DARMA - INFORMATION SYSTEM FOR ACCESSING LHC DATA WITHIN THE ALICE EXPERIMENT AT CERN

Ján JADLOVSKÝ, Milan TKÁČIK, Slávka JADLOVSKÁ, Michal PÁNIK
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics Technical University of Košice, Letná 9, 042 00 Košice, Slovakia, tel. 055/602 3175,
E-mails: jan.jadlovsky@tuke.sk, milan.tkacik@tuke.sk, slavka.jadlovska@tuke.sk, michal.panik@student.tuke.sk

ABSTRACT

The paper deals with the implementation of systems for remote data access in cooperation with the international laboratory CERN in Geneva. The first part discusses extensions of the DARMA information system for archive management. The main goal of this paper is to design and enhance the system with new functionalities and solve known shortcomings, and thus qualitatively increase the potential of the DARMA system in the context of the ALICE DCS (Detector Control System). The final part of the paper deals with the analysis of historical data and the design of a system for the anomaly and failure detection in the infrastructure of high voltage channels within the TRD detector (Transition Radiation Detector).

Keywords: *Database Information Systems, Oracle Relational Databases, Data analysis, Failure detection, ALICE Experiment at CERN*

1. INTRODUCTION

Currently, the DARMA system represents one of the supporting tools in the context of ALICE DCS (Detector Control System) [1]. The key task of the system is to provide remote access to the offline data for the scientists and technical stuff of the CERN laboratory in Geneva. The DARMA system serves as a web application for accessing data and managing download requests and data visualizations of archival offline data [2].

Access to the historical DCS data and conditions data is necessary in order to reconstruct collisions detected by the ALICE experiment. During the LHC Run 1, this access was provided by the AMANDA (ALICE MANager for Dcs Archives) system, which served as an interface between dedicated DCS database and remote clients but had very limited functionality [3]. Therefore, a new version called AMANDA 3 was implemented, which provided the possibility of concurrent access to DCS archives using multiple clients and thus improved data availability to more users [4]. However, AMANDA 3 suffered from several problems, like often failures while downloading large amounts of data and that is a reason why the DARMA system was needed.

The DARMA system is conceptually based on the AMANDA 3. The AMANDA 3 system was designed and implemented by the Center for Modern Control Techniques and Industrial Informatics (CMCT&II) at the Department of Cybernetics and Artificial Intelligence (DCAI), Technical University of Košice [4].

The new software design enhances the user experience with new functionalities and preserves all base requirements of the former system. DARMA system is built on top of several modern technologies and platforms such as ASP.NET framework and user-friendly user interface. The new design can be considered as a reaction to the drawbacks of the former system such as high maintenance requirements, unused potentials of the complex distributed design that increased the total complexity of the system without the expected benefits.

The newly implemented DARMA system as an information system is currently deployed in the infrastructure of the ALICE experiment and it is closely connected to the DCS system. Specified data collected from DCS are continuously stored in archival databases. The amount of the data depends on the rules of archiving. The rules are defined at the SCADA (Supervisory Control and Data Acquisition) level of the distributed infrastructure [3].

2. ARCHITECTURE OF THE DARMA SYSTEM

The structure of the DCS archive is specific to WinCC OA and requires deep knowledge of its architecture. There is no other simple solution to access DCS archives. The only offered method is to use the SCADA system and its panels, however this requires specific knowledge of its programming language and access to the isolated DCS network.
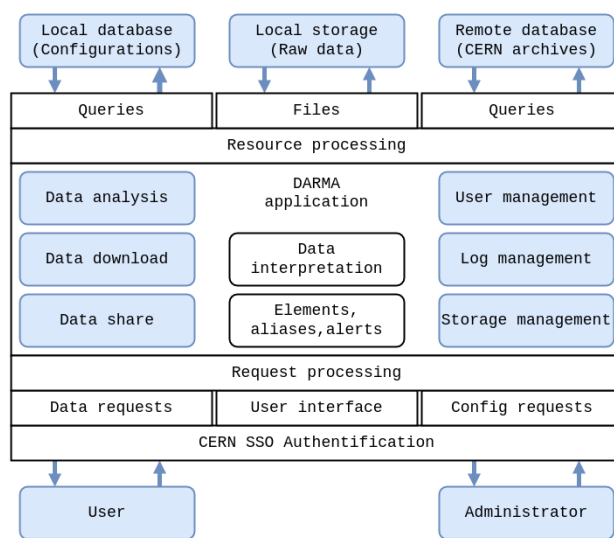

Fig. 1 DARMA structure diagram

The users are reluctant to develop their own tools and it would also be counterproductive. For example, each phys-

ical channel stored in database has its own unique name. This can represent for example the output channel of high voltage power supply. It can be connected to different parts of the detector. This technique is often used to balance the noise in setups with tens thousands of channels. To keep track of the connections, an additional name (alias) is assigned to each channel and reflects its use. While the name of the channel remains fixed, during the operation it can serve the different parts of detector. DARMA provides seamless method for tracking these changes and user can formulate a single request to retrieve voltage applied to a given module without the need to track the physical connections. This task will be performed by DARMA.

The DARMA system is built as an ASP.NET application with widely used MVC (Model-View-Controller) architecture design pattern. The top part of Figure 1 represents a Model behavior such as abstraction of data entities and data preservation layers. This layer contains a local MS SQL database, remote Oracle relational databases and temporary data composed of log files and selected data downloaded from remote databases. The middle part of the figure represents Controller behavior such as main logic, request processing, URL routing, user authentication, logging subsystem and visualization preprocessing. Finally, the bottom part corresponds with the presentation layer using HTML templates (ASP.NET Razor) and JavaScript interactive functionalities building together the screens of the system.

The DARMA system connects to the offline copies of the archival databases. Data in these relational databases are stored in predefined structure according to the WinCC OA schemas. Data are replicated to the several databases in the network to prevent data loss [5]. Offline copies are read-only and synchronized with the ALIONR database using Oracle Active Data Guard.

The DARMA system is intended for deployment on a dedicated web server. It is an ASP.NET application, so it is primarily intended for deployment on the Windows Server operating system. The minimum requirements for technical parameters of server and OS are given according to the minimum requirements of the application modules required by the DARMA, such as .NET framework 4.7, ASP.NET MVC extension 4.2, IIS 8.5 web services and SqlLocalDB 12.0 database. Since the DARMA was developed for needs of CERN users, it is designed to work only with relational databases that use CERN schemas. However, an open-source nature of all software developed for CERN makes it possible to modify the DARMA system to use other authentication methods and database schemas.

Fig. 2 Create request form

3. USER INTERFACE

Users of the DARMA system are mostly scientists and technical staff of CERN. Logging into the system is underlied by having a CERN account. Before working with DARMA, user has to be authenticated by SSO (Single Sign On) method. After login, user is authorized by one of the roles, USER or ADMIN. ADMIN user can perform more action with system then USER. System is divided into sections. USER can see sections Request, Share and Help. ADMIN can access also sections Manage and Logs.

In the main section - Request - user can see a list of his or her own download requests with a set of options per each of them. These actions allow to download, share, analyse, show detail and delete requests. User can create a new download request by using a form shown in Figure 2. This form allows to create a simple and even more complex data request by specifying its parameters, such as datasource, archive, datetime range, alert options, type of identifiers, list of identifiers, request name, datetime format, output identifier format and sort option. The system supports several interactive tools to simplify the process of creating requests, for instance importing and exporting the identifiers or filter box.

The process for downloading archival data communicates with the selected remote database and the current status is stored in the local system database. After successfully processing the request, the user has the option to download data to the client device, analyze the data in interactive graphs or share the data with other users. The request management process is logged in the system database and in the log file.

The section Shared contains a list of download requests that are shared for the user by other users. The section allows similar options to the main Request section. Deleting shared request will break sharing relation, but the original request will still be available to the request owner.

The Manage section contains two subsections: Download management and User management. The Download management subsection contains a list of all created requests along with the classical options for working with these requests. The subsection also contains basic information about the state of the local storage, such as free disk space or total disk capacity. Access roles for users can be set in the User management subsection. The system administrator with the ADMIN role can grant or revoke access rights to other registered users.

4. NEW FEATURES AND IMPLEMENTATIONS

Based on the feedback of users of the system, some new features were selected and implemented into the DARMA system. The most significant new feature is the ability to download data from multiple database sources that can be dynamically added and removed. Another added feature is the data download monitoring module, which creates a log file containing information about the amount of down-loaded records over time. The last new feature is the console batch client for downloading data, which provides the same functionality as the web interface, but is more convenient for downloading large amounts of data than the web interface.

4.1. Dynamic datasource configuration

So far, the DARMA system was able to communicate with only one archival database. With later development we added a possibility to dynamically manage new datasource connections. As a result, we extended the system with ALICEITSDB a ALIDCSSURF. Formerly the datasource was split based on corresponding detectors and we changed them into archives. Internal data model was updated by adding new necessary fields and extra information, especially in tables DOWNLOAD and CONFIGURATION, responsible for storing download request configuration. An administrator can easily add new datasources by a simple configuration in the Web.config file of ASP.NET application as shown in the Figure 3. Number of datasources is practically unlimited. The actual number has to be set in the field of $count\_of\_remote\_connections$ in Web.config file. The set of archives is defined by a specific stored procedure of remote database. If no archives are defined, only one default archive is shown.
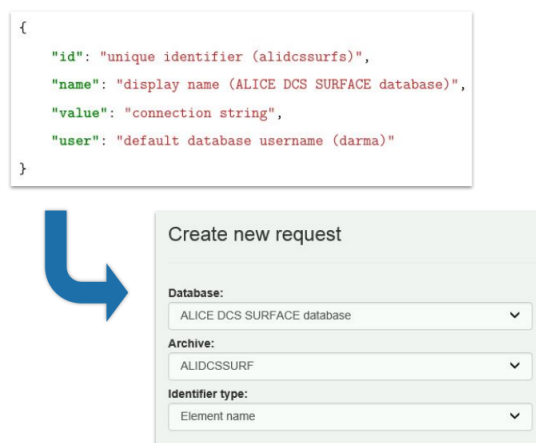

Fig. 3 Dynamic datasource configuration

In high energy experiments, the online database is installed on protected network and stores data coming from production systems in the ALICE cavern. However the commissioning of new detectors happens in external labs, or labs on the surface, not serviced by the DCS network. DCS also monitors parameters in detector labs (like environment, dust particles, etc.) and these are not stored on the main online database. To avoid operation of several DARMA instances in parallel, the new features add the possibility to access multiple databases from one DARMA server.
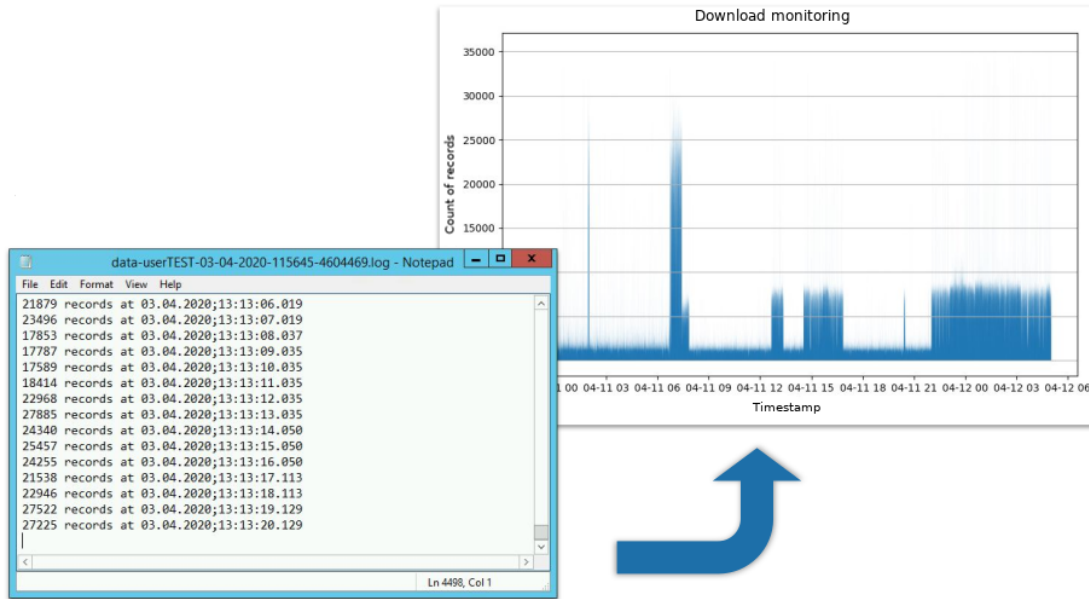
Fig. 4 Download monitoring

4.2. Download monitoring

We developed a new supporting module for download monitoring in order to simplify the maintenance for developers of the system. The new module stores an actual count of downloaded records in a given time frame. These records simplify especially searching for failure sources and increase transparency of the download process. Development of the module was initiated by a specific failure in the downloading process accruing within TRD archive.

The module regularly stores logs into a temporary file ending with a file extension .log. The length of the time frame can be defined in application configuration reffed with Web.config file. The log file can be downloaded by owner of the download request available in the Detail section. Figure 4 shows time dependence of downloaded records according to a log file.
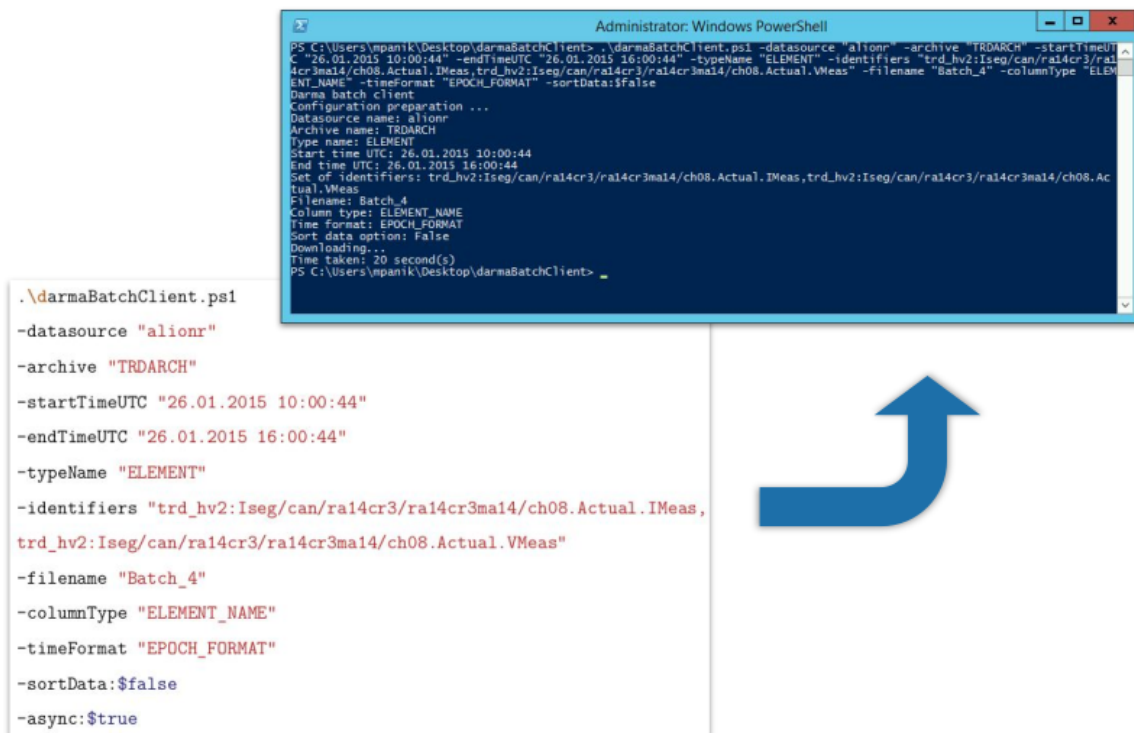

Fig. 5 Console Batch Client

4.3. Console Batch Client

The base functionalities of the DARMA system are available without necessity to use graphical user interface. Console Batch Client allows to fully cooperate with the server side of DARMA system to access remote archival data. Implementation of the Console Client is based on a MS PowerShell script and server API. The module provides a batch downloading based on client requests.

Batch client requires to specify a similar set of options to the base web interface of the DARMA system. Batch client can be also triggered by other processes in a auto-mated scenario. The module is useful when the client de-sires to run a lot of download requests according to a speci-fied set of parameters. An example of using this batch client is shown in Figure 5.
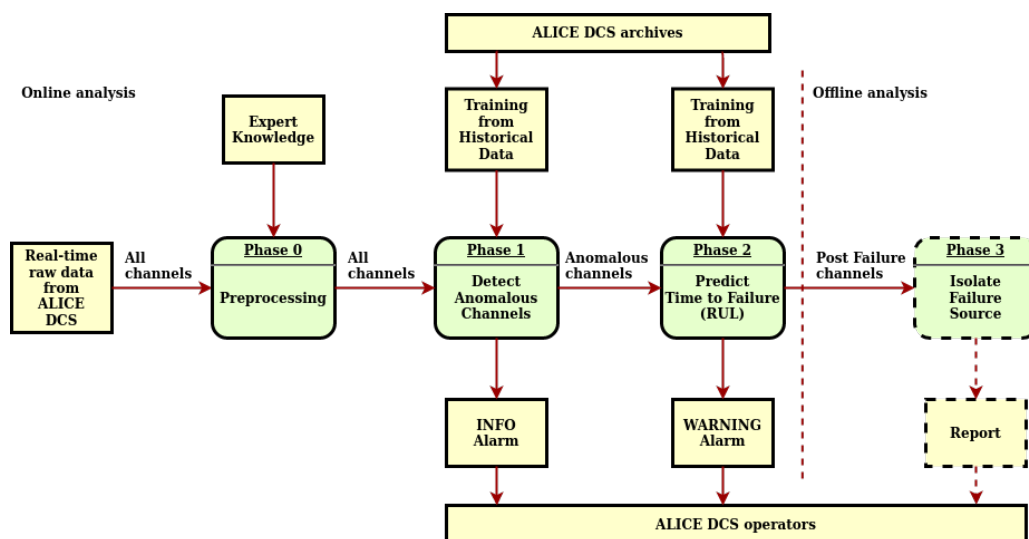

Fig. 6 Detection diagram

5. FAILURE DETECTION

Archival data obtained from DARMA system are used for further analysis. In this part we focus on design and im-plementation of a failure detection strategy in the context of high voltage channels of TRD detector of ALICE exper-iment [6].
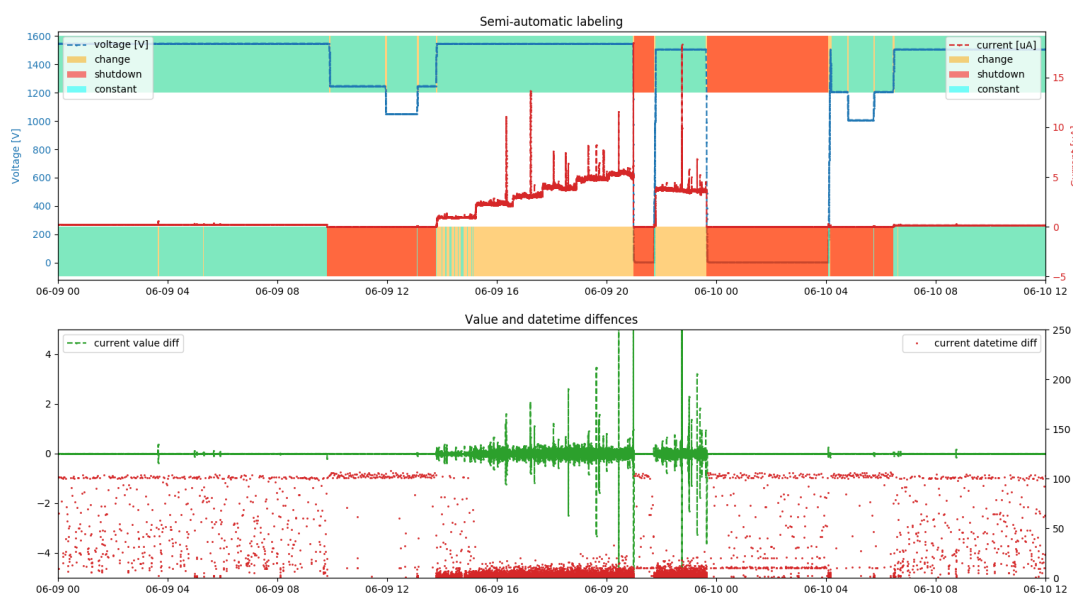

Fig. 7 Semi-labeling

5.1. Detection strategy

The main task here is to identify the conditions when a so called High Voltage trip (HV) occurs as a consequence of a hidden failure source. HV trip is a state when electrical voltage and current drop down immediately as a result of intervention of a set of safe fuses. A diagram in the Figure 6 describes the designed detection strategy consisting of preprocessing phase, anomaly detection phase, remaining-useful-life (RUL) prediction phase and offline failure source isolation phase. We used DARMA batch client to obtain the necessary sets of samples to build a training model. Proposed strategy was consulted with experts dealing with similar data from cryogenic subsystems of ATLAS experiment [7].

5.2. Implementation

The data we work with are quite challenging, since they can be characterized as unevenly-spaced, non-stationary and multivariate time series data. This limits a set of possible methods to use in such scenario. Designed algorithms are supposed to run in real-time so we focus on simple statistical and machine learning techniques [8].

In the first phase we obtain the true labels from semi-labeling process based on expert knowledge and archival alert information. Figure 7 shows the result. Afterwards, we calculate the rolling mean and variance of electrical current channel samples, since these changes highly correlate with the known anomalous behaviour. Channels of the same sector are joined together. If some of the channels statistically vary from others, we mark these channels as anomalous. We use a principle of Pearson chi-squared test to consider channels as significantly different.

The next step is to predict time when HV trip might occur. We create a set of sequences and normalize them by its working regime of voltage channels. Residual similarity model is trained from historical sequences and used to test novel sequences. Firstly, the model tries to fit sequences by 2nd order polynomial fit and calculate 20-nearest neighbors. From this set of neighbors and their true RUL values we can estimate the RUL of the new sequence as a median of historical values.

Fig. 8 Evaluation of the prediction error

The resulting predictions can be evaluated by measuring of the prediction error. Figure 8 shows how the the variance of the error decreases in time as the prediction is made closer to the actual failure occurrence.

6. CONCLUSION

This paper summarizes the latest development of the DARMA system, as well as the development of important components for the analysis of archived data. As part of the solution, the internal process of downloading data was streamlined, thus eliminating the problem of interrupting the downloading process, as one of the shortcomings of the original design. In the next phase of development, the system was extended by two new database sources, ALICEITSDB and ALIDCSSURF, making new archives available. This has significantly increased the quality and potential of the DARMA system in the context of the ALICE DCS infrastructure. As part of testing the functionality of the designed and implemented solutions, it was successfully verified that the DARMA system is able to query tens of Gigabytes of archival data from hundreds of identifiers from 2009 to the present, per a single request, without interrupting the download process. The work, this paper refers to is documented in form of system and user manuals providing extensive documentation material of the DARMA system, which can be used for future researchers. Possible extensions of the DARMA system in the future might be simplification of the deployment process and the continuous testing of components using CI / CD methods.

This solution is in production at CERN and is used by detectors regularly. During the LHC shutdown many detectors perform aging studies, that require retrieval of historical values from the database. The data stored in offline repositories contains values recorded during the active data-taking, but periods when the detector chambers operated on nominal setting largely extend these period. As corresponding data is not present in offline archives, DARMA is the optimal way for finding and retrieving this additional information. As an example a TRD aging study requires analysis of high voltage applied to the chambers since 2008. The data was retrieved by the DARMA and its size exceeds 5TB.

Currently, the DARMA is in daily use by detector groups and feedback is being regularly implemented in consecutive releases. Nowadays, DARMA is the only possibility to access ALICE conditions data stored in Oracle databases from networks external to the internal DCS network.
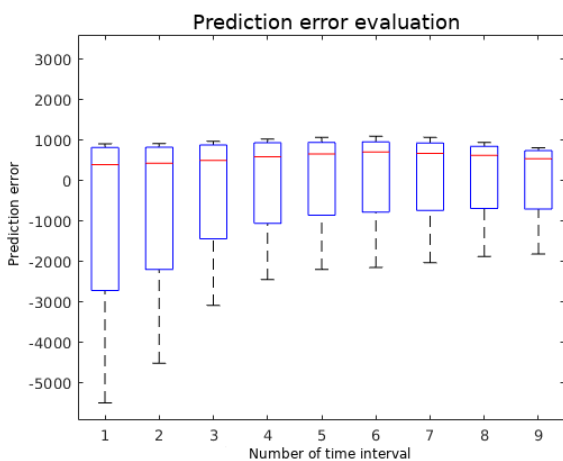
REFERENCES

[1] P. CHOCHULA – L. JIRDEN – A. AUGUSTI-NUS – G. DE CATALDO – C. TORCATO – P. ROSINSKY – L. WALLET – M. BOCCIOLO – L. CARDOSO: *The ALICE detector control system*, IEEE Transactions on Nuclear Science **57**, No. 2 (2010) p. 472-478. `https://ieeexplore.ieee.org/iel5/23/5446471/05446521.pdf`

[2] J. JADLOVSKÝ – S. JADLOVSKÁ – J. ČABALA and A. JADLOVSKÁ – J. ČERKALA – M. KOPČÍK – M. ORAVEC – D. VOŠČEK – M. TKÁČIK – E. HANC – P. CHOCHULA – P. BOND: *Information system for ALICE experiment data access*, ICALEPS 2017 , No. (2017) p. 1-4. `http://icalepcs2017.vrws.de/papers/thpha041.pdf`

[3] P. CHOCHULA – A. AUGUSTINUS – V. FEKETE – L. JIRDÉN – S. KAPUSTA – P. ROSINSKÝ: *Handling large data amounts in ALICE DCS*, Proc. ICALEPCS07, Knoxville, USA **591**, No. (2007) . `http://accelconf.web.cern.ch/AccelConf/ica07/PAPERS/RPPA36.PDF`

[4] J. JADLOVSKÝ – A. JADLOVSKÁ – J. SARNOVSKÝ – S. JAJČIŠIN and M. ČOPÍK – S. JADLOVSKÁ – P. PAPCUN – R. BIELEK – J. ČERKALA – M. KOPČÍK – P. CHOCHULA – A. AUGUSTINUS: *External access to ALICE controls conditions data*, Journal of Physics : Conference Series (JPCS) : 20th International Conference on Computing in High Energy and Nuclear Physics : CHEP 2013 , No. (2013) p. 1-5. `http://iopscience.iop.org/1742-6596/513/1/012015/pdf/1742-6596_513_1_012015.pdf`

[5] S. KAPUSTA – A. AUGUSTINUS – P. CHOCHULA – L. JIRDÉN – P. ROSINSKÝ: *Data Flow in ALICE Detector Control System*, Proc. of 11th Vienna Conf. on Instrumentation , No. (2007) .

[6] M. KWEON: *The Transition Radiation Detector for ALICE at LHC*, Nuclear Physics A , No. (2009) p. 1-4. `http://dx.doi.org/10.1016/j.nuclphysa.2009.10.047`

[7] F. TILARO – B. BRADU – M. GONZALES-BERGESX – M. ROSHCHIN – F. VARELA: *Model learning algorithms for anomaly detection in CERN control systems*, ICALEPS 2017 , No. (2018) p. 1-7. `https://cds.cern.ch/record/2305655`

[8] The MathWorks Inc: *Residual comparison-based similarity model for estimating remaining useful life*, The MathWorks Inc. , No. (2020) Online; cited 2-May-2020 `https://www.mathworks.com/help/predmaint/ref/residualsimilaritymodel.html`

BIOGRAPHIES

Ján Jadlovský was born on June 13, 1960. He graduated in the field of Technical Cybernetics at Electrotechnical faculty of Technical University of Košice with focus on robotics. In the workplace of the speech recognition he vindicated his dissertation work in 1990 and in 1994 he was habilitated for associate professor with his habilitation thesis dealing with Distributed Control Systems. Since 1994 he has worked at Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics (DCAI FEEI) TUKE as assistant professor. His teaching activities are focused on the issues of proposal and implementation of distributed systems for control of the production processes. His research activities are oriented towards distributed control systems, image recognition, complex functional diagnostics of the controllers, diagnostics of production control systems, design of the information and control systems with application of the latest information technology. Currently he is Team Leader for TUKE for participation of TUKE in the ALICE Collaboration at CERN. He is a chief executive of company KYBERNETIKA, s.r.o., Košice.

Milan Tkáčik is PhD student at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University of Košice, since year 2019. His PhD thesis is focused on development of methods and tools for design, modeling and realisation of Distributed Control Systems of Large Physical Experiments. He is focused on research and teaching activities in the fields of Distributed Control Systems and Mobile Robotics at the Center of Modern Control Techniques and Industrial Informatics. He is currently involved in development of the Detector Control System for the ALICE experiment at CERN, as a member of the research team of the project CERN at TUKE.

Slávka Jadlovská was born on June 25, 1988 in Košice, Slovakia. She graduated (MSc) in 2011 at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics, Technical University of Košice (DCAI FEEI TUKE). She defended her PhD thesis in the domain of Cybernetics in 2015 at the same University; her thesis title was Modeling and Optimal Control of Nonlinear Underactuated Dynamical Systems. Since 2015, she has worked at DCAI FEEI TUKE as Assistant Professor. Her main research activities include modeling and simulation of underactuated mechatronic systems, optimal and predictive control design for hybrid systems and the advanced application of MATLAB toolboxes in modeling and control of physical and robotic systems including the verification of simulation results on laboratory models. She was one of the establishing members of the participation of TUKE in the ALICE Collaboration at CERN, and is currently Deputy Team Leader for TUKE. She was awarded the Rector's prize for the best young scientist (under the age of 35) at TUKE in 2018. She is the author of scientific articles and contributions to various journals and international conference proceedings, and has co-authored a monograph.

Michal Pánik studied at the department of Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University of Košice. His diploma thesis (Implementation of an Information System for Accessing LHC Data within the ALICE-CERN Experiment) deals with the implementation of systems for remote data access in cooperation with the international laboratory CERN, including the implementation of new requested functionalities into the DARMA system. He defended his diploma thesis in the field of database information systems and data analysis in 2020.