Decentralized distributed data structure for bioanalytical laboratory setups

Kanwal Ashraf
Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology
Tallinn, Estonia
kanwal.ashraf@taltech.ee

Yannick Le Moullec
Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology
Tallinn, Estonia
yannick.lemoullec@taltech.ee

Tamás Pardy
Department of Chemistry and Biotechnology & Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology
Tallinn, Estonia
tamas.pardy@taltech.ee

ABSTRACT

The integration of information flow and management between different domains within laboratory setups has been under focus for many years. Communication technologies and data distribution play a key role in this integration. In this research, we propose a data distribution structure for bio-analytical laboratories based on a decentralized publish-subscribe model. In particular, we describe the main structure for data communication and the conceptual, logical, and physical data flow models. We believe this research can pave the way for enabling scalable and robust communication between the laboratory units.

CCS CONCEPTS

• **Information systems → Data structures**; • **Networks** → *Network architectures*; • **Computer systems organization → Distributed architectures**; *Real-time system architecture.*

KEYWORDS

Data distribution structure, scalable communication

1 INTRODUCTION

Scalable and high-performance communication technologies play a key role in real-time automated systems built upon efficient data exchange. The real-time data exchange via a publisher-subscriber architecture is more efficient as compared to a client-server architecture [Mastouri, mohamed anis and Hasnaoui 2007]. Different publish-subscribe middlewares, including MQTT, ROS and DDS, are used for efficient information flow and management for different applications including smart grids, health systems, air-traffic control, process industry and many more [Happ et al. 2017]. However, this concept is still emerging in biochemical and bioanalytical laboratories where client-based architectures [Gibbon 1996] are commonly used for information flow and management [Kang et al. 2018]. This research aims at integrating different domains including computational, communication, fluidic and biochemical domains of a high throughput bioanalytical laboratory setup using a decentralized communication architecture. The proposed structure is data-centric and system models are built around it using SAP PowerDesigner. Moreover, the data structure provides the flexibility to use the same concept for integrating different domains in other types of applications.

2 PROPOSED STRUCTURE/ARCHITECTURE

As mentioned earlier, the proposed data flow structure for bio-analytical laboratory setup is based on the well-known publish-subscribe model. The main elements of the structure include the entities, also known as topics, and different tasks running on the different devices in each unit. Each device in the laboratory will either subscribe or publish to a certain entity with a set of defined Quality of Service (QoS) parameters. These QoS parameters are part of the communication protocol and include delay and packet loss. The data flow between devices is highly decentralized and the same device in any unit could work both as subscriber and publisher for entities.

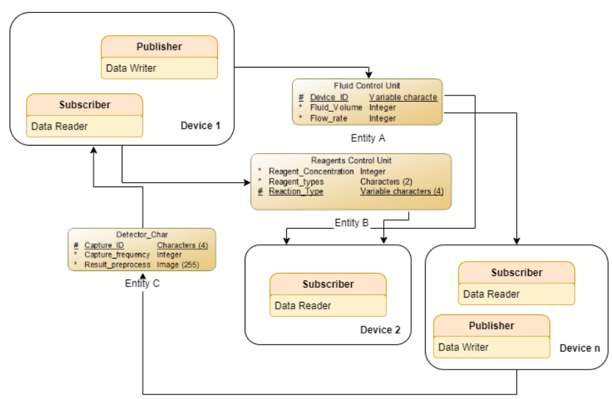

Figure 1: Main Structure for Data Communication. The communication-based data flow structure is decentralized and supports different devices in a bioanalytical laboratory unit.

For the use-case considered in this work, the bio-analytical laboratory setup is divided into three different units [Pärnamets et al. 2021], namely Biological Sample Processing Unit, Computational Unit, and Sensing Unit, with different data types.

Figure 1 shows the main decentralized communication-based data flow structure for different devices in a laboratory unit. SAP PowerDesigner is used to build the conceptual, logical, and physical data flow models for each unit. Figure 2 shows the conceptual model for the system, Figure 3 represents the logical data-flow, while Figure 4 depicts the physical data flow model for the system.

The conceptual model (Figure 2) depicts the information flow relation between different laboratory units which are defined as

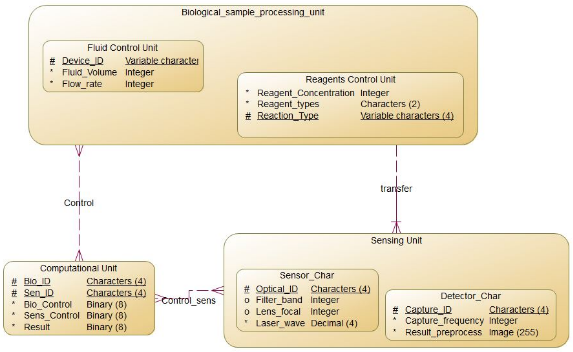

Figure 2: Conceptual Data Flow Model for Bio-analytical Laboratory setup

entities. The attributes for each entity are defined and many-to-many relations are used between different units. Each device is given an identification code and the data is categorized as either mandatory or primary. The logical data flow model is then built upon the conceptual model. Figure 3 shows the logical data flow model for the use-case bio-analytical laboratory setup, exposing the structure of the information flow.

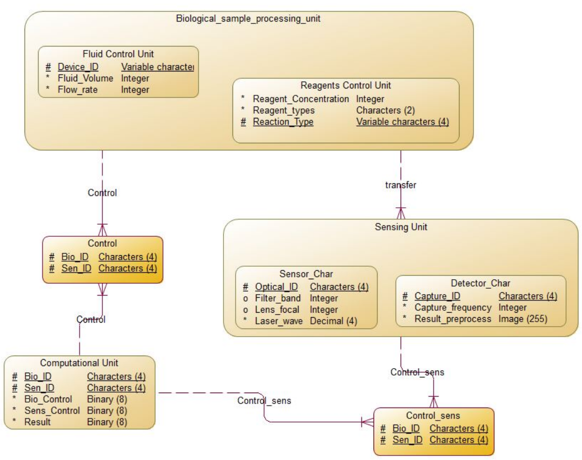

Figure 3: Logical Data Flow Model for Bio-analytical Laboratory setup

Next, by elaborating the logical data model, a more concrete physical data-flow model is generated, as shown in Figure 4. This physical data model includes more details that help analyze different objects and tables in the database and map the model onto the physical elements that will implement and execute the real system. To provide a comparison with centralized data flow structures, the performance evaluation of the purposed model will be performed in high traffic scenarios.

This research work is in progress and the resulting model is further aimed to be implemented in practice on different test units and the performance evaluation of the proposed data flow communication architecture is ongoing.

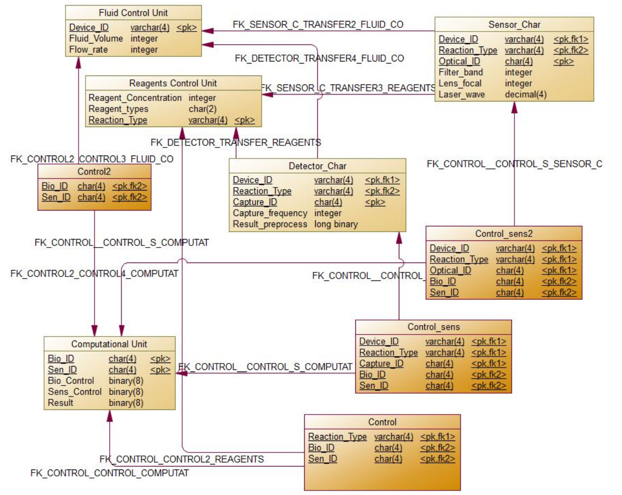

Figure 4: Physical Data Flow Model for Bio-analytical Laboratory setup

3 CONCLUSION

The proposed data flow structure supports the integration of the computational and communication domains with biological and fluidic domains for biochemical and bio-analytical laboratories. By implementing minor changes in entities, the proposed model could help in the robust, scalable, and efficient operation of other types of high throughput laboratory setups.

ACKNOWLEDGMENTS

This research was funded by the Estonian Science Agency ETAg, grant numbers PRG620 and TAR16013 Center of Excellence 'EX-CITE IT', the European Commission under Horizon 2020 ERA-chair Grant 'Cognitive Electronics COEL' (Agreement number: 668995; project TalTech code VFP15051).

REFERENCES

Gerst A. Gibbon. 1996. A brief history of LIMS. *Laboratory Automation and Information Management* 32, 1 (1996), 1–5. https://doi.org/10.1016/1381-141X(95)00024-K

Daniel Happ, Niels Karowski, Thomas Menzel, Vlado Handziski, and Adam Wolisz. 2017. Meeting IoT platform requirements with open pub/sub solutions. *Annales des Telecommunications/Annals of Telecommunications* 72, 1-2 (feb 2017), 41–52. https://doi.org/10.1007/s12243-016-0537-4

Wenjun Kang, Sabah Kadri, Rutika Puranik, Michelle N. Wurst, Sushant A. Patil, Ibro Mujacic, Sonia Benhamed, Nifang Niu, Chao Jie Zhen, Bekim Ameti, Bradley C. Long, Filipo Galbo, David Montes, Crystal Iracheta, Venessa L. Gamboa, Daisy Lopez, Michael Yourshaw, Carolyn A. Lawrence, Dara L. Aisner, Carrie Fitzpatrick, Megan E. McNerney, Y. Lynn Wang, Jorge Andrade, Samuel L. Volchenboum, Larissa V. Furtado, Lauren L. Ritterhouse, and Jeremy P. Segal. 2018. System for Informatics in the Molecular Pathology Laboratory: An Open-Source End-to-End Solution for Next-Generation Sequencing Clinical Data Management. *Journal of Molecular Diagnostics* 20, 4 (jul 2018), 522–532. https://doi.org/10.1016/j.jmoldx.2018.03.008

Salem Mastouri, mohamed anis and Hasnaoui. 2007. Performance of a Publish/Subscribe Middleware for the Real Time Distributed Control systems. (2007). https://www.researchgate.net/publication/239555403_Performance_of_a_PublishSubscribe_Middleware_for_the_Real_Time_Distributed_Control_systems

Kaiser Pärnamets, Tamas Pardy, Ants Koel, Toomas Rang, Ott Scheler, Yannick Le Moullec, and Fariha Afrin. 2021. Optical Detection Methods for High-Throughput Fluorescent Droplet Microflow Cytometry. *Micromachines* 12, 3 (2021). https://doi.org/10.3390/mi12030345