Railway line data optimization and formal modelling and verification of generation method

Zihan Zhang1*, Zhongtian Liu2
1,2 School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, 100044, China
*Corresponding author’s e-mail: 16120282@bjtu.edu.cn

Abstract. The research on upgrading and modification of the existing railway signal system found that the existing line basic datasheet and line data engineering datasheet cannot satisfy the application requirement, and it is necessary to study data optimization on the existing railway line. This paper presents a method for railway line data optimization and generation process verification. Based on the analysis of railway line data and the research of data update requirements, the hierarchical block storage model of line data is established by expanding the engineering datasheet, and design the data model generation process. The formal model of generation process is established by the model verification method combining UML with NuSMV model, which verifies the activity, certainty and transference of the important attributes in the model. The data model and generating process are improved by counterexample analysis to meet the requirements of signal system. The discussion of this paper provides useful solutions for the key problems encountered in the research and development of the existing railway signal system.

1. Introduction
China Railway Corporation organized the research on upgrading the existing railway signal system, which mainly resolved security problems in the existing infrastructure, such as the inflexible application of line data, safety risks and other problems existing in the replacement.

The mainstream research direction aims to upgrade and reform the existing railway signal system, enabling lower the railway line data to the ground for centralized management, thus achieving wirelessly data update by the train. Considering the safety of railway line data maintenance and the efficiency of the trains’ wirelessly data update, it is found that the data engineering datasheet cannot meet the line data’s safety requirement and efficiency of replacements. Therefore, it is necessary to expand and optimize the existing railway line database.

At present, there are few researches on railway line data and mature solutions are lacking in the field of study. Literature [1] studied the search algorithm generating column control digital track map with signal graph transformations. Literature [2] used the topology unit-based modelling method to model typical railway structured data, which verified that the access data are capable of meeting the requirements of the train control system. Literature [3-4] focused on the existing safety concerns of the on-board data and puts forward suggestions for improvements. Literature [5] utilized the stochastic Petri net to model and verify the completeness of Chinese Train Control System(CTCS) 3 train control data. The above academic works analysed train control data from different perspectives. However, the current research mainly focuses on the data matching methods and the field of data security verification. The existing line data optimization only describes the reconstruction suggestions without
giving specific optimization scheme. Subsequently, this results in inadequacy of specific schemes regarding line data being transferred to the ground for storage.

This paper proposes a new railway line data optimization model and data generation method. Through the comprehensive analysis of the data of the existing line and the analysis and research of the CTCS-2/3 line data, the relationship between the data is obtained. Then, corresponding to the application requirements, the engineering data table is expanded to obtain the data model structure that meets the requirements, and the data model generation method is designed. Finally, the UML language and NuSMV language are combined to model and validate the data model generation method scene. This paper’s research results attempt to provide a new method for data optimization storage and update, which is of high relevance for ensuring safe train operations.

2. Railway line data analysis
Railway lines are a specific type of data with a variety of line data and intrinsic relationship between each corresponding component. In the optimization of the data model, in addition to optimizing the data structure, the original data attributes and associations are kept unchanged [7].

2.1. Data Requirements Analysis
According to the existing problems of the aforementioned railway line signal system and the safety of the line data maintenance[8], the line data update requirements are summarized into three parts: update requirements, storage requirements and maintenance requirements.

Update requirement. Due to that the line cannot be suspended for data replacement, the entire line is decomposed into several areas, and the entire line is updated in sub-areas.

Storage requirements. Data is stored in such a way that all data is independent of each other and the data that needs to be updated is independent of other device data. The functional data is separated from the device itself, as well as the spatial topology data, and stored independently.

Maintenance requirements. Each update of the data device iterates over a new version number, which updates the version change of the corresponding data element in the data center file list. Consequently, this keeps the entire data model up to date.

2.2. Line data analysis
Line data type. The line data is the basic component of the railway line data, line data can be divided into the following three types according to different data levels: railway line data, track section data, and signal equipment data.

Line signal data characteristics. The line signal data description correlates to a substantial number of subjects. In relation with each element, they could mainly be divided into two categories: line signal equipment data and line functional data.

Spatial topological relationship of line data. The spatial structure of line data can be abstracted into a topological model composed of points, lines and planes. Signal data is abstracted into points data, track segments and lines are abstracted into lines data. Subsequently, the aforementioned primary sources are organized into a point-line relationship.

Based on the above analysis, the types and logical structure of existing line data are shown in figure 1.

3. Data model optimization design
Data model is the systematic extraction and abstraction of real-world characteristics, through which a adequate simulated expression could be obtained to describe its real-world counterpart[9]. According to the analysis of railway data lines, the data model should contain two types of data: signal attribute data and spatial topological data. Signal attribute data mainly includes signal controller, turnout and other functional signal equipment; spatial topological data storage signal device data association and data organization between lines. At the same time, data updating and storage requirements should be
considered to design the two types of data separately in the data model. The data model generation framework is shown in figure 2.

Figure 1. Logical structure of existing line data. Figure 2. Data model generation framework.

The topological relationship of signal equipment in the track section is mainly that one track segment corresponds to multiple signal equipment. Data hierarchy and its branching connections are expressed through tree structure systems, with nodes representing each data elements. The tree structure system is widely regarded to be the most suitable and logical structure to store line basic data. The logical representation of the model is mainly aimed at the logical rules of the signal device hierarchy, the specific signal data are adjusted according to the line requirements and the number of devices.

The four-layer data structure is shown in figure 3. Specifically, the first layer is the data set, which stores all railway lines under the national network. The second layer is a data block, which stores all lines under the station and interval envelope. The third layer is the packet, the storage station and the track section inside the interval. The fourth layer is data element, which stores signal device data. Each data element represents a signal equipment that defines the attributes of the data element according to a specific signal function, and each attribute is defined as a data entry point.

Figure 3. Data model logical structure.

Data element is the smallest inseparable unit in a data model. Due to space limitation, only signal controller data element is introduced in this paper. The signal controllers are divided into different categories according to their purpose. This paper only considers three most commonly used signal controller for design, and the signal controller data element model is shown in Table 1.
Table 1. A signal controller data element model structure.

| name  | Stuc_ID | ID | mileage | type | insulation section | direction | special signal |
|-------|---------|----|---------|------|--------------------|-----------|----------------|
| definition | Packet_ID | Controller_ID | Kilometer | post | Type of signal | Insulation_ID | Line direction | Special signal or not |
| Data type | int | int | double | int | int | int | bool |

4. Modelling and verification of data generation scenarios

4.1. Introduction to the data generation process scenario

First, through reading the basic datasheet and the supplementary information in the signal graph of a railway station, a set of complete ground data is formed through integration. Then, the railway line is decoupled. By decoupling the railway line, as well as adding the mathematical logic formula of the relationship of the junction station in the metadata, the data in units of lines are discretized and transformed into independent railway lines. The discrete lines are decomposed into different signal data lines according to the signal type. Identifying stations and sections in the railway lines, they are then divided into data blocks in accordance to different stations and sections. Thus, defining corresponding data element models for each type of data.

4.2. UML model of data generation scenarios

A UML model of data generation process is established based on data model generation method mentioned above. The data generation process is described by three diagrams in the UML model[10], and the model is converted into a formal model for verification according to the UML-NuSMV extension method. Due to space limitation, only class diagram and state diagram is listed.

Class diagram. Class diagram describes the static structure of the system, abstracting all the objects involved in the system and related relationships. By identifying the participating objects in the scene, six main classes are abstracted and the more complex classes are refined into subclasses, which is shown in figure 4.

State diagram. State diagram is a control structure analysis of a system object, which is usually used to describe system states. The state diagram is a representation of the state changes of each object in the system under different conditions, as shown in figure 5.

4.3. UML - NuSMV model transformation

NuSMV is a symbolic model checking language, which verifies the model by enumerating finite states[11-12]. The NuSMV model consists of a main module and sub-modules, all of which contain the declaration (VAR) and assignment (ASSIGN). The UML model to NuSMV model transformation method[13] is as follows:

Step 1. Convert a class in UML to a MODULE in the NuSMV model with the same definition name as the class name.

Step 2. Extract the state sets of the main module and sub-modules from the state diagram and declare them in different modules respectively.
Step 3. Transform the relationship between classes into the actions between modules, and describe the corresponding actions through the jump of module state and the change of conditions.

Step 4. Describe the transition relationship in the state diagram through an assignment statement. The first state in the state diagram is the initial state of the model, and the state triggered by conditional excitation is the next state.

Step 5. Use CTL and LTL to describe properties to be verified according to data requirements and state models.

4.4. NuSMV model of data generation scenarios
Based on the transformation method of UML model and NuSMV model, the above UML model is transformed into NuSMV model. Convert the static class established in UML model into main module and sub-module, and the code fragment of the main module as shown in figure 6:

![Main module code fragment of NuSMV model.](image)

Based on the symbolic model verification method, the data model generation process and the data requirement specification are compiled into the verification statement of the NuSMV model[14]. Through the different description of compute tree logic(CTL), three aspects of the model, the main realization function, trigger path and process state, are verified. Specific transformational grammar should be referenced to manuals.

4.5. Verification results and counter-example analysis
The NuSMV model of the line data model generation method is established, and requirement statement could be inputted for verification. If a counter-example(false) occurs in the verification, the part where the model is not valid according to the counter-example given by the model should be located, then analyzed for the error.

When a counterexample occurs, firstly, confirm that the formal verification model and the attributes to be verified are correct. Secondly, check the UML and NuSMV model transformation between whether there is a mistake. Finally, check if there is a problem with the UML model of the data generation method, and confirm that the data generation method is correct if all the above models are correctly expressed. Repeat the above steps until the problem is confirmed. Subsequently, modify the model, and finally make all the attributes of the model pass the verification to get the line data generation method that meets the requirements. The verification results of the main properties in the data model generation scenario are shown in figure 7.
5. Conclusion
This paper proposes a data model of optimizing the structure and data model of generating method. Through the study of the existing line datasheet based on the analysis of the data requirements and characteristics, the line data model is generated according to the optimization of user requirements, and the method of generating the model from the existing railway basic data to the optimized model is proposed. Then, a verification model of the data model generation scenario is constructed by combining the UML model and the NuSMV model. The correctness of data model design and generation is verified by the NuSMV model which can be automatically verified. The results show that the research results of this paper provide a reliable storage model for the optimization of the existing line data.

Acknowledgments
This work is supported by the Fundamental Research Funds for the Central Universities 2017JBM009.

References
[1] Wang H, Schmid F, Chen L, et al. (2013) A Topology-Based Model for Railway Train Control Systems[J]. IEEE Transactions on Intelligent Transportation Systems, 14(2):819-827.
[2] Zhou G, Zhao H. (2016) Data Safety Verification of Computer Interlocking in Urban Railway Signaling [J]. Journal of the China Railway Society, 38(8):63-72
[3] Wang Z, Wang M. (2011) Study of Optimum Structure of Basal Track Data Applied on LKJ Train Running Monitor Device [J]. Electric Drive for Locomotives, 62-64.
[4] Lu Z. (2016) Research on the Function Integration Optimization of Electric LKJ Vehicle Equipment[J]. Railway signalling & communication, 52(3):25-27.
[5] Cheng Y, Wang J. (2012) Research on modelling method for data completeness of train control[J]. Railway Computer Application, 21(7):12-15.
[6] Mo Z. (2016) Research on the overall technical scheme of CTCS-1 train control system [J]. China Railway, 37-43.
[7] Xiao L, Zhong E, Liu J. (2001) Spatial Conceptual Data Model of GIS [J]. Geomatics and Information Science of Wuhan University, 26(5):387-392.
[8] Yang Z. (2008) Train operation monitoring device (LKJ) data file preparation specification. China Railway Publishing House, Beijing.
[9] James Rumbaugh, Ivar Jacobson, Grady Booch. (2001) The Unified Modelling Language User Guide. China Machine Press, Beijing.
[10] Quelle J P, Sifakis J. (1982) Specification and verification of concurrent systems in CESAR. International Symposium on Programming pp 337-351
[11] Koh K Y, Seong P H. (2009) SMV model-based safety analysis of software requirements [J]. Reliability Engineering & System Safety, 94(2):320-331.

[12] Zhou Y P, Zhao L, Liu J T. (2014) Automatic Conversion Method of the Requirements Specification Verification for CTCS Level 3 [J]. Applied Mechanics & Materials, 563:248-257.

[13] McMillan, Kenneth L. (1993) Symbolic Model Checking. Springer US, New York.

[14] Alessandro C, Fausto G. (2000) NUSMV: a new symbolic model checker [J]. International Journal on Software Tools for Technology Transfer, 2(4):410-425.