

МЕТОД БАЛАНСУВАННЯ НАВАНТАЖЕННЯ У ВІРТУАЛЬНИХ МЕРЕЖАХ

Проблема балансування навантаження з метою підвищення ефективності використання віртуальних ресурсів актуалізувалася у зв'язку з масовим використанням хмарних обчислювальних систем на віртуальних платформах. Одним із аспектів проблеми є ефективне планування та розподіл завдань між віртуальними системами з метою оптимізації використання ресурсів та скорочення часу обчислень. Тобто часто виникає ситуація, при якій частина обчислювальних ресурсів простоює, в той час, як друга частина ресурсів перевантажена і, ймовірно, наявна велика кількість завдань на очікуванні свого виконання.

Необхідність балансування навантаження розподіленої системи виникає з наступних причин:

- неоднорідність структури обчислювального комплексу (наприклад, кластера), тобто різні вузли мають різну продуктивність;
- неоднорідність структури міжвузлової взаємодії, тобто лінії зв'язку які з'єднують вузли, можуть мати різні характеристики пропускну здатності.

Рішення задачі виявилось настільки актуальним що було запропоновано багато алгоритмів балансування навантаження. Всі ці алгоритми працюють або на підставі поточного стану системи, або на підставі статусу ініціатора алгоритму. Алгоритми які працюють на підставі статусу поділяються на статичні та динамічні, а ті, які працюють на підставі статусу можуть бути ініційовані відправником, отримувачем або бути комбінованою ініціації.

При цьому слід зазначити що властивості віртуальних середовищ суттєво залежать від того яке наповнення мають віртуальні машини які це середовище утворюють. Відповідно до цього можна стверджувати те, що ефективність будь-якого алгоритму балансування навантаження може суттєво залежати від конкретного середовища, де цей алгоритм застосовується. В даному випадку розглядається середовище SDN-NFV мереж.

Результати чисельних досліджень, проведені за останнє десятиліття свідчать про те, що трафік у комп'ютерних мережах має самоподібні властивості. Такий трафік відзначається значними затримками та втратою пакетів навіть у випадку, коли навантаження далеке від максимального. В зв'язку з цим почали активно розроблятися методи та алгоритми управління трафіком з урахуванням його самоподібних та мультифракторальних характеристик [1].

Система балансування навантаження будується на основі підсистеми регулювання та підсистеми керування і моніторингу. Підсистема регулювання реалізує алгоритм балансування, на підставі інформації щодо поточного стану системи, міри її завантаженості та стану динамічного розподілу трафіку. В підсистемі керування та моніторингу здійснюється збір та аналіз статистики щодо стану системи, визначення самоподібних та мультифракторальних властивостей вхідного трафіку, розрахунок розподілу потоків по вузлам мережі з урахуванням класифікації трафіку, завантаженості серверів та каналів.

Основним критерієм завантаженості розподіленої системи вважається сумарне значення дисбалансу компонентів системи. Запропонований метод балансування навантаження передбачає використання як внутрішнього так і зовнішнього моніторингу. Метод базується на застосуванні стратегії найкоротшої черги (*англ. Shortest Queue strategy*). При використанні цієї стратегії диспетчер балансування надсилає завдання на сервер із найменшою кількістю завдань у черзі. Якщо є кілька серверів із найменшою довжиною черги диспетчер випадково вибирає сервер із цього списку. Використання такої стратегії дозволяє вирівняти поточну кількість завдань на кожному сервері.

В роботі виконано дослідження запропонованого методу і загалом доведена його коректність у порівнянні з існуючими рішеннями.

ВИКОРИСТАНІ ДЖЕРЕЛА

1. Л.О.Кириченко, И.Н.Иванисенко, Т.А. Радивилова. *Анализ дисбаланса распределенной системы при самоподобной нагрузке.//Информационные технологии. Вісник ХНТУ – 2016 – Вип.3(58) – С.224-231.*