

УДК 004.738.52

Д.Л. Жусов

МОДЕЛЬ И АЛГОРИТМ ОБСЛУЖИВАНИЯ ПОЛЬЗОВАТЕЛЕЙ КОРПОРАТИВНОГО WEB-СЕРВЕРА

Предложены модель и алгоритм обслуживания пользователей корпоративного web-сервера. Обоснованы применение приоритетных правил обработки клиентских запросов и введение процедуры учета самоподобия входящего потока запросов. Произведена оценка эффективности предлагаемых решений.

Введение. В настоящее время в интересах министерств и ведомств РФ, различных организаций и предприятий функционирует большое количество корпоративных серверов службы гипертекста (web-серверов), обеспечивающих надежное и защищенное информационное взаимодействие. Их задачи отражают важнейшие направления практических аспектов реализации процесса его функционирования: обеспечение информационной безопасности и регламентированное представление информационных услуг организациям и населению с выполнением требований по их качеству.

Для корпоративных web-серверов характерно выполнение представительских функций (размещение информации об организации), а также предоставление возможности обсуждения, ответов на интересующие вопросы и т.п. при различных уровнях привилегий пользователя на обслуживание (приоритетах) и в условиях высокой загрузки сервера. В то же время главной задачей web-сервера как информационной системы является обеспечение своевременности представления требуемой информации и безопасности представляемой информации [1]. Сложность выполнения требований по своевременности предоставления информации обусловлена наличием подсистемы защиты web-сервера, необходимостью учета временных затрат на их реализацию, групп привилегированных пользователей, обладающих преимущественным правом по доступу к информации, и постоянно изменяющимся характером входящей нагрузки [2].

В связи с этим с целью адекватного представления процесса обслуживания пользователей корпоративного web-сервера в рамках исследований необходимо разработать модель, учитывающую данные особенности, и алгоритм, позволяющий повысить эффективность функционирования web-сервера.

Цель работы – разработать модель и алгоритм обслуживания пользователей корпоративного web-сервера, учитывающие особенности функционирования сервера и входящего сетевого трафика, провести оценку эффективности решений.

Постановка задачи. На содержательном уровне задачу на разработку модели можно сформулировать следующим образом: для наблюдаемого потока клиентских запросов к web-серверу, исходя из требований по качеству обслуживания пользователей, рассчитать размер очереди запросов K с учетом того, что для защиты сервера используются заранее определенные процедуры, такие что на их выполнение требуется время $T_{безоп}$, таким образом, чтобы общее время нахождения запроса в системе сервера

$$v = w_r + \mu^{-1} + T_{безоп} \quad (1)$$

не превышало требуемого значения v_{mp} .

На формальном уровне задача примет следующий вид:

При заданных λ_r , T_r , v_{mp} , $T_{безоп}$, P_{mp} необходимо рассчитать размер очереди запросов обслуживания K такой, чтобы:

$$\begin{cases} P(w_r \leq T_r) \geq P_{mp} \\ P_{отк} \leq P_{отк}^{mp} \\ v \leq v_{mp} \end{cases} \quad (2)$$

При постановке задачи использованы следующие обозначения:

T_r – требуемое время ожидания обслуживания запроса сервером;

v – общее время пребывания запроса на сервере;

v_{mp} – требуемое общее время пребывания запроса на сервере;

w_r – время ожидания обслуживания запроса сервером;

$T_{безоп}$ – время, затрачиваемое на выполнение процедур защиты;

μ^{-1} – среднее время обслуживания одного запроса;

λ_r – интенсивность поступления запросов по каждой из приоритетных групп пользователей.

Результаты исследования

Функциональная модель процесса обслуживания пользователей корпоративного web-сервера. Исследования [3, 4] показали, что web-сервер можно представить в виде системы массового обслуживания, на вход которой поступает поток запросов. При этом основным показателем для расчета эффективности функционирования системы является ν – время пребывания запроса на сервере:

$$\nu = w_r + \mu^{-1}. \quad (3)$$

В то же время для защищенного web-сервера для расчета времени нахождения запроса в системе необходимо учесть время, затрачиваемое на выполнение процедур защиты, что описывается выражением (1). При этом расчет $T_{безоп}$ осуществляется на предварительном этапе путем выполнения функций защиты.

В работах [5, 6] показано, что телетрафик в сети Интернет не является стационарным и обладает свойством самоподобия. В связи с этим постоянные значения такого параметра web-сервера, как размер очереди запросов, на всей продолжительности его функционирования не позволяют получить требуемые значения показателей качества обслуживания. Это обуславливает необходимость постоянного расчета значений параметров web-сервера с учетом требований по качеству обслуживания, на основании постоянно изменяющейся прогнозной оценки значений нагрузки web-сервера. Учитывая свойство самоподобия трафика, для реализации процедуры прогнозирования нагрузки web-сервера предполагается использовать математический аппарат теории временных рядов.

В соответствии с положениями теории случайных процессов в [7, 8] доказано, что даже при невыполнении условия стационарности входящего потока запросов, при расчетах его можно считать простейшим, а при применении аппарата временных рядов для прогнозирования нагрузки учитывается как нестационарность потока запросов, так и наличие последствия в потоке запросов.

Исходя из вышесказанного расчет вероятностно-временных характеристик (ВВХ) процесса обслуживания запросов пользователей может быть произведен на основе математического аппарата теории массового обслуживания с учетом того, что для обработки поступает поток запросов λ_{ex} от $r=1..R$ приоритетных классов пользователей. При этом интенсивность обслуживания запросов пользователей с учетом выполнения процедур защиты может быть описана выражением:

$$\mu_{общ} = \mu_{безоп} + \mu_{обсл}, \quad (4)$$

где $\mu_{безоп}$ – интенсивность выполнения операций по обеспечению безопасного доступа; $\mu_{обсл}$ – интенсивность обслуживания запросов пользователей.

Вышесказанное позволяет представить функциональную модель процесса обслуживания пользователей защищенного корпоративного web-сервера следующим образом (рисунок 1).

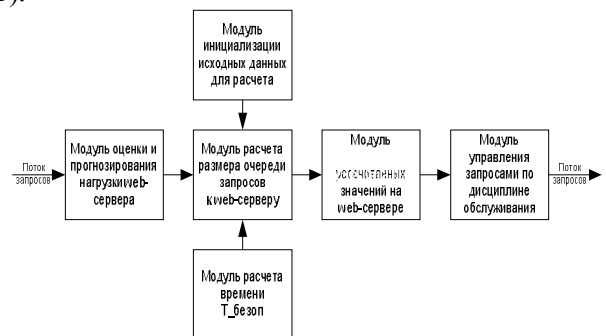


Рисунок 1 – Функциональная модель процесса обслуживания пользователей корпоративного web-сервера

На предварительном этапе в рамках инициализации исходных данных для проведения расчетов путем проведения нагрузочного тестирования определяются параметры web-сервера (количество потоков для обслуживания запросов пользователей, размер буфера) [9]. Путем тестирования средств защиты, установленных на web-сервере, вычисляется значение времени анализа поступающих запросов $T_{безоп}$ на предмет наличия в них признаков компьютерных атак.

В процессе функционирования web-сервера на основе прогнозных оценок величины нагрузки, поступающей на обработку, и значений, рассчитанных на предварительном этапе, производится расчет размера очереди запросов к web-серверу с учетом предъявляемых требований по качеству обслуживания разноприоритетных пользователей. При расчетах учитывается факт применения динамического метода формирования web-страниц, что усложняет процесс оценки значений размера очереди запросов ввиду апри-

орной неопределенности о времени обработки запроса. Для устранения данного противоречия в качестве времени обработки запросов принято значение, рекомендованное международными организациями [10]. Затем рассчитанное значение размера очереди запросов устанавливается в конфигурационном файле web-сервера. Дальнейшее управление потоком запросов к web-серверу осуществляется на основе установленных значений и в приоритетном порядке в соответствии с дисциплиной обслуживания.

Аналитическая модель процесса обслуживания пользователей корпоративного web-сервера. Анализ процесса функционирования корпоративных web-серверов показал, что для них характерны следующие особенности:

- в общем случае нагрузка от высокоприоритетных пользователей мала;
- основную нагрузку web-сервера создают запросы пользователей низких приоритетов;
- необходимо минимизировать отказы в обслуживании запросам всех пользователей при обеспечении качества обслуживания каждой из приоритетных групп пользователей.

Данные особенности влияют на выбор дисциплины обслуживания запросов. Анализ существующих вариантов показал, что наиболее пригодной является дисциплина с относительными приоритетами, которая наиболее просто обеспечивает требования к качеству обслуживания низкоприоритетных пользователей при одновременном выполнении требований к обслуживанию высокоприоритетных пользователей. В то же время в процессе функционирования web-сервера возможны ситуации, когда нагрузка от высокоприоритетных пользователей существенно возрастает, при этом применение дисциплины обслуживания с относительными приоритетами не позволяет обеспечить требуемое значение показателя "вероятность своевременной обработки запроса". В этом случае необходимо применять дисциплину обслуживания с относительными приоритетами и резервированием ресурсов [11].

Критерием необходимости управления параметром web-сервера является выражение:

$$P(w_r \leq T_r) < P_{mp}, \quad (5)$$

где P_{mp} – требуемое значение вероятности $P(w_r \leq T_r)$, устанавливаемое в соответствии с регламентирующими документами и спецификой работы сервера.

При выполнении данного условия расчет размера очереди запросов (размера буфера) web-сервера осуществляется для прогнозной оценки входящей нагрузки на основе анализа ВВХ web-

сервера, полученных по следующим выражениям [11].

Вероятность того, что запрос r -го приоритета, поступивший в произвольный момент времени на обработку web-серверу, попадет на ожидание и среднее время его ожидания не превысит допустимого значения:

$$P(w_r \leq T_r) = 1 - A \cdot \sum_{i=0}^{K_r-1} \left(\frac{\rho_r}{V_r} \right)^i \sum_{j=0}^i \frac{(\overline{\mu} V_r T_r)^j}{j!}, \quad (6)$$

$$\text{где } A = \frac{P_{V_r}(r, V_r) \exp(-\overline{\mu} V_r T_r)}{1 - P_{V_r+K_r}(r, V_r)};$$

V_r – число процессов, которые могут быть одновременно запущены на сервере с целью обработки запросов пользователей соответствующего приоритета;

K_r – размер очереди запросов соответствующего приоритета;

T_r – требуемое значение времени ожидания обслуживания для запросов пользователей соответствующего приоритетного класса;

ρ_r – суммарная интенсивность поступающей нагрузки сервера запросами $\overline{1, r}$ приоритетов;

$P_i(r, V_r)$ – вероятность того, что на web-сервере в произвольный момент времени находится i запросов соответствующих приоритетов.

Вероятность отказа в обслуживании запросов r -го приоритета:

$$P_{отк}(r, V_r) = P_{V_r+K_r}(r, V_r) + B, \quad (7)$$

где $B = P_{выт}(r, V_r)[1 - P_{V_r+K_r}(r, V_r)]$;

$$P_{выт}(r, V_r) = \begin{cases} 0, & r = 1 \\ \frac{1}{\rho_r^*} \cdot \sum_{m=1}^{r-1} C_m, & r = \overline{2, R} \end{cases} \quad \text{– вероятность,}$$

характеризующая отношение среднего числа вытесненных из очереди запросов r -го приоритета к среднему числу поступивших запросов r -го приоритета (при этом вытеснение запросов r -го приоритета определяется запросами с приоритетом $\overline{1, r-1}$);

$$C_m = \rho_m^* \cdot [P_{V_r+K_r}(r, V_r) - P_{V_r+K_r}(m, V_r)];$$

$$\rho_r^* = \frac{X_r \cdot \lambda_r}{\mu_r} \quad \text{– интенсивность поступающей}$$

нагрузки r -го приоритета.

Качество расчета значения размера очереди запросов к web-серверу определяется выполнением условий (2). При этом среднее время пребывания запросов в системе web-сервера складывается из среднего времени ожидания обслуживания w_r , среднего времени

обслуживания μ^{-1} и времени, затрачиваемого на обеспечение безопасного доступа к ресурсу $T_{безоп}$ (1), или иначе - с учетом введенных выше обозначений:

$$v_r = w_r + \bar{\mu}_{общ}^{-1}. \quad (8)$$

Проведенный эксперимент показал, что соответствующие аналитические выражения адекватно описывают процесс обслуживания разноприоритетных запросов к web-серверу.

Эффективность функционирования системы оценивается путем расчета обобщенного показателя эффективности, который наиболее полно описывается векторным критерием [12, 13]: $P[v_r \leq v_{r_{mp}}]$ – вероятностью своевременной доставки принятых к обслуживанию запросов приоритетов $r = \overline{1, R}$. При этом обобщенный показатель эффективности процесса функционирования web-сервера представляет собой аддитивную свертку векторного критерия:

$$P[v \leq v_{mp}] = \sum_{r=1}^R \frac{\lambda_r \cdot X_r}{\Lambda_R} \cdot P[v_r \leq v_{r_{mp}}] \quad (9)$$

и определяет степень выполнения требований по качеству обслуживания.

Разработанная модель обслуживания пользователей корпоративного web-сервера, учитывающая, в отличие от известных, требования по своевременности обработки клиентских запросов и затраты на реализацию процедур обеспечения безопасного доступа к ресурсам web-сервера, позволила разработать предложения по структуре корпоративного web-сервера (рисунок 2).



Рисунок 2 – Структурная схема корпоративного web-сервера

В зависимости от задач, стоящих перед конкретным сервером, каждый модуль может иметь различное построение. Однако в общем случае расчет параметров web-сервера в модуле управления должен производиться в соответствии с описанными выше выражениями (1)-(9).

Алгоритм обслуживания пользователей корпоративного web-сервера. На основе разработанной модели предлагается следующий алгоритм обслуживания пользователей корпоративного web-сервера (рисунок 3).

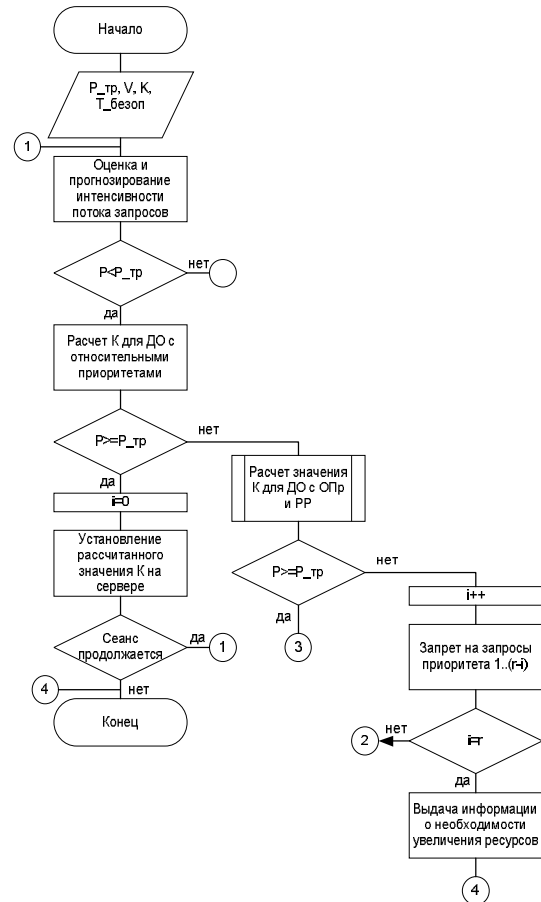


Рисунок 3 – Алгоритм обслуживания пользователей корпоративного web-сервера

Перед началом работы алгоритма в качестве исходных данных задаются значения $P_{тр}$, K , V , $T_{безоп}$. При этом значение $P_{тр}$ определяется на основе нормативных документов либо субъективно администратором сервера, а остальные исходные данные инициализируются путем проведения предварительного нагрузочного тестирования программно-аппаратной части web-сервера [9]. На этапе функционирования web-сервера на основе модели [14] осуществляется процедура оценки и прогнозирования интенсивности входящего потока запросов по каждой из приоритетных групп пользователей. Далее для полученных прогнозных значений рассчитывается значение показателя эффективности функционирования. Если рассчитанное значение показателя эффективности меньше требуемого, выполняется процедура расчета значения размера очереди запросов к web-серверу для дисциплины обслуживания (ДО) с относительными приоритетами. В случае повышенной нагрузки к web-серверу при невыполнении требований по своевременности обработки запросов для данной дисциплины обслуживания происходит пересчет значения размера очереди запросов для дисцип-

лины обслуживания с относительными приоритетами и резервированием ресурсов (ОПр и РР). В алгоритме предусмотрен случай, когда ни одна из дисциплин обслуживания не может удовлетворить требованиям по качеству обслуживания пользователей. В этом случае необходимо ограничить поток запросов от низкоприоритетных пользователей, который снова получает доступ при снижении общей нагрузки на сервер.

Проведенный эксперимент показал, что путем применения разработанных решений эффективность функционирования web-сервера повышается не менее чем на 5 % по показателю "своевременность обработки запросов" для высокоприоритетных пользователей.

Таким образом, в результате исследований разработана модель, описывающая процесс обслуживания разноприоритетных пользователей web-сервера, которая позволила разработать структуру web-сервера с учетом требований по своевременности предоставления информации и безопасности функционирования сервера, а также алгоритм обслуживания пользователей корпоративного web-сервера; проведена оценка эффективности предложенных решений.

Заключение. Решение задачи обеспечения качества обслуживания разноприоритетных пользователей корпоративного web-сервера позволяет более эффективно использовать ресурсы web-сервера для обеспечения пользователей информацией. В расчетах, основанных на применении математического аппарата теории массового обслуживания, учитываются прогнозные оценки входящей нагрузки web-сервера, что позволяет управлять размером очереди запросов в целях выполнения заданных требований по своевременности обработки запросов.

Научная новизна предлагаемых решений заключается в комплексном подходе к рассмотрению процесса обслуживания пользователей корпоративного web-сервера, учитывающем временные затраты на выполнение процедур по защите web-сервера от компьютерных атак, свойства телекоммуникационного трафика к web-серверу и особенности обслуживания разноприоритетных групп пользователей.

Практическая значимость работы определяется совокупностью полученных результатов, являющихся решением актуальной научной задачи, направленной на повышение эффективности функционирования корпоративных web-серверов и возможности практического внедрения алгоритма обслуживания пользователей корпоративного web-сервера в состав действующих программно-аппаратных комплексов.

Библиографический список

1. ГОСТ РВ 51987–2002. Типовые требования и показатели качества функционирования информационных систем. Госстандарт России. М. – 2002
2. ITU-T. Terms and definitions related to quality of service and network performance including dependability. Recommendation E.800. – Geneva, 1994
3. Д. Менаске, В. Алмейда. Производительность web-служб. Анализ, оценка и планирование: Пер. с англ. – СПб: ООО "ДиаСофтЮП", 2003. – 480 с.
4. Л. Клейнрок. Вычислительные системы с очередями. Пер. с англ. Под ред. Б. С. Цыбакова. – М.: Мир, 1997. – 600 с.
5. S.S. Lam A carrier sense multiple access protocol for local networks // Computer Networks, vol. 4, n. 1, pp. 21-32
6. Городецкий А.Я., Заборовский В.С. Информатика. Фрактальные процессы в компьютерных сетях: Учеб. пособие. СПб.: изд-во СПбГТУ, 2000. – 102 с.
7. Венцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения: Учеб. пособие для вузов. – 2-е изд., стер. – М.: Высш. шк., 2000. – 480 с.: ил.
8. Боровков А.А. Теория вероятностей. – М.: Наука, 1986. – 432 с.
9. Васинев Д.А., Жусов Д.Л., Тарасенко Д.Н. Способ нагрузочного тестирования службы гипертекста // Перспективы развития средств связи в силовых структурах, обеспечение информационной безопасности в системах связи. Материалы 3-й межвузовской научно-практической конференции в 2-х частях (21 марта 2007 года). Ч.2. / Под общ. ред. старшего преподавателя кафедры № 4 ГПИ ФСБ России, подполковника Зубкова В. В. – Голицино: ГПИ ФСБ России, 2007. – 132 с., С. 59-61.
10. Битнер В.И., Попов Г.Н. Нормирование качества телекоммуникационных услуг: Учеб. пособие / Под ред. профессора В.П. Шувалова. – М.: Горячая линия – Телеком, 2004. – 312 с., ил.
11. Михалевич И.Ф., Сычев К.И. Модели систем массового обслуживания в практических задачах анализа систем мобильной связи: Монография. – Орел: Академия ФАПСИ, 2003. – 211с.
12. Терентьев В.М., Санин Ю.В. Анализ эффективности функционирования автоматизированных систем многоканальной радиосвязи. – Л.: ВАС, 1992. – 80 с.
13. Щекотихин В.М., Терновой И.Л. Оптимизация ресурсов телекоммуникационных систем / Под ред. В.М. Терентьева – Орел: ВИПС, 1999. – 137 с.
14. Борисенко Н.П., Васинев Д.А., Жусов Д.Л. Модель контроля и управления производительностью web-сервера // Высокие технологии, фундаментальные и прикладные исследования, образование. Т.7: Сборник Второй международной научно-практической конференции "Исследование, разработка и применение высоких технологий в промышленности" / Под ред. А.П. Кудинова, Г.Г. Матвиенко, В.Ф. Самохина. СПб.: Изд-во Политехн. ун-та, 2006, 607 с., С. 69-71.