

РАЗРАБОТКА ПРОТОТИПА СРЕДСТВА ГОРИЗОНТАЛЬНОГО МАСШТАБИРОВАНИЯ ДЛЯ УСТРОЙСТВ КОНЕЧНОГО ПОЛЬЗОВАТЕЛЯ

Антонов М.А. (Университет ИТМО),

Кореньков Ю.Д. (Университет ИТМО)

Научный руководитель – к.т.н., доцент, Кореньков Ю.Д.
(Университет ИТМО)

Введение. Несмотря на интенсивное развитие аппаратной составляющей вычислительных систем, часто возникают ситуации, когда программное обеспечение требует большего количества ресурсов, чем может предоставить система. И поскольку вертикальное масштабирование менее эффективно, для большинства случаев приходится прибегать к горизонтальному. Однако поддерживать большое количество узлов на протяжении всего времени работы приложения не целесообразно, поскольку нагрузка носит неламинарный характер. Таким образом, становится ясно, что для эффективного использования имеющихся ресурсов необходимо решить задачу планирования и динамического изменения состава вычислительной сети. Сейчас существует некоторое количество решений, но все они являются отдельно работающим инфраструктурным программным обеспечением, использование устройств конечного пользователя в котором затруднительно, например, Kubernetes.

Основная часть. В основе предлагаемого решения лежит идея службы-супервизора, объединяющей устройства пользователя в оверлейную одноранговую p2p-сеть – слабо-связанный кластер, не имеющий централизованного управления. Связь между узлами данного кластера поддерживается только по мере возможности, когда устройства пользователя доступны. Основные задачи службы-гипервизора:

1. Сбор информации об уровне нагрузки устройств
2. Планирование операций балансировки для приложений кластера.
3. Управление жизненным циклом подчинённых приложений [1][2][3]

Приложение под управлением супервизора должно реализовывать специальный программный интерфейс для каждого из своих модулей. Данный интерфейс позволяет гипервизору управлять модулем – элементом выполняющегося приложения, а также предоставляет возможность для сериализации и десериализации его состояния, что является основой операций миграции элементов приложения в процессе балансировки нагрузки. Решение о выборе устройств, задействуемых в каждом отдельном акте балансировки, принимается на основе собранной супервизором информации о доступных ресурсах и связности между узлами. [1][2]

Связность между модулями прикладного приложения при этом поддерживается посредством инфраструктуры удалённых вызовов процедур, прозрачной к расположению этих модулей. Функционирование данной инфраструктуры обеспечивается благодаря распределенному между устройствами пользователя реестру, отвечающему за управление метаданными о подчинённых приложениях и статистикой о устройствах пользователя. [4][5]

Мониторинг узлов сети происходит в фоновом режиме по средствам платформу-зависимых интерфейсов, взаимодействие с которыми осуществляется с использованием JNI. Основными характеристиками являются доступные ресурсы процессора, памяти, а также подсистемы ввода-вывода, поскольку данные факторы являются ключевыми для оценки производительности устройства.

Защита каналов передачи данных между узлами основана на использовании SSL-соединений в сочетании с автоматически генерируемыми сертификатами, управление которыми автоматизировано на базе комбинации профилей пользователя и узла. При этом процедура добавления нового устройства в слабо-связанный кластер упрощена за счёт

передачи информации о конечной точке для первичного сетевого соединения через такие средства как QR-code и NFC.

В результате работы был разработан прототип системы горизонтального масштабирования для устройств конечного пользователя, предназначенной для автоматизированной балансировки нагрузки компонентов приложений. Созданный прототип состоит из службы-супервизора и библиотеки для прикладных приложений, поддерживающих автоматизированную балансировку нагрузки. Было создано демонстрационное приложение на базе движка символьных вычислений с поддержкой балансировки нагрузки для вычислительного ядра, показавшее применимость предложенного решения.

Список использованных источников:

1. Kuzenkova E., Korenkov I., Loginov I., Ilina A., Dergachev A. Automated Scalability System for Distributed Applications in Heterogeneous Environment//CEUR Workshop Proceedings, 2020, Vol. 2590, pp. 1-8
2. Kuzenkova E., Korenkov I., Loginov I., Ilina A., Dergachev A. Development of Horizontal Scaling Management System for Applications in Heterogeneous Environment//XI Международная научно-практическая конференция молодых ученых «Программная инженерия и компьютерная техника» (Майоровские чтения-2019), MICSECS 2019 SAINT-PETERSBURG The Majorov International Conference on Software Engineering and Computer Systems Международная конференция. (Санкт-Петербург, 12-13 декабря 2019г.): сборник трудов, 2020, No. XI, pp. 207-210
3. Telegram Open Network / Dr. Nikolai Durov. <https://ton.org/ton.pdf> (23.11.2022).
4. Grosso W., Java R. M. I., Sebastopol C. A. O'Reilly Media //Sebastopol, United States. – 2001.
5. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications / Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan. https://pdos.csail.mit.edu/papers/chord:sigcomm01/chord_sigcomm.pdf (25.11.2022).