УДК 004.413

РАЗРАБОТКА ОТКАЗОУСТОЙЧИВОГО МЕХАНИЗМА МЕЖКЛАСТОРНОЙ КООРДИНАЦИИ В KUBERNETES НА ОСНОВЕ BFT-КОНСЕНСУСА

Емелин Д.И. (Университет ИТМО)
Научный руководитель – Лисицина В.В.
(Университет ИТМО)

Введение. Kubernetes является стандартом для оркестрации контейнеризированных приложений, обеспечивая автоматизацию развертывания, масштабирования и управления. Критически важным компонентом Kubernetes является управляющая плоскость, которая отвечает за поддержание желаемого состояния кластера. Для обеспечения отказоустойчивости управляющей плоскости используется протокол консенсуса Raft, реализованный с помощью распределенного хранилище etcd. Ограничения протокола Raft в том, что он обеспечивает устойчивость только, когда узел просто перестает отвечать. В современных условиях, характеризующихся ростом кибератак, аппаратными сбоями и программными ошибками, возникает необходимость защиты от более сложного класса отказов – византийских , когда узел может вести себя произвольно: искажать данные, задерживать сообщения или действовать злонамеренно. Настоящая работа посвящена интеграции протокола византийской отказоустойчивости BFT-SMaRt в архитектуру Kubernetes для устранения этого фундаментального недостатка.

Основная часть. Целью работы является повышение отказоустойчивости управляющей плоскости Kubernetes, обеспечивающей поддержку как аварийных, так и византийских отказов. Для достижения поставленной цели необходимо решить задачи по интеграции BFT-протокола в существующую архитектуру, разработке модуля консенсуса и проведению тестирования устойчивости системы к различным типам атак.Архитектура разработанного решения включает три основных слоя.

Первый слой – слой оркестрации. Он сохраняет стандартную функциональность Kubernetes и включает набор мастер-узлов, на которых развернуты ключевые компоненты управляющей плоскости: API Server, принимающий и обрабатывающий запросы, Controller Manager, отвечающий за контроль состояния системы, и Scheduler, распределяющий нагрузку на рабочие узлы. Рабочие узлы, в свою очередь, выполняют пользовательские нагрузки и подчиняются решениям управляющей плоскости.

Второй слой – слой BFT-консенсуса. Он представляет собой распределенную прослойку, развернутую поверх мастер-узлов. Компоненты этого слоя функционируют изолированно от основных процессов Kubernetes и взаимодействуют между собой по защищенным каналам связи. Для обеспечения византийской отказоустойчивости используется библиотека BFT-SMaRt, реализующая протоколы достижения консенсуса в условиях потенциально недоверенной среды. Каждый мастер-узел содержит экземпляр BFT-модуля, который участвует в голосовании и подтверждении транзакций.

Третий слой – слой хранения. Он представлен распределенным хранилищем etcd, которое в стандартной конфигурации Kubernetes использует протокол Raft для обеспечения отказоустойчивости. В рамках разработанной архитектуры синхронизация состояния etcd между узлами осуществляется через слой BFT-консенсуса. Все операции записи и чтения ключевых данных проходят через механизмы византийского согласования, что позволяет сохранять целостность и консистентность состояния системы даже при наличии доступа к ограниченному числу скомпрометированных

узлов. Хранилище остается централизованным с точки зрения логики, но его состояние распространяется и подтверждается через BFT-протокол.

Ключевой особенностью разработанной системы является замена встроенного механизма консенсуса на протокол, способный корректно функционировать при наличии до f византийских узлов в кластере при условии наличия 3f+1 узлов. Это обеспечивает сохранение свойств безопасности и отказоустойчивости управляющей плоскости даже при DDoS-атаках, компрометации части узлов или наличии программных ошибок, вызывающих недетерминированное поведение. В системе реализованы механизмы обнаружения и изоляции византийских узлов, криптографической проверки полномочий лидера и защиты целостности данных в etcd.

Выводы. В ходе работы разработано архитектурное решение по интеграции протокола BFT в среду Kubernetes. Предложенный подход основан на трехуровневой модели, обеспечивающем разделение ответственности между компонентами системы. Предложенная архитектура сохраняет совместимость с существующими компонентами Kubernetes и не требует модификации их исходного кода, что обеспечивает возможность ее внедрения в существующие инфраструктуры.

Список использованных источников:

1. Barger A., Manevich Y., Meir H., Tock Y. A Byzantine Fault-Tolerant Consensus Library for Hyperledger Fabric // 2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). – IEEE, 2021. – С. 1–9.

2. Bessani A., Sousa J., Alchieri E.E.P. State machine replication for the masses with BFT-SMART // 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. – IEEE, 2014. – С. 355–362.

3. Castro M., Liskov B. Practical Byzantine fault tolerance // OsDI. – Т. 99. – 1999. – № 1999. – С. 173–186.

4. Ongaro D., Ousterhout J. In search of an understandable consensus algorithm // 2014 USENIX Annual Technical Conference (USENIX ATC 14). – 2014. – С. 305–319.

Емелин.Д.И. (автор)

Лисицина.В.В. (научный руководитель)