Verification and generation of geographical data using a domain theory (revised extended abstract)

Lars-Henrik Eriksson
lhe@it.uu.se

Department of Information Technology*
Uppsala University
Box 337
SE-751 05 UPPSALA, Sweden

Abstract. Verification and generation of interlocking geographical data using a domain theory for railway signalling is described. Examples are taken from the methodology used industrially by Industrilogik L4i AB.

Railway interlockings form a family of systems where the individual systems have identical functions on an abstract level, as they implement general signalling principles. On the concrete level, differences in function between different interlockings is determined by the particular physical layout and other properties – both abstract and concrete (such as the maximum speed permitted through particular points) – of the track system controlled by the interlocking. A formal description of these properties is called the *geographical data* of the particular interlocking. (This sense of geographical data is similar, but not indentical, to the one used in work on formal verification of geographical data of the british SSI interlockings [4] [5].)

Using geographical data, generic requirements specifications that describe general signalling principles can be specialised to give a requirements specification for a particular interlocking installation. Similarly, interlockings can be implemented using generic modules (either in software or hardware) which are configured using geographical data to give a specialised implementation for a particular site. An example of interlockings working using this principle are Bombardier Transportation EBILOCK family of interlockings.

Given that the precise requirements of a generic specification, as well as the precise behaviour of a generic interlocking, are critically dependent on the geographical data, the correctness of the geographical data is of primary importance. Some kinds of geographical data – let us call them "primary" geographical data – are direct descriptions of the physical track structure and its concrete properties. Clearly, this data can not be formally verified, but its internal consistency

* The work presented herein was done while the author was employed by Industrilogik L4i AB, Box 3470, SE-103 69 STOCKHOLM, Sweden. I wish to thank my former colleagues for their involvement in this work.

– e.g. that it describes a physically possible track system – can be checked using a domain theory for rail systems.

Other kinds of geographical data – let us call them "secondary" geographical data – are data that are wholly or in part determined by the primary geographical data. One example is the description of all possible routes through the track system – a route typically being defined as a path through the track system on which a train could run, beginning and ending at a signal. Another example is the various kinds of protection areas required around a route to prevent possible collision with trains or vehicles close to the route. The construction and verification of secondary geographical data is of critical importance to the safety of the interlocking, while being one of the most time-consuming and error prone tasks in the interlocking design process.

Given a sufficiently complete domain theory and generic requirements specification, secondary geographical data can be formally verified or automatically generated given a set of primary geographical data. In this presentation, I will illustrate how this is done in the formal specification and verification methodology used for industrial projects by Industrilogik L4i AB (e.g. [1][2][3]). The sample domain theory axioms are adapted from generic formal specifications developed by Industrilogik for Swedish and Norwegian railway signalling.

The track system is represented as a set of "units", a unit being a set of points, a linear piece of track, a buffer stop, crossing, etc. A relation *connectsTo* describes which units are adjacent to each other. The predicate *points* is true of units that are points. For every set of points, the relations $leftBranch$ and $rightBranch$ describe what units are reached from the facing points, taking the left or right direction, respectively. There is also a set of signals. Every signal is assumed to be located at the boundary between two units. Relations *ahead* and $inRear$ describes the location and direction of a signal by giving the unit ahead of the signal (the unit the signal is facing) and the unit in rear of the signal. Fragments of a domain theory for the track system is given by the following predicate logic formulae:

1 $\forall u1, u2 \in UNITS\ (connectsTo(u1, u2) \rightarrow connectsTo(u2, u1))$
2 $\forall u \in UNITS\ \neg connectsTo(u, u)$
3 $\forall w, u \in UNITS\ (points(w) \wedge rightBranch(u, w) \rightarrow connectsTo(u, w))$
4 $\forall w \in UNITS\ (points(w) \rightarrow \exists u1, u2, u3 \in UNITS\ (connectsTo(w, u1) \wedge connectsTo(w, u2) \wedge connectsTo(w, u3) \wedge u1 \neq u2 \wedge u1 \neq u3 \wedge u2 \neq u3 \wedge \forall u4 \in UNITS\ (connectsTo(w, u4) \rightarrow u1 = u4 \vee u2 = u4 \vee u3 = u4)))$
5 $\forall s \in SIGNALS\ \exists u \in UNITS\ (ahead(s, u) \wedge \forall u1 \in UNITS(ahead(s, u1) \rightarrow u = u1))$

Formulae (1) and (2) state that the *connectsTo* relation is symmetric and irreflexive. Formula (3) states that the unit reached by going right through facing points must be adjacent to the points. Formula (4) states that a set of points is adjacent to exactly three different units. Formula (5) states that a signal is ahead of exactly one unit.

A particular set of primary geographical data determines a logical interpretation of the predicates and sets. Since the sets will be finite, it is possible to

directly compute the truth value of each of these axioms. If the data is consistent, the interpretation will be a model, i.e. every axiom will compute to true.

Now, consider routes as pieces of secondary geographical data. Routes are principally sets of units. To avoid having to quantify over sets, every route is represented by an identifier in the set $ROUTES$, while the relation $partOf$ relates each unit to identifiers of any routes it is part of. The direction of a route is determined using the relation $before$ which relates a route identifier to the unit immediately preceding the route. The defined predicate $first$ characterises the first unit of a route. Fragments of the theory for routes is given by the formulae:

6 $\forall r \in ROUTES\, \exists u \in UNITS\, (before(r,u) \wedge \forall u1 \in UNITS\, (before(r,u1) \rightarrow u = u1))$

7 $\forall r \in ROUTES\ \forall u \in UNITS\ (before(r,u) \rightarrow \neg partOf(r,u) \wedge \exists u1 \in UNITS\ (partOf(r,u1) \wedge connectsTo(u,u1)))$

8 $first(r,u) \equiv partOf(r,u) \wedge \forall u1 \in UNITS\, (before(r,u1) \rightarrow connectsTo(u,u1))$

9 $\forall r \in ROUTES\, \exists s \in SIGNALS\, (\forall u \in UNITS\, (ahead(s,u) \rightarrow before(r,u)) \wedge \forall u \in UNITS\ (inRear(s,u) \rightarrow first(r,u)))$

10 $\forall r1, r2 \in ROUTES\, (conflict(r1,r2) \leftrightarrow r1 \neq r2 \wedge \exists u \in UNITS\, (partOf(u,r1) \wedge partOf(u,r2))$

Formula (6) states that there is exactly one unit located before each route, while (7) states that this unit is in fact adjacent to the first unit of the route while not being part of the route itself. Formula (8) defines the auxiliary predicate $first$. Formula (9) states that there must be a signal at the beginning of the route, facing the unit before the route. Formula (10) states that two routes are in conflict if they have some unit in common.

The secondary data can be verified in the same manner as the primary data. However it is also possible to automatically generate the secondary data. Primary data gives a "partial interpretation" of the domain axioms where secondary data predicates are undetermined. Since the sets are finite, this essentially creates a propositional satisfiability problem which can be solved using a SAT solver. The SAT solver would generate truth assignments to the secondary data predicates, effectively creating correct secondary geographical data.

A problem is that the number of routes is not known in advance, while the number of elements of the set $ROUTES$ must be known in order to create a SAT problem. One possibility is making a conservative estimate of the maximum number of possible routes. Another one is to include only one route, but generate the complete set of routes by finding successive solutions to the SAT problem. The latter approach is implemented in the SST/SVT formal methods toolset used by Bombardier Transportation for interlocking software development.

These techniques presuppose the existence of a complete domain theory for railway track systems and signalling, which shows that such a theory has a concrete practical use.

References

1. Eriksson, L-H.: Using Formal Methods in a Retrospective Safety Case, In Heisel, M., Liggesmeyer, P., Wittmann, S. (eds.): Computer Safety, Reliability, and Security – 23rd International Conference, SAFECOMP 2004, Springer Lecture Notes in Computer Science 3219, Springer-Verlag (2004)
2. Eriksson, L-H.: Specifying Railway Interlocking Requirements for Practical Use, In Schoitsch, E. (ed.): Proceedings of the 15th International Conference on Computer Safety, Reliability and Security (SAFECOMP'96), Springer-Verlag (1996)
3. Eriksson, L-H. and Johansson, K.: Using formal methods for quality assurance of interlocking systems, In Mellit, B. et.al. (eds.): Computers in Railways IV, Computational Mechanics publications (1998).
4. Morley, M. J., Safety Assurance in Interlocking Design, Ph.D. thesis, University of Edinburgh (1996).
5. Simpson, A. C., Woodcock, J. C. P., and Davies J. W.: The mechanical verification of Solid State Interlocking geographic data. In Groves, L. and Reeves, S. (eds.), Proceedings of Formal Methods Pacific, Wellington, New Zealand, 9–11 July, pages 223–242. Springer-Verlag (1997).