

УДК 004.75

ОЦЕНКА ВЕРОЯТНОСТИ ЭРЛАНГОВСКОГО СТАРЕНИЯ ИНФОРМАЦИИ

Б. Я. Советов,

доктор техн. наук, профессор

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

М. О. Колбанёв,

доктор техн. наук, профессор

Санкт-Петербургский государственный экономический университет

Т. М. Татарникова,

доктор техн. наук, доцент

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Рассматривается процесс предоставления информационной услуги, который регламентируется соглашением о качестве обслуживания. Предлагается процедура вычисления вероятности своевременной обработки запроса пользователя как одной из основных характеристик качества обслуживания. Приводится методика оценки вероятности своевременного предоставления информационных услуг центрами обработки данных. Показано, что искомая вероятность может быть вычислена, если преобразование Лапласа — Стильтьеса функции распределения времени обработки запроса пользователя представляет собой произведение или частное двух функций, является суперпозицией двух функций или задано неявно.

Ключевые слова — качество обслуживания клиента, информационная услуга, показатели качества предоставления услуг, центр обработки данных, вероятность своевременного предоставления информационной услуги.

Введение

В последние годы государственными органами и бизнесом все шире используются центры обработки данных (ЦОД), обеспечивающие надежное хранение данных, круглосуточную поддержку бизнес-процессов, качественное предоставление других информационных услуг [1, 2].

Процесс предоставления услуг регламентируется соглашениями о качестве обслуживания (Service Level Agreement — SLA). SLA — это часть договора между провайдером и клиентом, в котором приводится система показателей для контроля качества предоставляемых услуг, а также штрафы за нарушение соглашения и вознаграждения в случае успешного его выполнения.

Вероятность эрланговского старения информации как характеристика качества обслуживания

SLA не регламентирует содержательную сторону процесса информационного взаимо-

действия, а относится только к процессу предоставления информационной услуги. Качество этого процесса оценивается вероятностью своевременной обработки запроса пользователя, под которой понимается вероятность того, что ответ на запрос будет получен за допустимое время [3]. При моделировании сложных информационных систем и сетей, подобных ЦОД, эту характеристику обычно называют вероятностью старения информации или вероятностью своевременного предоставления информационной услуги ($P_{\text{усл}}$). В общем случае $P_{\text{усл}}$ определяется выражением

$$P_{\text{усл}} = \int_0^{\infty} \varphi_{\text{доп}}(x) dT(x), \quad (1)$$

где $\varphi_{\text{доп}}(x)$ — функция распределения (ФР) допустимого времени реализации моделируемого процесса предоставления информационной услуги, устанавливаемая SLA; $T(x)$ — ФР времени реализации этого процесса.

При нормировании вероятности $P_{\text{усл}}$, как правило, выбирается постоянное ограничение на время предоставления услуги $\varphi_{\text{доп}}(x)$. Приме-

ром такой нормы может служить рекомендация Q.543ITU-T, в которой один из показателей качества обслуживания вводится как 95%-квантили распределений времени своевременной обработки запросов пользователей. В ряде работ получены результаты для так называемого экспоненциального старения, когда функция $\varphi_{\text{доп}}(x)$ имеет показательное распределение.

Наиболее общим случаем является использование в качестве ограничения обобщенного распределения Эрланга, т. е. функции вида

$$\varphi(x) = 1 - \sum_{r=1}^n \eta_r e^{-k_r \gamma_r x} \sum_{i=0}^{k_r-1} \frac{(k_r \gamma_r x)^i}{i!}, \quad (2)$$

$$\sum_{r=1}^{n_j} \eta_r = 1, \quad \frac{1}{\gamma} = \sum_{r=1}^n \eta_r / \gamma_r,$$

где $1/\gamma$ — среднее допустимое время предоставления услуги. Эту характеристику можно назвать вероятностью эрланговского старения информации ЦОД. Путем соответствующего выбора значений параметров n , η_r , k_r и γ_r функцией (2) можно достаточно точно аппроксимировать любую реальную ФР допустимого времени предоставления услуги. Так, например, при $n=1$ и $k_r \rightarrow \infty$ функция (2) стремится к постоянному ограничению, а при $n=1$ и $k_r=1$ превращается в экспоненциальное распределение.

Подставляя (2) в (1), после ряда преобразований получаем

$$P_{\text{усл}} = 1 - \sum_{r=1}^n \eta_r \sum_{i=0}^{k_r-1} (-1)^i \frac{s_0^i}{i!} t^{(i)}(s_0), \quad (3)$$

где $s_0 = k_r \gamma_r$; $t^{(i)}(s_0)$ — значение i -й производной преобразования Лапласа — Стильтьеса (ПЛС) функции распределения времени предоставления услуги в точке s_0 .

Процедура вычисления вероятности $P_{\text{усл}}$

Рассмотрим лежащую в основе расчета вероятностей $P_{\text{усл}}$ по формуле (3) процедуру вычисления производных от ПЛС ФР $t(s)$ в точке s_0 , предполагая, что функция $t(s)$ может представлять собой произведение или частное двух функций, может быть сложной функцией (суперпозицией двух функций) или может быть задана неявно.

Пусть $x(s)$ и $y(s)$ n раз дифференцируемы в точке $s=s_0$. Тогда для n -й производной $Z_{s_0}^{(n)}$ функции $Z(s)=x(s)y(s)$ в точке s_0 справедлива формула Лейбница

$$Z_{s_0}^{(n)} = \sum_{i=0}^n c_n^i x_{s_0}^{(n-i)} y_{s_0}^{(i)}, \quad (4)$$

где $x_{s_0}^{(i)}$ и $y_{s_0}^{(i)}$ — значения производных функций $x(s)$ и $y(s)$ в точке s_0 и предполагается, что

$$x_{s_0}^{(0)} = x(s_0), \quad y_{s_0}^{(0)} = y(s_0) \quad \text{и} \quad Z_{s_0}^{(0)} = Z(s_0);$$

$$c_n^i = \frac{n!}{i!(n-i)!} \text{ — биномиальный коэффициент.}$$

Если функция $Z(s)=x(s)/y(s)$, то n -я производная $Z_{s_0}^{(n)}$ может быть вычислена рекуррентно. Умножим левую и правую части последнего уравнения на $y(s)$ и возьмем n -ю производную от левой и правой частей полученного равенства. Тогда, применяя (4), получим

$$x_{s_0}^{(n)} = \sum_{i=0}^n c_n^i y_{s_0}^{(n-i)} Z_{s_0}^{(i)}, \quad (5)$$

откуда следует рекуррентное соотношение

$$Z_{s_0}^{(n)} = \frac{x_{s_0}^{(n)} - \sum_{i=0}^{n-1} c_n^i y_{s_0}^{(n-i)} Z_{s_0}^{(i)}}{y(s_0)}, \quad (6)$$

которое позволяет вычислить $Z_{s_0}^{(n)}$ через производные $Z(s)$ в точке s_0 более низких порядков.

Если функция $Z(s)=y[x(s)]$ и известны значения $y_{x_0}^{(i)} (i=\overline{1, n})$ производных по x функции $y(x)$ в точке $x_0=x(s_0)$, то $Z_{s_0}^{(n)}$ может быть вычислена следующим образом:

$$Z_{s_0}^{(1)} = y_x^{(1)} x_s^{(1)}. \quad (7)$$

Переходя к производным порядка n и применяя (4), получаем

$$Z_{s_0}^{(n)} = \sum_{i=0}^{n-1} c_{n-1}^i \left[y_x^{(1)} \right]_{s_0}^{(n-1-i)} x_{s_0}^{(i+1)}, \quad (8)$$

где $\left[y_x^{(1)} \right]_{s_0}^{(i)}$ — i -я производная по s функции $dy(x)/dx$ в точке s_0 .

Формула (8) позволяет вычислить $Z_{s_0}^{(n)}$, если известны производные $x_{s_0}^{(1)}, \dots, x_{s_0}^{(n)}$ и $y_{x_0}^{(1)}, \left[y_x^{(1)} \right]_{s_0}^{(1)}, \dots, \left[y_x^{(1)} \right]_{s_0}^{(n-1)}$, причем последние также зависят от производных функции $x(s)$ порядка не выше $(n-1)$ и могут быть вычислены рекуррентно:

$$\begin{aligned} \left[y_x^{(m)} \right]_{s_0}^{(1)} &= y_x^{(m+1)} x_{s_0}^{(1)}, \quad m = \overline{1, n-1}; \\ &\dots\dots\dots \\ \left[y_x^{(m)} \right]_{s_0}^{(k)} &= \sum_{v=0}^{k-1} c_{k-1}^v \left[y_x^{(m+1)} \right]_{s_0}^{(k-1-v)} x_{s_0}^{(v+1)}, \quad m = \overline{1, n-k}; \\ &\dots\dots\dots \\ \left[y_x^{(1)} \right]_{s_0}^{(n-1)} &= \sum_{v=0}^{n-2} c_{n-2}^v \left[y_x^{(n)} \right]_{s_0}^{(n-2-v)} x_{s_0}^{(v+1)}, \quad (9) \end{aligned}$$

где $\left[y_x^{(m)} \right]_{s_0}^{(k)}$ — k -я производная по s функции $d^m y(x)/dx^m$ в точке s_0 .

Пусть теперь $x(s) = y[x(s)] + s$. Тогда после дифференцирования левой и правой частей и приведения подобных для первой производной $x_{s_0}^{(1)}$ получаем

$$x_{s_0}^{(1)} = 1 / \left(1 - y_x^{(1)} \right), \quad (10)$$

а для производных более высокого порядка непосредственно из (8) после замены $Z(s)$ на $x(s)$ получаем следующую рекуррентную формулу:

$$x_{s_0}^{(n)} = \frac{\sum_{i=0}^{n-2} c_{n-1}^i \left[y_x^{(1)} \right]_{s_0}^{(n-1-i)} x_{s_0}^{(i+1)}}{1 - y_x^{(1)}}, \quad (11)$$

где производные $\left[y_x^{(1)} \right]_{s_0}^{(k)}$ ($k = \overline{1, n-1}$) могут быть рассчитаны по формуле (9).

Таким образом, для того чтобы вычислить n -ю производную ($n > 1$) функции $x(s) = y[x(s)] + s$, необходимо вычислить:

$x_{s_0}^{(1)}$ по формуле (10);

$\left[y_x^{(1)} \right]_{s_0}^{(1)}, \dots, \left[y_x^{(n-1)} \right]_{s_0}^{(1)}$ по формуле (9) при

$k = 1$ и $m = 1, 2, \dots, n-1$;

$x_{s_0}^{(2)}$ по формуле (11);

$\left[y_x^{(1)} \right]_{s_0}^{(2)}, \dots, \left[y_x^{(n-2)} \right]_{s_0}^{(2)}$ по формуле (9) при

$k = 1$ и $m = 1, 2, \dots, n-2$;

$x_{s_0}^{(3)}$ по формуле (11)

и так далее, пока не будут получены значения

$$\left[y_x^{(1)} \right]_{s_0}^{(n-1)} \text{ и } x_{s_0}^{(n)}.$$

Приложение вычисления вероятности эрланговского старения информации для ЦОД

Проиллюстрируем изложенную процедуру вычисления вероятности эрланговского старения на примере ЦОД, который содержит серверы, обменивающиеся блоками данных взаимодействия (БДВ). Предположим, что обмен осуществляется через среду множественного доступа, а при обработке данных серверами используется дисциплина обслуживания с относительными приоритетами. Структура современного ЦОД подробно рассматривается в работе [4].

В этом случае для ПЛС ФР времени $t_h(s)$ процесса предоставления услуги в ответ на БДВ с относительным приоритетом h имеем

$$t_h(s) = \omega_h^{\text{обп}}(s) \beta_h(s) \omega^{\text{дост}}(s), \quad (12)$$

где $\omega_h^{\text{обп}}(s)$ и $\beta_h(s)$ — соответственно ПЛС ФР времени ожидания начала обработки и времени обработки в сервере БДВ с приоритетом h ; $\omega^{\text{дост}}(s)$ — ПЛС ФР времени доставки БДВ к серверу через среду множественного доступа.

Для $\omega_h^{\text{обп}}(s)$ справедливы выражения

$$\omega_h^{\text{обп}}(s) = \frac{(1-\rho)G_h(s) + \sum_{j:h_i>h} \lambda_j [1-\beta_j(G_h(s))]}{s - \sum_{j:h_i=h} \lambda_j [1-\beta_j(G_h(s))]}, \quad (13)$$

$$G_h(s) = s + \sum_{j:h_i<h} \lambda_j [1-\beta_j(G_h(s))], \quad (14)$$

где ρ — загрузка сервера; λ_j — интенсивность поступления БДВ, имеющих приоритет j .

Преобразование $\omega^{\text{дост}}(s)$ определяется протоколами доступа управляющих элементов к среде передачи и управления логическим каналом и параметрами физического уровня и среды передачи [5]:

$$\omega^{\text{дост}}(s) = \frac{s(1-\rho_y)g(e^{sT})}{s - \lambda_y + \lambda_y g(e^{sT})}, \quad (15)$$

где ρ_y и λ_y — соответственно загрузка и интенсивность потока БДВ на среду множественного доступа; T — временной параметр среды множественного доступа; $g(e^{sT})$ — ПЛС ФР интервала обслуживания БДВ для выбранной совокупности указанных протоколов.

Методика вычисления вероятности своевременного предоставления информационной услуги

В соответствии с разработанным методом вычисление вероятности $P_{\text{усл}}$ осуществляется в следующем порядке.

1. Найти значения производных $\beta_j^{(i)}(s_0)$ от ПЛС $\beta_j(s)$ в точке s_0 ($i=1, k_r-1$). Если, например, время обработки БДУ с приоритетом j распределено по показательному закону со средним значением \bar{t}_j , то

$$\beta_j(s) = 1 / (1 + \bar{t}_j s) \quad (16)$$

и искомые производные

$$\beta_j^{(i)}(s) = (-1)^i i! \frac{(\bar{t}_j)^i}{(1 + \bar{t}_j s_0)^{i+1}} \quad (17)$$

2. Найти значения производных $\beta_j^{(i)}(G_0)$ функций $\beta_j(G_h(s))$ по $G_h(s)$ в точке $G_0 = G_h(s_0)$. Решая (14) методом последовательных приближений при $s=s_0$, находим G_0 , а затем — искомые значения $\beta_j^{(i)}(G_0)$ по формуле (17) путем замены s_0 на G_0 .

3. Найти значения производных $\omega_h^{\text{обp}(i)}(s_0)$ от преобразования $\omega_h^{\text{обp}(i)}(s)$ в точке s_0 . Для этого обозначим числитель и знаменатель формулы (13) $\omega_{h,1}^{\text{обp}}(s)$ и $\omega_{h,2}^{\text{обp}}(s)$ соответственно.

Очевидно, что производные $\omega_{h,1}^{\text{обp}}(s_0)$ и $\omega_{h,2}^{\text{обp}}(s_0)$ равны

$$\omega_{h,1}^{\text{обp}(i)}(s_0) = (1 - \rho) G_h^{(i)}(s_0) - \sum_{j:h_j > h} \lambda_j [\beta_j(G_h(s))]_{s_0}^{(i)}; \quad (18)$$

$$\omega_{h,2}^{\text{обp}(i)} = \begin{cases} 1 + \sum_{j:h_j = h} \lambda_j [\beta_j(G_h(s))]_{s_0}^{(i)} & \text{при } i=1; \\ \sum_{j:h_j = h} \lambda_j [\beta_j(G_h(s))]_{s_0}^{(i)} & \text{при } i > 1. \end{cases} \quad (19)$$

В свою очередь производные $G_h^{(i)}(s_0)$ функции (14) могут быть вычислены из (9)—(11), если положить $x(s) = G_h(s)$, а производные по s $[\beta_j(G_h(s))]_{s_0}^{(i)}$ сложной функции $\beta_j(G_h(s))$ — по (8) и (9), если положить $Z(s) = \beta_j(G_h(s))$, $x(s) = G_h(s)$.

Значения производных $\omega_h^{\text{обp}(i)}(s_0)$ могут быть найдены по формуле (6) с заменой функций $Z(s)$, $x(s)$, $y(s)$ и их производных на функции $\omega_h^{\text{обp}}(s)$, $\omega_{h,1}^{\text{обp}}(s)$, $\omega_{h,2}^{\text{обp}}(s)$ и их производные соответственно.

4. Аналогично п. 1—3 найти (k_r-1) -ю производную от (15).

5. Найти производные $t_h^{(i)}(s_0)$ от ПЛС ФР (12), применяя последовательно (4) к следующим

парам функций: $\beta_h(s)$ и $\omega_h^{\text{обp}}(s)$, $[\omega_h^{\text{обp}}(s)\beta_h(s)]$ и

$\omega^{\text{дост}}(s)$, вычисляя таким образом производные произведения трех функций.

После вычисления производных $t_h^{(i)}(s_0)$, ($i=1, k_r-1$) искомая вероятность эрланговского старения $P_{\text{усл}}$ может быть найдена по формуле (3).

Заключение

Рассмотрен подход к вычислению вероятности старения информации, основанный на рекуррентных процедурах расчета производных большого порядка от ПЛС функции распределения времени формирования информационных услуг. Предложенный метод может использоваться для вычисления вероятности старения информации для центров обработки данных и других информационных систем при формировании и контроле выполнения SLA.

Литература

1. Советов Б. Я., Цехановский В. В. Информационные технологии. – М.: Юрайт, 2012. – 263 с.
2. Советов Б. Я., Цехановский В. В., Чертовской В. Д. Интеллектуальные системы и технологии. – М.: Академия, 2013. – 320 с.
3. Колбанев М. О., Яковлев С. А. Модели и методы оценки характеристик обработки информации в интеллектуальных сетях связи. – СПб.: СПбГУ, 2002. – 230 с.
4. Колбанев М. О., Татарникова Т. М., Воробьев А. И. Модель балансировки нагрузки в вычислительном кластере центра обработки данных // Информационно-управляющие системы. 2012. № 3(58). С. 37–41.
5. Колбанев М. О., Татарникова Т. М., Воробьев А. И. Модель обработки клиентских запросов // Телекоммуникации. 2013. № 9. С. 42–48.