

Управление распределенными хранилищами данных на основе метаданных, умных контрактов и блокчейн-технологии

А.П. Демичев, А.П. Крюков

*Научно-исследовательский институт ядерной физики имени Д.В.Скобельцына,
Московский государственный университет имени М.В.Ломоносова,
119991, Москва, Ленинские горы, 1, стр.2*

Н.В. Приходько

*Новгородский государственный университет имени Ярослава Мудрого
173003, Великий Новгород, ул. Большая Санкт-Петербургская, д. 41*

Аннотация

В настоящее время осуществление крупных научных, инженерных и бизнес-проектов связано, как правило, с необходимостью хранения и обработки больших объемов данных. Это приводит к необходимости развивать новые, более экономичные и надежные, архитектуры и принципы работы информационных систем, в том числе систем хранения данных. Экстремальными вариантами архитектурных решений для таких систем являются полностью централизованные хранилища и хранилища на основе одноранговых P2P-сетей [1]. Однако, во многих случаях такие решения оказываются неприемлемыми, например из-за их высокой стоимости или низкой надежности, а оптимальным является промежуточное между такими экстремальными вариантами решение. Для его осуществления организации — участники крупного проекта — объединяют свои локальные ресурсы хранения в единый распределенный пул и, при необходимости, дополнительно арендуют облачные ресурсы хранения, возможно у нескольких провайдеров. Особенно выгодным с экономической и технической точек зрения такое решение может оказаться в случае, когда появляется потребность в хранении больших объемов данных в течение ограниченного срока осуществления какой-либо проекта и в ситуации, когда проект объединяет многих организационно несвязанных между собой участников. В общем случае такой распределенный пул хранилищ образует динамически меняющуюся среду (по мере необходимости могут подключаться новые хранилища или отключаться ранее входившие в пул). Задача заключается в том, чтобы объединить все эти хранилища и данные в них в единую систему в динамически меняющейся среде, а также обеспечить реализацию взаимных политик доступа к данным участвующих сторон. Например, владелец файла с данными (пользователь, создавший эти данные или организация, которой они принадлежат) должен иметь возможность управлять правами доступа к нему для других пользователей. Это подразумевает наличие способов децентрализованного управления правами доступа к данным в такой динамически меняющейся среде, обеспечения консенсуса участвующих сторон относительно содержания и порядка операций с данными и обеспечения надежной неизменяемой записи истории совершенных операций, то есть метаданных провенанса (МДП), для разбора и разрешения возможных коллизий между участниками проекта, а также владельцами хранилищ. Коллизии могут быть связаны с вопросами приоритета при получении результатов обработки данных, использования результатов, нарушении прав

доступа и т. п. Другими словами, необходимо обеспечить инструментарий для поддержки осуществления бизнес-процессов хранения и обмена данными в распределенной среде и при наличии административно несвязанных или слабо связанных организаций, участвующих в совместных проектах, или просто обменивающихся данными на определенных условиях.

Необходимо отметить, что хотя за последние годы был осуществлен целый ряд проектов по созданию систем для поддержки и управления метаданными, включая провенанс данных, но подавляющее большинство реализованных решений являются централизованными (см., например, обзор [2] и ссылки в нем), что плохо соответствует случаю использования распределенной динамически меняющейся среды. С другой стороны, в последнее время в разных прикладных областях большую популярность приобрели — благодаря наличию ряда важных преимуществ — распределенные реестры на основе технологии блокчейна [3,4]. В самое последнее время на основе блокчейна появились разработки и для систем управления МДП [5,6]. Однако, они рассчитаны на работу с одним хранилищем, не решают проблему обеспечения бизнес-процесса обмена данными между административно различными организациями и управления доступом к данным.

В данной работе предложен новый подход к построению системы управления метаданными провенанса и правами доступа к данным, основанный на интеграции блокчейн-технологии, смарт-контрактов и управления данными на основе метаданных. Разработаны принципы и алгоритмы работы такой системы, названной ProvHL (Provenance HyperLedger), которая является отказоустойчивой, безопасной, надежной с точки зрения сохранности и защищенности записей метаданных провенанса от случайных или намеренных искажений. Исследованы вопросы оптимального выбора типа блокчейна для такой системы, а также выбора блокчейн-платформы. А именно, предложено использовать эксклюзивный (permissioned) тип блокчейна и блокчейн-платформу Hyperledger Fabric (HLF, <https://www.hyperledger.org>) [7], на основе которой реализуется система ProvHL.

В настоящее время на базе НИИЯФ МГУ создан полигон, на котором развернут предварительный вариант прототипа ProvHL для реализации разработанных принципов и отработки алгоритмов работы системы. Создание системы ProvHL производственного уровня позволит существенно повысить качество и надежность результатов, получаемых на основе обработки и анализа данных в распределенной компьютерной среде.

Исследование выполнено за счет гранта Российского научного фонда (проект № 18-11-00075).

Литература

1 A. P. Kryukov and A. P. Demichev “Decentralized Data Storages: Technologies of Construction,” *Programming and Computer Software*, vol. 44, No. 5, pp. 303–315, 2018 (в печати).

2 F. Zafar et al. “Trustworthy Data: A Survey, Taxonomy and Future Trends of Secure Provenance Schemes,” *Journal of Network and Computer Applications*, vol. 94, pp.50-68, 2017.

3 E. Staff "Blockchains: The Great Chain of Being Sure about Things". *The Economist*; <https://www.economist.com/news/briefing/21677228-technology-behind-bitcoin-lets-people-who-do-not-know-or-trust-each-other-build-dependable>, 2016 (Retrieved 09.08.2018)

4 A. Baliga “Understanding Blockchain Consensus Models,” Tech. rep., Persistent Systems Ltd, 2017, pp. 1-14.

5 A. Ramachandran and M. Kantarcioglu “SmartProvenance: A Distributed, Blockchain Based Data Provenance System,” in CODASPY. '18: The 8th ACM Conference on Data and Application Security and Privacy, March 19 - 21, 2018. Tempe, AZ, USA.

6 X. Liang et al. "Provchain: A Blockchain-based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability," in Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 468-477. IEEE Press, 2017.

7 E. Androulaki et al. “Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains,” in Proceedings of the Thirteenth EuroSys Conference, April 23-26, Porto, Portugal, Article No. 30 , ACM, 2018.