 1994

6. *Postel J.* Request for Comments 1591 – Domain Name System Structure and Delegation. Network Working Group, 1994

7. *Fielding R., Gettys J.* Request for Comments 2616 – HyperText Transfer Protocol – HTTP/1.1. Network Working Group, 1999

8. *Сердюк В.А.* Математическая модель поведенческого метода обнаружения информационных атак // Тезисы. Международная молодежная научная конференция "XXIX Гагаринские чтения". – М.: МАТИ. Т.5. – С.5–6.

9. *Сердюк В.А.* Математическая модель поведенческого метода обнаружения сетевых атак, базирующаяся на конечных автоматах-распознавателях // Моделирование. Теория, методы и средства: Материалы IV Междунар. науч.-практ.конф., г. Новочеркасск, 9

апр. 2004 г.: В 4 ч. Юж.-Рос. гос. техн. ун-т (НПИ). Новочеркасск: ЮРГТУ, 2004. Ч. 4. – С. 8–13.

10. *Бухарин В.В.* Применение верификации для специального программного обеспечения в автоматизированных системах управления. Системы связи. Анализ. Синтез. Управление. Вып. № 15/ Под ред. В. П. Постюшкова.– СПб.: Тема–2005.С.9–14

11. *Журавлев Ю.И.* Распознавание. Классификация. Прогноз. Математические методы и их применение. Вып. 2. – М. Наука, 1989 – 72 с.

12. *Замыслов М.А., Михайленко С.Б., Замыслов Е.М., Замыслов А.М.* Применение методов теории нечетких множеств и отношений для решения задачи выбора потребителем наиболее предпочтительных поставщиков по совокупности показателей //Информационные технологии, № 5, 2002, с.28-33

13. *Орловский С.А.* Проблемы принятия решений при нечеткой исходной информации. – М. Наука, 1981. – 208 с.