A New Approach to Data Sharing and Distributed Ledger Technology: A Clinical Trial Use Case

David F. Ferraiolo, Joanna F. DeFranco, D. RIchard Kuhn, Joshua D. Roberts

Distributed systems have always presented complex challenges, and technology trends are in many ways making the software designer's job more difficult. In particular, today's systems must successfully handle:

- *Privacy*: Regulations such as GDPR require much stronger security for personal data, *and* that system owners must delete all personal data on a user's request.
- *Access control complexity*: Modern access control often uses rules that depend on data from sources outside the organization, requiring high performance networks with data integrity guarantees.
- *"Internet of Things" and ubiquitous sensor nodes*: Data sources can include building sensors, smart watches, medical sensors, and many other sensor types.

In short, systems must handle more data from disparate sources, and at the same time be responsive to new and emerging privacy and data retention requirements.. Recent developments in distributed ledger technology (DLT) have been applied to some of these challenges, with only limited success. A key problem is the immutability property of blockchain (the most prominent form of DLT), which conflicts with the requirement to allow deletion of user data.

Additionally, as data sources multiply, systems must handle many more data formats and APIs. Providing data access to these often incompatible platforms can be costly and error-prone. Fortunately, some new developments in DLT and federated data access make it possible to improve security and performance at the same time. In this paper, we show how these new tools can be used in a clinical trial data management problem, illustrating their potential for solving a wide range of distributed data problems.

Use Case – Clinical Data

Clinical trials have significant challenges resulting from the use of centralized data storage. By some estimates, up to 80% of clinical study research is not reproducible [1]. It is also estimated that fewer than half of clinical trials comply with FDA regulations to make the trial data available [2]. Use of conventional data storage methods contributes towards missing data and difficulty verifying source data. In fact, source data verification is 20-30% of a clinical trial budget [3].

The inability to effectively share data can be devastating. Recently there was a retraction of two articles published in The Lancet New England Journal of Medicine, one of the world's top medical journals [4]. One of the articles discussed Covid-19 treatment medications where the study results were based on the data from 671 hospitals on six different continents. The data could not be validated and was inaccessible for peer-review. Not only was the article retracted, but this publication/results caused other studies and progress on the same topic to be halted. The other article retracted was regarding the safety of blood pressure medication for people with Covid-19 – obviously, the speed of these results was crucial.

With the need for rapid progress in antivirals and other medications, conventional FDA rules are being modified regarding clinical trials [9], and there is still a need to ensure protection of proprietary and personally identifiable information (PII) to allow data sharing that could lead to effective remedies.

Much of this data is stored in database management systems (DBMSs) and of mutual interest to health care and biotechnology organizations. However, rapid sharing of DBMS data among users in different organizations, referred to as *relying parties* (RP), may be obstructed because: (1) DBMSs with different formats and record schemas are difficult to transfer, consume and interpret correctly among RPs, (2) the data is often sensitive, with authorization controlled by policies of the originating RP.

In addition, authorized access to data transmitted is hard to verify and significant risks arise from granting local access to unknown users with unknown credentials to databases with varying authorization mechanisms. This situation contributes to an unwillingness or inability to share critical data.

There is an opportunity to solve the database sharing problem of clinical trial data, while protecting proprietary, PII and other sensitive data through the integration of two proven NIST technologies: Next Generation Database Access Control (NDAC) and the data block matrix. These technologies collectively allow local access to DBMS data, down to the field level, in a federated environment, under local policies, as opposed to the transfer of data.

This system is an overlay of existing DBMS infrastructure and thus nonintrusive. In this short paper we will describe our solution to share access to existing disparate DBMSs.

A Trusted Federated System

The underlying Attribute-Based Access Control (ABAC) model is an access control method where access to resources is granted or denied based on assigned attributes and a set of policies that are specified in terms of those attributes [5]. Therefore, this effort recognizes user attributes as the “currency” for establishing access to resources in a federation of RPs when there exists a catalog of overlapping user attributes for achieving access to resources of mutual interest. Such a catalog (e.g., snomed-ct [6]), for example, may include roles and responsibilities used in the clinical trial industry for allowing a consistent nomenclature to access and select portions of trial information.

The effort further leverages NDAC, a NIST developed ABAC technology to deliver a standard means for accessing DBMS resources by imposing access control over database queries as middleware [7]. Consequentially, NDAC eliminates the need to utilize a mishmash of access controls implemented in conventional applications and/or in instances of proprietary DBMS mechanisms. Protection is achieved through translation of a user’s query to a permitted query. Thus, users may fetch entire data sets and NDAC restricts access to the subset of data permissible for the user in accordance with his/her attributes, down to the field level. The NDAC user attributes of each RP would need at least in part to include the attributes defined in the catalog.

A New Approach to DLT

A key property of blockchain that makes it an attractive technology for distributed systems is the integrity guarantees provided by its chained hashes of data blocks. But blockchain is not the only form of DLT. NDAC makes use of a centralized *data block matrix* which provides hashed data integrity protection, with the capability of editing or deleting records [8]. This technology provides blockchain’s integrity protection but solves the problem of deleting private data. Within NDAC, it also provides performance improvements by substituting protected local data for potentially large volumes of access control data. When a user is assigned to a user attribute in an NDAC system of a RP (and in the catalog), that assignment is also reflected in the data block matrix. Subsequent changes to those assignments would also be reflected.

Establishing the privileges for RPs to read from and write to the block matrix provides the basis for trust in the federation. Trust is further bolstered, using NIST's reference implementation of Next Generation Access Control (an ANSI/INCITS ABAC standard) by imposing policy over user-to-attribute assignments within a RP, under a governance policy. The governance policy feature provides a uniform and agreed upon approach for the management of user-to-attribute assignments (i.e., by whom, and under what authority).

As shown in Figure 1, if a user (e.g., researcher) from RP-y wants access to trial information in RP-x, the user could issue a consent request. In response, RP-x would conduct a policy review to determine the minimum set of attributes necessary for reading that information in their NDAC system, check the data block matrix for those attributes, and if found, automatically onboard the user as a temporary user via local assignments to those attributes in the NDAC system. Prior to onboarding, RP-x could selectively reduce access (e.g., time, eliminating fields). This system also uses a federated identity management approach for establishing sessions across the federation.

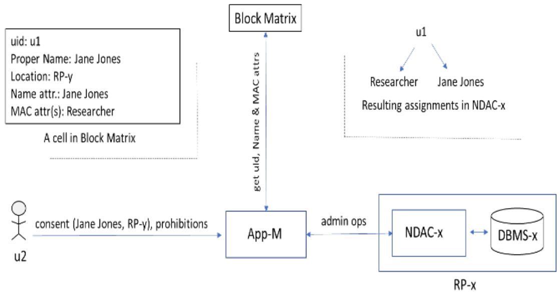

Figure 1. Sequence of events permitting policy-preserving access to database resources stored in RP-x by a user in RP-y.

NDAC is not only a nonintrusive integration with existing DBMS technology but also offers several advantages to instill confidence in sharing medical trial information. Among others the approach provides "purpose-based access" to targeted resources across federations, a basis of trust across the federation and, allows global access to DBMS resources, down to the field level according to local policies via consent.

References

[1] Benchoufi, M., Ravaud,P., "Blockchain technology for improving clinical research quality," Trials, Volume 18, No. 335, 2017.

[2] Anderson, M., Chiswell, K., Peterson, E., Tasneem A., Topping, J., Califf, R., "Compliance with Results Reporting at ClinicalTrials.gov," The New England Journal of Medicine, Volume 372, No. 11, 2015.

[3] Funning, S., Grahnen, A., Eriksson, K, Kettis-Linblad, A., "Quality Assurance within the Scope of Good Clinical Practice (GCP) – What is the Cost of GCP-related Activities? A Survey within the Swedish Association of Pharmaceutical Industry (LIF)'s Members," Quality Assurance Journal, Volume 12, 2009.

[4] Joseph, A., "Lancet, New England Journal retract Covid-19 studies, including one that raised safety concerns about malaria drugs," https://www.statnews.com/2020/06/04/lancet-retracts-major-covid-19-paper-that-raised-safety-concerns-about-malaria-drugs/ , June 2020.

[5] Ferraiolo, D., Chandramouli, R., Hu, V., Kuhn, R., "A Comparison of Attribute Based Access Control (ABAC) Standards for Data Service Applications," NIST Special Publication 800-178, October 2016.

[6] SNOMED International, "SNOMED CT Release File Specifications," SNOMED CT Release File Specifications - Release File Specification (ihtsdotools.org), February 6, 2020.

[7] Ferraiolo, D., Gavrila, S., Katwala, G., Roberts, J., "Imposing Fine-grain Next Generation Access Control over Database Queries," ABAC'17, Scottsdale, AZ, 2017, DOI: http://dx.doi.org/10.1145/3041048.3041050.

[8] Kuhn, R., Yaga, D., & Voas, J. (2019). Rethinking distributed ledger technology. *Computer*, *52*(2), 68-72.

[9] https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf#page=82

Biographies

David F. Ferraiolo (david.ferraiolo@nist.gov) is the manager of the Secure Systems and Applications group of the Computer Security Division, at the National Institute of Standards and Technology. He led the development of an Attribute-Based Access Control (ABAC) authorization system. This system serves as a research platform in support of a NIST led ANSI/INCITS standard under the title of Next Generation Access Control and Next Generation Database Access Control.

Joanna F. DeFranco (jfd104@psu.edu) is an associate professor of software engineering at the Pennsylvania State University and a guest researcher at the National Institute of Standards and Technology. She has worked as an Electronics Engineer for the Navy as well as a Software Engineer at Motorola. Dr. DeFranco is a senior member of the IEEE and on the advisory panel of IEEE Computer Magazine.

D. Richard Kuhn (kuhn@nist.gov) is a computer scientist in the Computer Security Division at NIST and is a Fellow of the IEEE. His primary research interests are in combinatorial methods for software verification and assurance.

Joshua D. Roberts (joshua.roberts@nist.gov) is a software developer at the National Institute of Standards and Technology. He is leading the effort to develop reference implementations of the technologies outlined in this paper.