

МОДЕЛЮВАННЯ СИСТЕМИ ОБРОБКИ ІНТЕНСИВНИХ РОЗГАЛУЖЕНИХ ІНФОРМАЦІЙНИХ ПОТОКІВ

М. С. Стремєцька^{1, а}, М. М. Савчук¹

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
Фізико-технічний інститут

Анотація

В роботі запропонована та досліджується модель багаторівневої ієрархічної системи обробки інтенсивних інформаційних потоків для обслуговування національної системи оплати комунальних послуг. Проводиться детальний аналіз функціонування та обчислення характеристик моделі системи, яка описує процеси надходження та обробки вимог на різних рівнях. Для обґрунтування практичної ефективності побудованої багаторівневої ієрархічної системи використовуються математичний апарат теорії масового обслуговування та технічні характеристики сучасних засобів обробки і передачі інформаційних пакетів по мережам.

Ключові слова: багаторівневі ієрархічні системи обробки інформації, системи масового обслуговування (СМО), інтенсивність вхідного потоку заявок, інтенсивність обслуговування заявок, завантаженість системи, обсяг черги

Вступ

При сучасних темпах зростання інформаційного обміну в усіх галузях суспільного життя стає вкрай актуальною задача побудови великих розгалужених мереж передачі інформації, що можуть забезпечити захищене з'єднання між центрами обробки даних та кінцевими користувачами. Важливо, щоб такі мережі мали високу відмовостійкість і забезпечували швидку та ефективну обробку інформаційних запитів, особливо якщо це стосується критичної інфраструктури, – такої, наприклад, як національна система оплати комунальних послуг.

Побудова сучасної системи передачі даних в масштабах всієї держави є революційним кроком у вирішенні ряду економічних та соціальних завдань і у майбутньому зможе замінити складні та неефективні механізми обміну даними між державними органами та населенням України. Це дозволить скоротити кількість інституцій для прийому звернень громадян, і перенести всі подібні запити в площину електронного документообігу, що допоможе зекономити ресурси і особистий час українців, виключивши людський фактор і бюрократію в органах державної влади. Звільнені ресурси від ліквідації таких установ державні компанії та профільні міністерства зможуть направити на створення нових ефективних центрів обслуговування громадян з сучасним клієнт-орієнтованим підходом. В свою чергу, держава зможе вирішувати задачі інформування громадян про реформи, нововведення і т.д., а також мати повну аналітику про потреби українців, а отже, забезпечити всі умови для їх задоволення.

Розглянемо мережу передачі даних на прикладі національної системи оплати комунальних послуг [1].

Для моделювання такої системи в рамках національного масштабу необхідно, в першу чергу, оцінити кількість рівнів ієрархії та описати логіку передачі даних на цих рівнях. Розглянемо варіант 3-рівневої архітектури:

- 1) Національний рівень;
- 2) Регіональний рівень;
- 3) Клієнтський рівень.

На національному рівні система складається з потужного центру обробки даних (ЦОД) та адміністративного апарату, що забезпечує менеджмент та контроль роботи всієї структури. На регіональному рівні пропонується розмістити представництва установ, що підтримуватимуть роботу на рівні області або багатомільйонного міста та матимуть достатньо обчислювальних потужностей для забезпечення обміну інформацією між кінцевими користувачами та своїми представництвами на регіональному рівні (далі потоки інформації будуть передаватись по захищеному VPN-каналі (Virtual Private Network) до рівня №1). В свою чергу, клієнтський рівень представляє собою встановлення спеціально розробленого веб-додатка на смартфони, планшети чи інші персональні гаджети, за допомогою якого мешканці всієї України зможуть передавати інформацію про оплату комунальних послуг до регіонального центру обробки даних.

Такий поділ на рівні зумовлений, в першу чергу, тим, щоб зменшити навантаження на систему, і по-друге, щоб диверсифікувати ризики, пов'язані з інформаційною безпекою. Адже важливо, що інформація, яка передається в цих каналах зв'язку, є критичною: містить персональні, фінансові дані і пов'язана з комунальними господарствами всієї держави, – тому має бути добре захищеною.

^а mira_strem@gmail.com

Виділені VPN з'єднання з національним ЦОДом матимуть також постачальники комунальних послуг: вони швидко і легко зможуть отримати всю інформацію про сплату рахунків і поточні показники лічильників своїх користувачів. Розглянемо ситуацію, коли таких комунальних підприємств є 4 види і всі вони згруповані між собою за видом комунальних послуг: постачання електроенергії, газопостачання, водопостачання та послуги ЖЕКів.

Метою роботи є побудова моделі багаторівневої ієрархічної системи обробки інтенсивних інформаційних потоків, характеристика процесів надходження та обробки заявок в термінах СМО, розрахунок завантаженості системи та довжини черги на різних рівнях ієрархії відповідно до технічних характеристик обслуговуваних пристроїв та пропускної здатності каналів передачі даних.

1. Методика дослідження

Опишемо нашу модель в термінах СМО [2, 3]. За Кендаллом класифікуємо систему наступним чином:

$$M|M|1|\infty.$$

Вхідний потік такої системи є елементарним пуассонівським потоком, обслуговування проводиться одним приладом, а час обслуговування має показниковий розподіл. Кількість місць для очікування не обмежена. Черга в системі організована за принципом First In, First Out (FIFO) [4]. Така СМО часто позначається як система $M|M|1$ з очікуванням і має назву моделі Ерланга.

Таким чином, для характеристики процесів надходження та обробки заявок необхідно встановити:

- інтенсивність вхідного потоку λ ;
- інтенсивність обслуговування μ ;
- середній обсяг черги.

Для початку розглянемо кількість вимог $Q(t)$ в системі в момент часу t . Легко помітити, що для даної системи процес $Q(t)$ є однорідним ланцюгом Маркова з неперервним часом. Нехай початковим станом ланцюга $Q(t) \in k$, тоді позначимо

$$p_i(t) = P\{Q(t) = i | Q(0) = k\}.$$

Розглянемо стаціонарний режим системи $M|M|1|\infty$. Припускаючи, що система знаходиться в стаціонарному режимі, $p_i(t) \equiv p_i$, можна отримати систему рівнянь:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0; \\ -(\lambda + \mu)p_i + \lambda p_{i-1} + \mu p_{i+1} = 0, i \geq 1. \end{cases} \quad (1)$$

З системи рівнянь (1) знаходимо

$$p_i = \frac{\lambda}{\mu} p_{i-1} = \left(\frac{\lambda}{\mu}\right)^i p_0.$$

Так як повинна виконуватись умова нормування $\sum_{i=0}^{\infty} p_i = 1$, то необхідною умовою існування стаціонарного режиму буде нерівність $\lambda < \mu$. В цьому випадку $p_0 = 1 - \frac{\lambda}{\mu}$, а $p_i = (1 - \frac{\lambda}{\mu}) \cdot (\frac{\lambda}{\mu})^i$.

Відношення $\rho = \frac{\lambda}{\mu}$ називають завантаженістю системи. Можна показати, що нерівність $\rho < 1$ є і

Рахунок — Блокнот

Файл Правка Формат Вид Справка

Рахунок: за е/енергію згідно з рахунком №24628061

Отримувач:
ПАТ "Київенерго"
р/р: 26038301201
МФО: 322669
ЄДРПОУ: 00131305

Споживач:
о/р 08006300213
КС: 332

Індекс XXXXX, м. XXXXX, адреса XXXXX

Показник лічильника:
на 01/04/2016
XX.XXXX

Сума сплати:
XXXX

Рис. 1. Рахунок за комунальні послуги

достатньою умовою існування стаціонарного режиму. Знання величини p_i дозволяє знайти середнє значення кількості вимог в стаціонарному режимі:

$$MQ(t) = \sum_{i=0}^{\infty} i \cdot p_i = \sum_{i=0}^{\infty} i \cdot (1-\rho)\rho^i = \frac{\rho}{1-\rho}. \quad (2)$$

Формула (2) показує, що при значеннях ρ , близьких до 1, середня кількість вимог швидко зростає (а при $\rho = 1$ стаціонарний режим взагалі відсутній!)

2. Результати дослідження

Щоб розрахувати пропускну здатність каналів передачі даних і потужність обладнання в ЦОД на різних рівнях ієрархії, потрібно розрахувати розмір пакетів, що будуть передаватись в системі.

Для цього було оцінено приблизний розмір платіжної інформації про оплату окремої комунальної послуги на прикладі квитанції за електроенергію від ПАТ «Київенерго» (Рис. 1).

Обсяг такого повідомлення 280 байт. Стандартний розмір пакету, що передається без фрагментації (MTU – Maximum Transmission Unit) – 1500 байт. Середній розмір Overhead, який додається за рахунок заголовків пакетів TCP – 20 байт, IPsec – 50–60 байт, що в сумі складає не більше 150 байт при передачі повідомлення через TCP з'єднання та два VPN тунелі. Отже, в залишку маємо не менше 1300 байт для передачі інформативної частини повідомлення (PDU – Protocol Data Unit), що набагато більше, ніж вказаний обсяг платіжки у 280 байт. Це дозволяє внести додаткову інформацію в наше повідомлення, зокрема використати додаткові засоби криптографічного захисту [5]: ЦП, шифрування і т.д.

Що стосується потужності обладнання в центрі обробки даних на національному рівні, то вузьким місцем у даній системі буде міжмережевий екран, що виконуватиме функції захисту і перевірки вхідного трафіку: IDS – Intrusion Detection System, IPS – Intrusion Prevention System, побудова захищених VPN-тунелів, підтримка демілітаризованих зон DMZ, щоб розділити пакети різного змісту та ізолювати їх перед розпакуванням.

Для прикладу, розглянемо міжмережевий екран Cisco ASA серії 5500-X нового покоління – ASA 5555-X. Такий пристрій підтримує VPN з'єднання з шифруванням 3DES/AES, пропускною здатністю 700 Мбіт/сек. і загальною кількістю VPN сеансів між вузлами мережі з клієнтом IPSec IKEv1 – 5000 підключень.

Таким чином, знаючи максимальний розмір одного повідомлення та ряд технічних характеристик обслуговуючого пристрою, можемо розрахувати загальну інтенсивність обслуговування повідомлень μ_1 та середню інтенсивність обслуговування заявок по окремому каналу μ_2 . Нехай візьмемо максимальний обсяг пакета 1500 байт = 12000 біт = $12 \cdot 10^{-3}$ Мбіт. Сумарна пропускна здатність пристрою – 700 Мбіт/сек. Тож загальна інтенсивність обслуговування повідомлень складатиме:

$$\mu_1 = \frac{700}{12 \cdot 10^{-3}} \approx 58000 \text{ повідомлень/сек.}$$

Аналогічно, якщо припустити, що захищених VPN з'єднань буде близько 30-ти (обласні центри, багатомільйонні міста та 4 керуючі компанії по наданню комунальних послуг), то середня інтенсивність обслуговування заявок по кожному з каналів складатиме:

$$\mu_2 = \frac{58000}{30} \approx 1930 \text{ повідомлень/сек.}$$

Розрахуємо тепер інтенсивність вхідного потоку повідомлень λ для національного рівня ієрархії. Припускаємо, що середньостатистична сім'я складається з трьох осіб, а кількість платіжок на одну сім'ю складає 4 штуки: ЖЕК, вода, електроенергія та газ. Час t для обробки даних приймемо 10 год. = 36000 сек.

Таким чином, для знаходження інтенсивності вхідного потоку λ застосуємо формулу:

$$\lambda = \frac{4 \cdot N}{3 \cdot t}, \quad (3)$$

де N – населення відповідного регіону (території).

Для сумарного потоку заявок на національному рівні та для окремих регіональних каналів (обчислення проведено для областей України з найбільшим населенням і, відповідно, – з найбільшою кількістю заявок) інтенсивність вхідного потоку λ виглядатиме наступним чином (Табл. 1).

Табл. 1. Інтенсивність обслуговування заявок та інтенсивність вхідного потоку

	N , ос.	μ_i	λ
Україна	42760516	58000	1583,72
м. Київ	2906569	1900	107,65
Одеська обл.	2390289	1900	88,53
Дніпропетровська обл.	3254884	1900	120,55
Донецька обл.	4265145	1900	157,97
Київська обл.	1732235	1900	64,16
Луганська обл.	2205389	1900	81,68
Львівська обл.	2534174	1900	93,86
Харківська обл.	2718616	1900	100,69

Як бачимо, для розглянутих каналів передачі даних виконується умова $\lambda < \mu$ (аналогічно ця умова виконуватиметься і для інших регіонів України, так як їх населення на порядок менше, ніж тих, що ми розглянули вище). Звідси випливає, що завантаженість системи $\rho = \frac{\lambda}{\mu} < 1$ – тобто умова існування стаціонарного режиму виконується, що дозволяє проводити дослідження характеристик системи в зазначеному режимі далі.

Аби система на практиці дійсно працювала без відмов, цікаво дослідити розмір черги, щоб розрахувати кількість необхідної пам'яті для зберігання заявок. Розрахуємо середню довжину черги L_{queue} за формулою математичного сподівання:

$$\begin{aligned}
 L_{queue} &= 0 \cdot p_0 + 0 \cdot p_1 + \sum_{i=2}^{\infty} p_i \cdot (i-1) = \\
 &= p_0 \sum_{i=2}^{\infty} \rho^i (i-1) = p_0 \sum_{i=1}^{\infty} i \rho^{i+1} = \\
 &= p_0 \rho^2 \sum_{i=1}^{\infty} i \cdot \rho^{i-1} = p_0 \rho^2 \sum_{i=1}^{\infty} (\rho^i)' = \\
 &= p_0 \rho^2 \left(\sum_{i=1}^{\infty} \rho^i \right)' = p_0 \rho^2 \left(\frac{\rho}{1-\rho} \right)' = \\
 &= p_0 \rho^2 \left(\frac{(1-\rho) + \rho}{(1-\rho)^2} \right) = \frac{p_0 \cdot \rho^2}{(1-\rho)^2} = \\
 &= \frac{(1-\rho) \cdot \rho^2}{(1-\rho)^2} = \frac{\rho^2}{1-\rho} > 0. \quad (4)
 \end{aligned}$$

Тож розрахуємо завантаженість системи ρ , середню кількість заявок за формулою (2) та середню довжину черги за формулою (4) для обраних регіонів (Табл. 2):

Як бачимо, середня довжина черги настільки мала, що не досягає навіть 1. Максимальна кількість вимог, очевидно, буде більшою, ніж середня, проте обчислення демонструють, що навіть для Донецької області, яка має найбільше значення ρ , ймовірність того, що в черзі буде більше 11 вимог, складає $9 \cdot 10^{-15}$. Ймовірністю такої події на практиці можна знехтувати, адже вона в порівнянні з ймовірністю втрати вимоги через відмову апаратури чи помилку оператора набагато менша. А незначний обсяг пам'яті, який необхідний для зберігання одного повідомлення, дозволяє помістити сотні таких повідомлень в одному мегабайті пам'яті і не завдає помітної шкоди для роботи системи.

Також за допомогою виведеної формули (4) було встановлено залежність середньої довжини черги від завантаженості системи ρ . Графіки показують, що з ростом ρ середня довжина черги зростає за експоненційним законом [6], зокрема при $\rho = 0,95$ середня довжина черги досягає вже 18 (Рис. 2).

Висновки

Побудовано модель багаторівневої ієрархічної системи обробки інтенсивних інформаційних потоків, що враховує необхідні обсяги заявок та технічні характеристики сучасних пристроїв обробки даних.

Табл. 2. Середня кількість вимог в системі та середня довжина черги

	μ_i	λ	ρ	$MQ(t)$	L_{queue}
Україна	58000	1583,72	0,027306	0,02807209	0,000766524
м. Київ	1900	107,65	0,056658	0,06006123	0,003402965
Одеська обл.	1900	88,53	0,046594	0,04887146	0,002277133
Дніпропетровська обл.	1900	120,55	0,063448	0,067746407	0,004298376
Донецька обл.	1900	157,97	0,083141	0,090680518	0,00753929
Київська обл.	1900	64,16	0,033767	0,034946805	0,001180041
Луганська обл.	1900	81,68	0,04299	0,044921203	0,001931164
Львівська обл.	1900	93,86	0,049399	0,051966186	0,002567083
Харківська обл.	1900	100,69	0,052994	0,055960036	0,002965572



Рис. 2. Залежність середньої довжини черги від завантаженості системи

Виконано детальний аналіз параметрів системи, що описують процеси надходження та обробки вимог в термінах СМО. Проведено дослідження впливу завантаженості системи на середню довжину черги та доведено, що обране мережеве обладнання ефективно справляється з поставленими задачами і ймовірність виникнення черг довжиною більше 11 заявок не перевищує $9 \cdot 10^{-15}$. В подальшому планується провести оцінку вартості такої системи з метою оформлення комплексної пропозиції щодо можливого впровадження запропонованого рішення для побудови багаторівневої ієрархічної системи обробки інтенсивних інформаційних потоків у реальній системі обробки даних в рамках національного масштабу.

Перелік використаних джерел

1. Стремєцька М. С. Глобальна мережа передачі інформації з криптографічним захистом засобами легкої криптографії / М. С. Стремєцька // Матеріали XIII Всеукраїнської науково-практичної конференції «Теоретичні та прикладні проблеми фізики, математики та інформатики», 21–23 травня 2015 р.— Київ: ПОЛІТЕХНІКА, 2015. — С. 148–150.
2. Анисимов В. В. Элементы теории массового обслуживания и асимптотического анализа систем: учебное пособие для вузов / В. В. Анисимов, О. К. Закусило, В. С. Донченко. — К. : Вища шк. Головное изд-во, 1987.— 248 с.
3. Коба О. В. Математичні методи моделювання систем: навч. посіб. / О. В. Коба, С. В. Пустова. — К. : НАУ-друк, 2012.— 192 с.
4. Гнеденко Б. В. Введение в теорию массового обслуживания / Б. В. Гнеденко, И. Н. Коваленко. — М.: Наука. Гл. ред. физ.-мат. лит.— 1987.— 336 с.
5. Смарт Н. Криптография / Н. Смарт. — М.: Техносфера, 2005. — 528 с.
6. Клейнрок Л. Теория массового обслуживания / Пер. с англ. И. И. Грушко // Л. Клейнрок; Пер. И. И. Грушко; ред. В. И. Нейман. — М.: Машиностроение, 1979. — 432 с.