

Ф. І. Шилов, М. А. Скулиш, А. Сафарян

Національний технічний університет України “КПІ імені І. Сікорського”, Київ, Україна

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МЕТОДУ ОПТИМАЛЬНОГО ВИБОРУ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ ДЛЯ БІЛІНГОВИХ СИСТЕМ

Якість послуг і дохід операторів мобільного зв'язку суттєво залежить від продуктивності та ефективності білінгових систем, що обробляють потоки вхідних заявок на обслуговування. Щоб на вхідних інтерфейсах білінгової підсистеми не виникали черги оброблюваних даних, необхідно контролювати об'єм вхідного потоку, підтримувати оптимальне навантаження і своєчасно планувати розширення білінгових підсистем, що обробляють такий потік. У даній статті запропоновано математичну модель пошуку оптимального навантаження на систему обслуговування із раннім виявленням перевантаження, яка дозволить забезпечити обслуговування із заданими показниками ефективності, а саме із заданою ймовірністю час обслуговування сервісів у системі онлайн тарифікації. На основі статистичних даних оператора зв'язку розроблено імітаційну модель системи обслуговування для різних масштабів, яка показала як при перевищенні допустимого розрахованого значення оптимального навантаження збільшується кількість відмов у обслуговування.

Ключові слова: якість обслуговування, система онлайн тарифікації, алгоритм раннього виявлення перевантажень, оптимізація ресурсів мережі, динамічне керування ресурсами, хмарні обчислення.

Вступ

На сьогоднішній день спостерігається перенасичення ринку пропозиціями мобільних операторів. Переважаюча частина населення вже підключена до одного або, навіть, кількох мобільних операторів одночасно. Конкурентоспроможність у більшій степені залежить від здатності утримати існуючих абонентів, ніж від залучення нових.

Лояльність абонентів напряму залежить від рівня якості наданих послуг; пропозиції надання нових сервісів; швидкості роботи сервісів. Якість повинна, якщо не стабільно покращуватись, то утримуватись на достатньо високому рівні.

У сучасному світі об'єми інформації, що передається, невідмінно збільшуються, що, у свою чергу, потребує постійного розширення пропускних можливостей мобільної мережі. Як наслідок, навантаження на внутрішні підсистеми обробки трафіку збільшуються, при цьому критично важливо уникнути перенавантажень на компонентах мережі, втрати даних і черг на обробку замовлень.

Дохід оператора мобільного зв'язку напряму залежить від організації процесу контролю фінансового стану рахунку (балансу) абонента - точності списання коштів абонента у режимі реального часу. Це дозволяє уникнути дебіторської заборгованості і підвищити комфорт користування послугами оператора.

Описані вище фактори визначають роль білінгової системи, як одного з головних компонентів в організації всього бізнес-процесу.

Якість послуг і дохід операторів мобільного зв'язку суттєво залежить від того, наскільки правильно розрахована продуктивність та ефективність білінгових систем, що обробляють потоки вхідних заявок на обслуговування викликів тарифікації. Для білінгових підсистем, які надають послуги по списанню коштів абонента у режимі реального часу, важливою характеристикою роботи є обробка вхідних заявок в найкоротші терміни, без створення черги.

Щоб на вхідних інтерфейсах білінгової підси-

стеми обслуговування викликів тарифікації не виникали черги оброблюваних даних, необхідно контролювати об'єм вхідного потоку, підтримувати оптимальне навантаження і своєчасно планувати розширення білінгових підсистем, що обробляють такий потік. Вибраний метод дозволяє провести розрахунок рекомендованого значення інтенсивності вхідного потоку на існуючій білінговій підсистемі, порівняти його з доступними метриками продуктивності і спланувати рекомендоване розширення білінгових модулів при прогнозованому збільшенні вхідного потоку.

Результати досліджень

Метод формування вхідного потоку навантаження для ефективного використання ресурсів обслуговування. Основна ідея методу полягає в тому, щоб виходячи з ергодичного розподілу для можливих станів системи сформувати вимоги до середнього значення вхідного навантаження, що дозволить максимально ефективно використовувати наявні фізичні ресурси обслуговування вхідного потоку заявок. Процес обслуговування моделюється як n -канальний обслуговуючий пристрій, час обслуговування заявки у каналі є випадковою величиною розподіленою за законом Пуассона.

Вхідні дані.

n – кількість каналів для одночасного обслуговування заявок.

μ – інтенсивність обслуговування заявки,

G – кількість ресурсів залучених для обслуговування заявок,

ν^g – об'єм g -го ресурсу необхідний для обслуговування у блоці однієї заявки, $g = 1 \dots G$;

I^g – доступний об'єм ресурсу g -го ресурсу який спільно використовується заявками.

s – допустима кількість запитів у черзі на обслуговування.

R – відсоток заявок, які обслуговуються у системі не більше допустимого часу затримки, визначається експертами.

l – кількість запитів у черзі, до досягнення якої блокується надходження запитів до системи, відповідно до алгоритмів раннього попередження перевантажень.

Вихідні дані.

λ – рекомендоване значення для інтенсивності вхідного потоку, який буде направлено на обслуговування у s-каналний обслуговуючий пристрій.

Застосування запропонованого методу складається з двох етапів.

Етап 1. Для багатоканальної системи обслуговування, відповідно до моделей К. Жернового, необхідно знайти ергодичний розподіл кількості заявок у системі за формулами:

$$p_0 = \frac{1 - \beta^{s-l}}{A_n(\alpha, \beta)}, \quad \beta \neq 1, \quad \alpha = \lambda/\mu, \quad \beta = \lambda/n$$

$$A_n(\alpha, \beta) = (1 - \beta^{s-l}) \sum_{k=0}^n \frac{\alpha^k}{k!} + \frac{\alpha^n}{n!} \left(\frac{\beta - \beta^{s-l+1}}{1 - \beta} - (s-l)\beta^{s+1} \right)$$

$$p_k = \frac{\alpha^k}{k!} p_0 \quad (k = \overline{1, n})$$

$$p_{n+k} = \frac{\alpha^n}{n!} \beta^k p_0 \quad (k = \overline{1, l})$$

$$p_{n+k} = \frac{\alpha^n \beta^k - \beta^s}{n! (1 - \beta^{s-l})} p_0 \quad (k = \overline{l+1, s-1})$$

$$p_{n+s} = \frac{\alpha^n (1 - \beta) \beta^s}{n! A_n(\alpha, \beta)} \quad (k = \overline{n+l+1, n+s-1})$$

Якщо $\beta = 1, \alpha = \lambda/\mu, \beta = \lambda/n$, тоді

$$p_k = \frac{n^k}{k!} p_0 \quad (k = \overline{1, n}); \quad p_{n+k} = \frac{n^n}{k!} p_0 \quad (k = \overline{1, l});$$

$$p_{n+s} = \frac{n^n}{n!} p_0 (s-l-1) p_{n+s} \Rightarrow$$

$$\Rightarrow p_{n+s} = \frac{n^n}{n! (s-l)} p_0 \quad (k = \overline{n+l+1, n+s-1});$$

$$p_{n+k} = \frac{n^k}{n!} p_0 - (k-l) p_{n+s} = \frac{n^k}{n!} \frac{s-k}{s-l} p_0 \quad (k = \overline{l+1, s-1})$$

Етап 2. Розв'язання оптимізаційної задачі пошуку максимального вантаження, що забезпечить виконання умов на допустиму кількість ресурсів обслуговування.

$$\lambda \rightarrow \max \begin{cases} 4 * \left(\sum_{i=1}^n i v_k^g p_i + n v_k^g \sum_{i=n+1}^s p_i \right) \leq V^g, g = \overline{1, G} \\ \sum_{i=1}^s p_i \leq R \end{cases}$$

У білінгвовій підсистемі балансування вхідного потоку відбувається з використанням модифікованої схеми циклічного розбору (Round Robin), для розрахунку рекомендованого значення інтенсивності вхідного потоку всієї системи буде достатньо проаналізувати метрики з одного DOCS сервера і масштабувати отримані значення на всю підсистему.

Для підстановки вхідних даних буде використана метрика DOCS-OCS-Congestion, зібрана на стороні одного DOCS сервера (Diameter-сервер). В

табл. 1 приведено відповідність метрик і лічильників білінгової підсистеми вхідним даним обраної моделі. Всі значення були взяті на підсистемі обслуговування викликів тарифікації інтернет трафіку в момент найбільшої завантаженості ВНТА.

Таблиця 1. Відповідність метрик системи вхідним даним моделі

Параметр	Відповідність метрики	Одночасне значення
n	DOCS-OCS-Congestion	500
μ	DOCS-OCS-Congestion	ВНТА=403/сек
G	DOCS-OCS-Congestion	2048Mb
v_k^g	DOCS-OCS-Congestion	4Mb
V^g	DOCS-OCS-Congestion	1612Mb
s	DOCS-OCS-Congestion	500/сек
R	Визначено експертом	90%
l	DOCS-OCS-Congestion	450

Розрахунок оптимального значення інтенсивності вхідного потоку для існуючої системи. Для розрахунку оптимального значення λ проведено аналіз доступних у білінгвовій підсистемі метрик для найбільш завантаженого дня в 2017 році (24.11.2017 «Чорна п'ятниця»). Максимальне значення інтенсивності вхідного потоку було розраховано у години найбільшого навантаження, на момент, коли метрики сигналізували про відмови в обслуговуванні (REJECTS) для відомої кількості сесій. Під час аналізу, для кожного запису була розрахована інтенсивність обслуговування заявки (μ) і одночасне значення інтенсивності вхідного потоку (λ). Базуючись на отриманих даних, можна зробити висновок, що максимальному допустимому значенню інтенсивності вхідного потоку для єдиного OCS процесу, при якому не виникає деградації сервісу (відсутній REJECTS) відповідає $\lambda = 860$ при $\mu = 430$. Дані розрахунку оптимального значення λ і μ занесені в табл. 2.

Графічне відображення розрахованої інтенсивності обслуговування заявки (μ) і одночасного значення інтенсивності вхідного потоку (λ) для єдиного OCS процесу зображено на рис. 1.

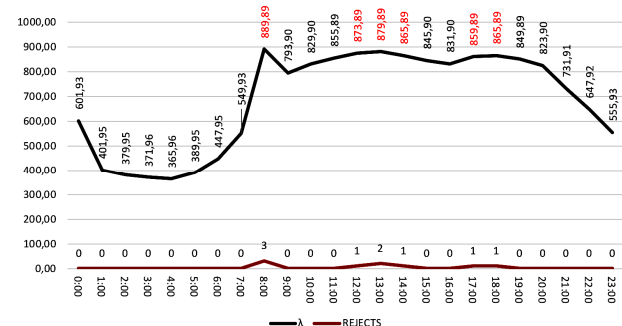


Рис. 1. Графік інтенсивності значення λ за 24.11.2017

Розрахунок оптимального значення інтенсивності вхідного потоку для системи, що масштабується. Відповідно до запропонованого розробником системи розширення фізичних компонентів, при масштабуванні системи – кількість OCS процесів,

що виконуються на одному DOCS сервері, буде збільшено у 5 разів, при цьому конфігурація OCS процесу змінена не буде. Таким чином, для системи,

що масштабується значення констант n , G , v_i^g , V_g , s , R , l , залишаться незмінними, а загальна кількість DOCS серверів буде збільшена.

Таблиця 2. Дані розрахунку оптимального значення λ і μ

DATE	TIME	GPRS1	GPRS2	REJECTS	μ	λ
24.11.2017	00:00	1211112	16102756	0	301	601.93
24.11.2017	01:00	977895	10609138	0	201	401.95
24.11.2017	02:00	962188	9984962	0	190	379.95
24.11.2017	03:00	971583	9719210	0	186	371.96
24.11.2017	04:00	973601	9570451	0	183	365.96
24.11.2017	05:00	1020312	10224994	0	195	389.95
24.11.2017	06:00	1157775	11722724	0	224	447.95
24.11.2017	07:00	1346571	14490450	0	275	549.93
24.11.2017	08:00	1749654	23858101	3	445	889.89
24.11.2017	09:00	1631822	21215782	0	397	793.90
24.11.2017	10:00	1693628	22199421	0	415	829.90
24.11.2017	11:00	1753250	22878271	0	428	855.89
24.11.2017	12:00	1817606	23379187	1	437	873.89
24.11.2017	13:00	1873052	23470265	2	440	879.89
24.11.2017	14:00	1907032	23036244	1	433	865.89
24.11.2017	15:00	1920535	22464995	1	431	861.89
24.11.2017	16:00	1945217	22005529	0	416	831.90
24.11.2017	17:00	1978970	22815321	0	430	859.89
24.11.2017	18:00	1986669	22961832	1	433	865.89
24.11.2017	19:00	1966866	22527160	0	425	849.89
24.11.2017	20:00	1958086	21745802	0	412	823.90
24.11.2017	21:00	1919809	19183564	0	366	731.91
24.11.2017	22:00	1831609	16819098	0	324	647.92
24.11.2017	23:00	1731484	14263415	0	278	555.93

Базуючись на розрахованому для такої системи значенні λ , можна розрахувати оптимальну кількість OCS процесів необхідних для підсистеми обслуговування викликів тарифікації інтернет трафіку, яка дозволить забезпечити доступність сервісів і цілісність даних, знизить вірогідність втрат і перенавантажень у системі після масштабування. Враховуючи, що один OCS процес існуючої білінгової підсистеми до масштабування, при оптимальному значенні $\lambda = 860$ здатен обробляти трафік з інтенсивністю $\mu = 430$ заявок в секунду, всіх 16 DOCS серверів з одиничним запущеним OCS процесом, оптимальна кількість трафіка, що обробляється буде дорівнювати 6880 сесій за секунду, що відповідає 24 768 000 сесіям в годину, опрацьованим всією підсистемою обслуговування викликів тарифікації інтернет трафіку.

Згідно з планами оператора мобільного зв'язку, інтенсивність вхідного потоку заявок повинна буде бути збільшеною до 50 000 000 сесій за годину. Виходячи з цього розробник білінгової системи дозволить розширити кількість OCS процесів до 30 штук зі збереженням конфігурації, що має забезпечити обробку зростаючого потоку вхідних заявок на білінгову систему. На рис. 2 зображено заплановане розширення білінгової підсистеми. Як видно, кількість DOCS серверів було зменшено до 6 штук, при цьому на кожному DOCS сервері запущено по п'ять OCS – процесів, які обробляють логіку дзвінка.

Оскільки, значення інтенсивності вхідного потоку за годину і кількість паралельно працюючих OCS процесів у новій системі відомі, розраховуємо інтенсивність обробки трафіку за секунду, окремо

на кожен процес. Використовуючи математичну модель, розраховуємо прогнозоване значення інтенсивності вхідного потоку λ . При 30 одночасно працюючих OCS процесах інтенсивність обробки трафіку за секунду буде $\mu = 463$. У табл. 3 відображено вхідні параметри для підставлення в імітаційну модель. Підставивши ці дані в імітаційну модель, отримуємо $\lambda = 925.89$, що значно перевищує рекомендоване значення ($\lambda = 860$).

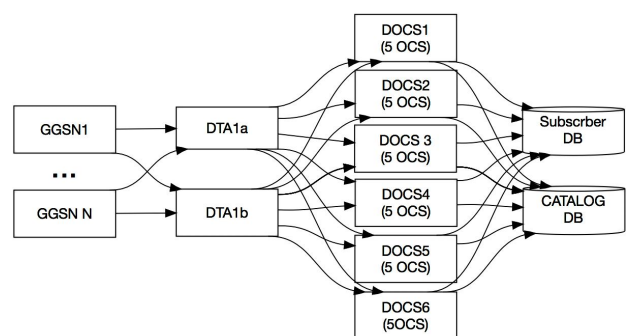


Рис. 2. Заплановане розширення білінгової підсистеми

З вищеприведеного слідує, що запропонований варіант масштабування білінгової підсистеми буде недостатнім для задоволення вимог. Більше того, фізична поломка одного DOCS сервера спричинить багатократне збільшення трафіку на DOCS серверах, які продовжать роботу, що у свою чергу, обов'язково, відобразиться на кількості неопрацьованих замовлень і може призвести до повної відмови всієї білінгової підсистеми.

Таблиця 3. Вхідні параметри для підставлення в імітаційну модель

Параметр	Значення
n	500
μ	463/сек
G	2048Mb
V_k^g	4Mb
V^g	1612Mb
s	500/сек
R	90%
l	450

Керуючись принципом надмірності при архітектурному проектуванні розширення системи, додаємо два DOCS сервера. У такому випадку загальна кількість DOCS серверів, підключених до кластеру, буде дорівнювати 8, а кількість OCS процесів буде збільшено з 30 до 40 штук.

При наявності 40 OCS процесів, інтенсивність обробки трафіку за секунду для кожного OCS процесу буде $\mu = 347$. Після підставлення отриманого значення в імітаційну модель, маємо $\lambda = 693.91$, що задовольняє розраховане рекомендоване значення ($\lambda = 860$). Навіть у випадку виходу з ладу одного DOCS сервера, прогнозоване значення інтенсивності вхідного потоку не перевищить рекомендоване значення і буде дорівнювати $\lambda = 793.90$, що перевищує відмовостійкість системи. На рис. 3 зображено покращене розширення з додаванням двох додаткових DOCS серверів.

Розрахунок номінального значення інтенсивності вхідного потоку після масштабування білінгової системи. Під час проведення робіт по масштабу-

ванню системи були враховані рекомендації по додаванню двох додаткових DOCS серверів. Білінгова. Підсистема була розширена до 8 DOCS серверів. Після переключення трафіку на нову підсистему, недовіків у роботі нової, розширеної підсистеми виявлено не було. Оператором мобільного зв'язку було прийнято рішення збільшити потік вхідних заявок до запланованої величини – орієнтовно, 50 000 000 заявок за годину. Система успішно витримала навантаження і продовжує роботу у штатному режимі. Через кілька місяців, після збільшення потоку вхідних заявок було проведено порівняльний розрахунок інтенсивності вхідного потоку λ , для підтвердження дієздатності обраного методу. Отримані дані оптимального значення λ і μ , після розширення системи, занесені у табл. 4.

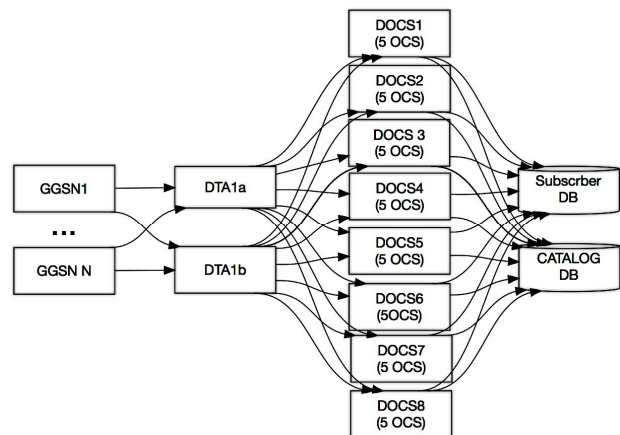


Рис. 3. Покращене розширення білінгової підсистеми

Таблиця 4. Дані порівняльного розрахунку λ і μ після розширення системи

DATE	TIME	GPRS1	GPRS2	REJECTS	μ	λ
11.04.2018	00:00	2725002	32205511	0	243	485.94
11.04.2018	01:00	2200264	21218275	0	163	325.96
11.04.2018	02:00	2164923	19969924	0	154	07.96
11.04.2018	03:00	2186062	19438419	0	150	299.96
11.04.2018	04:00	2190601	19140901	0	148	295.96
11.04.2018	05:00	2295701	20449988	0	158	315.96
11.04.2018	06:00	2604993	23445448	0	181	361.96
11.04.2018	07:00	3029785	28980900	0	222	443.95
11.04.2018	08:00	3936721	47716201	0	359	717.91
11.04.2018	09:00	3671600	42431565	0	320	639.92
11.04.2018	10:00	3810664	44398842	0	335	669.92
11.04.2018	11:00	3944812	45756542	0	345	689.92
11.04.2018	12:00	4089614	46758374	0	353	705.91
11.04.2018	13:00	4214366	46940530	0	355	709.91
11.04.2018	14:00	4290823	46072488	0	350	699.91
11.04.2018	15:00	4321204	44929989	0	342	683.92
11.04.2018	16:00	4376737	44011058	0	336	671.92
11.04.2018	17:00	4452683	44030643	0	337	673.92
11.04.2018	18:00	4470005	43737265	0	335	669.92
11.04.2018	19:00	4425450	42867923	0	328	655.92
11.04.2018	20:00	4405694	41305206	0	317	633.92
11.04.2018	21:00	4319571	38367128	0	296	591.93
11.04.2018	22:00	4121121	33638196	0	262	523.94
11.04.2018	23:00	3895840	28526830	0	225	449.95

Графічне відображення розрахованої інтенсивності обслуговування заявки (μ) і одночасного значення інтенсивності вхідного потоку (λ), для єдино-

го OCS процесу, після масштабування системи, зображено на рис. 4. Після завершення усіх необхідних розрахунків бачимо, що інтенсивність вхідного

поток не перевищує розраховане рекомендоване значення $\lambda = 860$, відмови в обслуговуванні – цілком відсутні.

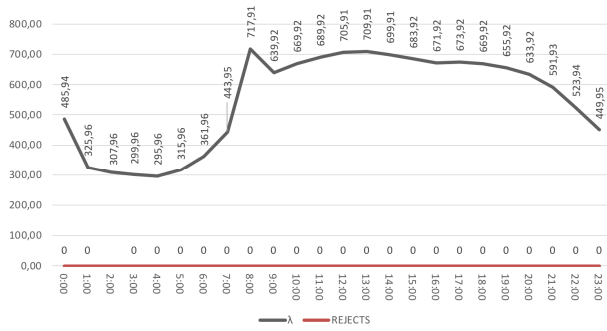


Рис. 4. Графік інтенсивності вхідного потоку (λ) за 11.04.2018

Для повноти аналізу проведемо розрахунок інтенсивності вхідного потоку так, якби рекомендоване додаткове розширення не було проведене, а загальна кількість DOCS серверів залишалась рівною шести. Отримані дані прогнозованого значення λ і μ , після розширення системи, але з неповною кількістю DOCS серверів, занесені в табл. 5.

Графічне відображення прогнозованої інтенсивності обслуговування заявки (μ) і одночасного значення інтенсивності вхідного потоку (λ), для єдиного OCS процесу, після масштабування системи, але без запропонованого додаткового розширення зображено на рис. 5. Розрахунок прогнозу показує багатократне перевищення розрахованого рекомендованого значення ($\lambda = 860$). Також за допомогою методу екстраполяції були прогнозовані відмови в обслуговуванні (REJECTS).

Таблиця 5. Дані порівняльного розрахунку λ і μ після розширення системи, при кількості DOCS серверів, рівній 6

DATE	TIME	GPRS1	GPRS2	Передбач. REJECTS	μ	λ
11.04.2018	00:00	2725001,5	32205511	0	323	645.92
11.04.2018	01:00	2200264,3	21218275	0	217	433.95
11.04.2018	02:00	2164922,5	19969924	0	205	409.95
11.04.2018	03:00	2186061,8	19438419	0	200	399.95
11.04.2018	04:00	2190601,4	19140901	0	198	395.95
11.04.2018	05:00	2295701,2	20449988	0	211	421.95
11.04.2018	06:00	2604993,3	23445448	0	241	481.94
11.04.2018	07:00	3029785,2	28980900	0	296	591.93
11.04.2018	08:00	3936721,1	47716201	100	478	955.88
11.04.2018	09:00	3671600,4	42431565	100	427	853.89
11.04.2018	10:00	3810664	44398842	300	446	891.89
11.04.2018	11:00	3944812,3	45756542	300	460	919.89
11.04.2018	12:00	4089614,1	46758374	400	471	941.88
11.04.2018	13:00	4214366	46940530	450	474	947.88
11.04.2018	14:00	4290822,9	46072488	400	466	931.88
11.04.2018	15:00	4321203,9	44929989	300	456	911.89
11.04.2018	16:00	4376737,3	44011058	300	448	895.89
11.04.2018	17:00	4452683,3	44030643	300	449	897.89
11.04.2018	18:00	4470004,5	43737265	300	446	891.89
11.04.2018	19:00	4425449,6	42867923	200	438	875.89
11.04.2018	20:00	4405693,5	41305206	0	423	845.90
11.04.2018	21:00	4319571,1	38367128	0	395	789.90
11.04.2018	22:00	4121120,9	33638196	0	350	699.91
11.04.2018	23:00	3895840	28526830	0	300	599.93

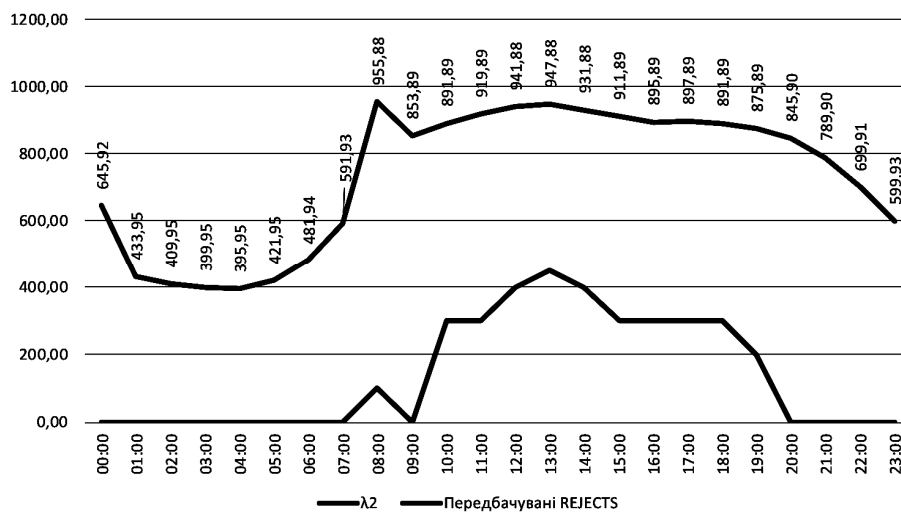


Рис. 5. Графік інтенсивності вхідного потоку (λ)

Після проведеного розширення білінгової підсистеми обслуговування викликів тарифікації, можна зробити висновок, що запропонований метод вибору обчислювальних ресурсів для обслуговування білінгових систем за умови коливання навантаження є оптимальним, початковий вибір цього методу був зроблений вірно.

Висновки

1. Запропонований метод і модель оптимального вибору обчислювальних ресурсів, для обслуговування білінгових систем за умови коливання навантаження є ефективними, у повній мірі враховують особливості архітектурних параметрів і можуть бути

застосовані у дослідженні для масштабування системи і контролю навантаження білінгової системи.

2. У ході проведення дослідження масштабування білінгової системи було виявлено, що запропонований варіант масштабування білінгової системи не здатний у повній мірі обробляти запланований до збільшення вхідний потік заявок.

3. З огляду на виявлений у ході дослідження суттєвий недолік масштабування білінгової системи, було запропоновано і реалізовано додаткове збільшення ємності DOCS серверів, що дозволило забезпечити доступність сервісів і збереження цілісності даних, підвищило відмовостійкість і знизило вірогідність перевантажень системи.

СПИСОК ЛІТЕРАТУРИ

1. Скулиш М. А., Заставенко А. А. Метод распределения ресурсов сервера оператора мобильной связи // Вісник Національного технічного університету України КПІ. Серія: Радіотехніка. Радіоапаратобудування. – 2015. – №. 60.
2. Скулиш М. А. Організація роботи групи серверів для забезпечення потреб розподіленої системи тарифікації послуг // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – №. 5. – С. 58-64.
3. Larisa G., Mariia S., Andriy R. Control strategy of the input stream on the online charging system in peak load moments // Microwave & Telecommunication Technology (CriMiCo), 2014 24th Int. Crimean Conf – IEEE, 2014. – С. 312-313.
4. Жерновий Ю.В. Імітаційне моделювання систем масового обслуговування. Львів: ЛНУ імені Івана Франка, 2007.
5. Жерновий К. Ю., Жерновий Ю. В. Система M0/G/1 с гістерезисним перемиканням інтенсивності обслуговування. Інформаційні процеси, 2012, т. 12, No 3, стр. 176–190.
6. Жерновий К.Ю., Жерновий Ю. В. Системи M0/G/1/m і M0/G/1 з часом обслуговування, що залежить від довжини черги. – Львів: Видавничий центр ЛНУ імені Івана Франка // Информационные процессы, Т. 13, No 2, 2013, с. 76–90.
7. А. Н.. Однолинейные системы массового обслуживания : учебное пособие / А. Н. Соколов, Н. А. Соколов. – СПб. : Изд-во «Теледом» ГОУВПО СПбГУТ, 2010. – 112 с.
8. P. Eronen, T. Hiller, G. Zorn. RFC 4072 Diameter Extensible Authentication Protocol (EAP) Application. August 2005. — Режим доступу: <https://tools.ietf.org/html/rfc4072> — Дата доступу: 14.05.18. — IETF RFC Tools
9. Pat R. Calhoun, John Loughney, Jari Arkko, Erik Guttman, Glen Zorn. RFC 3588 Diameter Base Protocol. September 2003. — Режим доступу: <https://tools.ietf.org/html/rfc3588> — Дата доступу: 14.05.18. — IETF RFC Tools.

Рецензент: д-р техн. наук, проф. Г. А. Кучук,

Національний технічний університет “Харківський політехнічний інститут”, Харків

Received (Надійшла) 21.02.2018

Accepted for publication (Прийнята до друку) 18.04.2018

Исследование эффективности метода оптимального выбора вычислительных ресурсов для биллинговых систем

Ф. И. Шилов, М. А. Скулиш, А. Сафарян

Качество услуг и доход операторов мобильной связи существенно зависят от производительности и эффективности биллинговых систем, обрабатывающих потоки входящих заявок на обслуживание. Чтобы на входных интерфейсах биллинговой подсистемы не возникали очереди обрабатываемых данных, необходимо контролировать объем входящего потока, поддерживать оптимальную нагрузку и своевременно планировать расширение биллинговых подсистем, обрабатывающих такой поток. В данной статье предложена математическая модель поиска оптимальной нагрузки на систему обслуживания с ранним выявлением перегрузки, которая позволит обеспечить обслуживание с заданными показателями эффективности, а именно с заданной вероятностью обслуживания сервисов в системе онлайн тарификации. На основе статистических данных оператора связи разработана имитационная модель системы обслуживания для различных масштабов, которая показала как при превышении допустимого рассчитанного значения оптимальной нагрузки увеличивается количество отказов в обслуживании.

Ключевые слова: качество обслуживания, система онлайн тарификации, алгоритм раннего выявления перегрузок, оптимизация ресурсов сети, динамическое управление ресурсами, облачные вычисления.

Investigation of the method of computing resources optimal choice for billing systems effectiveness

F. Shilov, M. Skulysh, A. Safaryan

The quality of services and the income of mobile operators significantly depend on the performance and efficiency of billing systems that process incoming request flows. To avoid queues on the input interfaces of the billing subsystem, it is necessary to control the amount of incoming traffic, maintain the optimal load, and plan the expansion of billing subsystems processing such a flow in a timely manner. In this paper, we propose a mathematical model for finding the optimal load on a service system with an early detection of congestion, which will allow servicing with specified performance indicators, namely, with a given probability of servicing in the online charging system. On the basis of statistical simulation operator service has been designed for different load reduction, which showed the number of service failures grows in violation of the optimal load.

Keywords: QoS, OCS, RAD, network resource allocation, dynamic resource control, Cloud computing.