Stochastic prediction of train delays in real-time using Bayesian networks

P. Kecman [1]   F. Corman [2]   A. Peterson [1]   M. Joborn [1]

[1]Department of Science and Technology, Linköping University

[2]Section Transport Engineering & Logistics, Maritime and Transport Technology, Delft University of Technology

21 July 2015

Linköping University

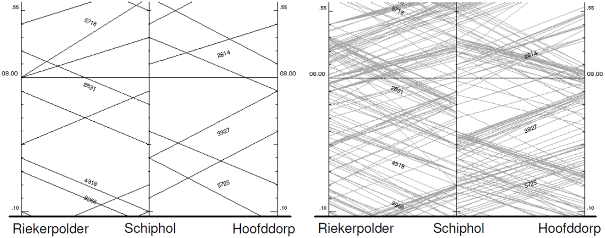
Outline

Outline

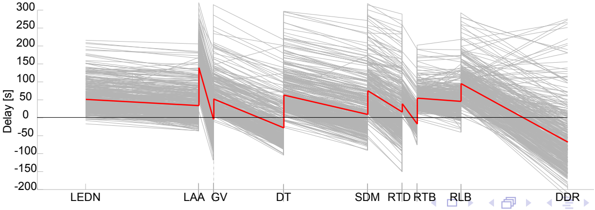
1. Introduction
2. Stochastic model based on Bayesian networks
3. Computational experiments
4. Current research
5. Conclusions

Uncertainty in railway traffic

Railway traffic typically operates according to a timetable, however...



Source: D'Ariano, PhD thesis



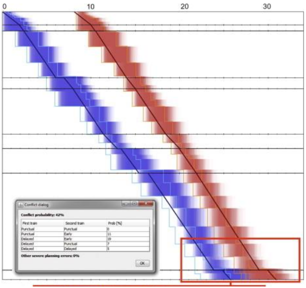
Importance of tackling uncertainty

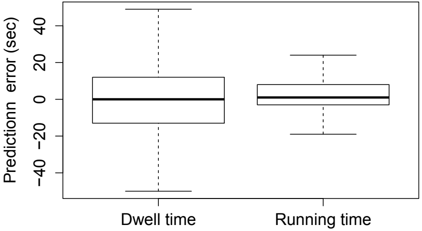
Quality of railway service depends on accurate predictions of future train movements on many levels:

- Traffic can be controlled pro-actively by taking actions that prevent delays and delay propagation
- Timetable, passenger transfer plans, rolling-stock and crew circulation plans can be kept up-to-date
- Passengers can be provided with accurate information: in-vehicle, on-platform, online


Existing approaches

- Deterministic models require frequent updates of train positions and detailed data
- Stochastic models are based on fixed distributions and do not exploit information available in real time





Source: Medeossi et al (2011) JRTPM

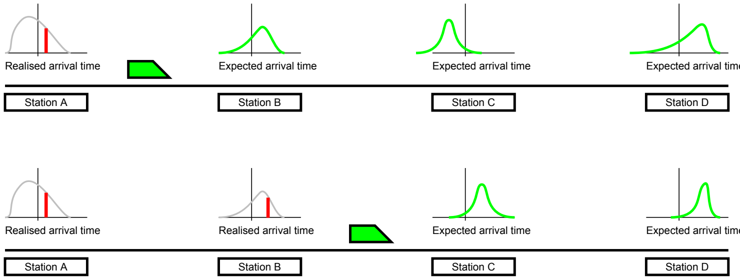
Research objective

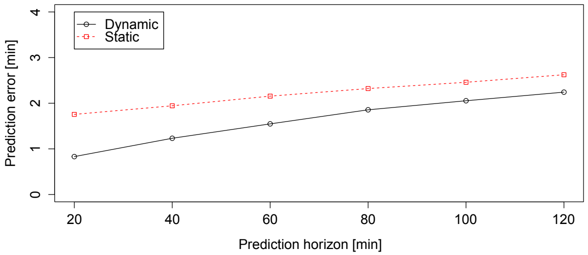
- How can incoming real-time information stream be used to reduce uncertainty?
- Develop a model that captures dynamics of uncertainty over time and gives predictions that are:
  - Accurate and reliable
  - Dynamic and responsive
  - Stable over longer prediction horizons

THE AUTHOR OF THE WINDOWS FILE COPY DIALOG VISITS SOME FRIENDS.

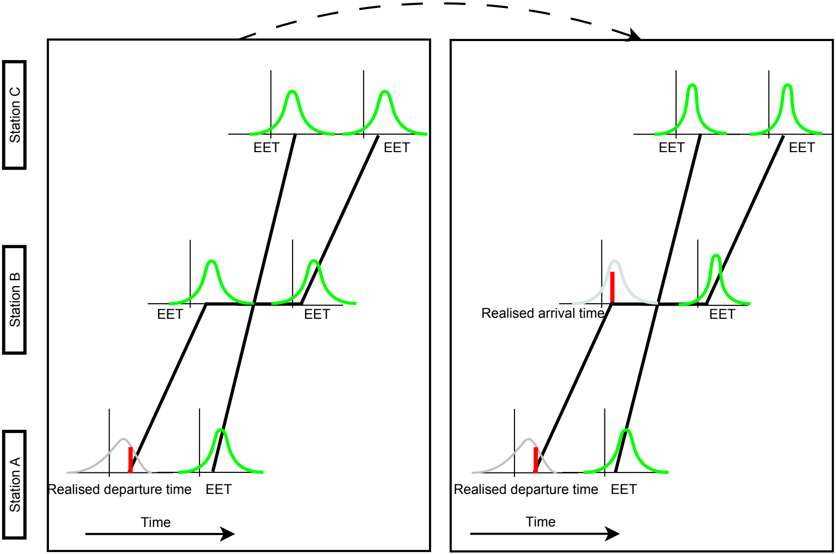
Source: xkcd.com/612

The first step - train delay evolution as a Markov process





The second step - stochastic delay propagation

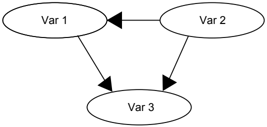


Outline

Bayesian networks
Definition

- Bayesian networks are graphical models for reasoning under uncertainty
- Variables and the conditional dependencies between them are represented with a directed acyclic graph $G = (N, A)$
- Direction of an arc reflects the causality relationship between two variables (nodes)
- Conditional probability distributions are computed for each node

	Var 1	
Var 2	T	F
F	0.4	0.6
T	0.01	0.99



Var 2	
T	F
0.2	0.8

		Var 3	
Var 1	Var 2	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

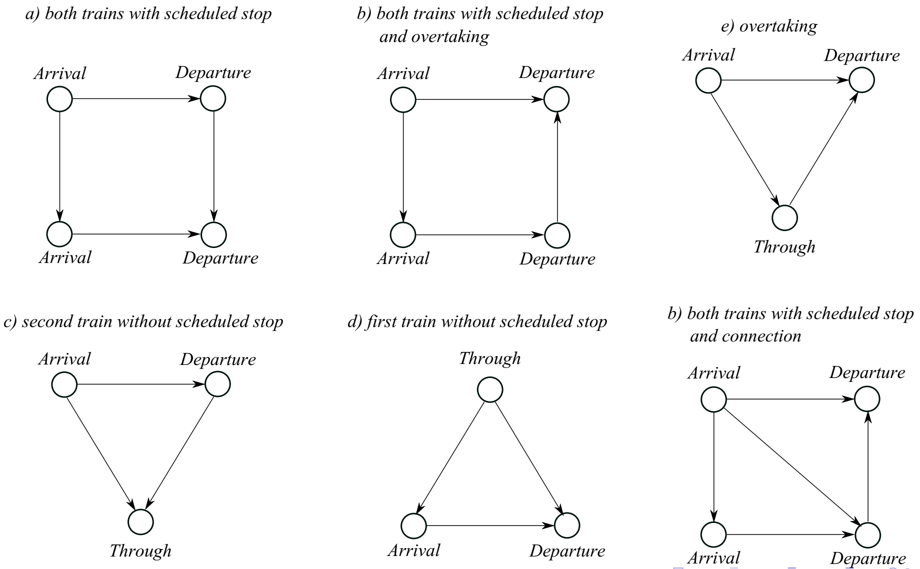
Properties

- Compact representation of a joint probability distribution
- The structure of the network, i.e., the directed arcs between the nodes and conditional probability distributions, can either be learned from the data or determined by expert knowledge
- Markov property
- Inference - Probabilistic beliefs updated automatically as new information about a variable becomes available
  - Conditional probability updates - marginal posterior probability
  - Maximum posterior density - configuration q of the variables in Q that has the highest posterior probability

Inference

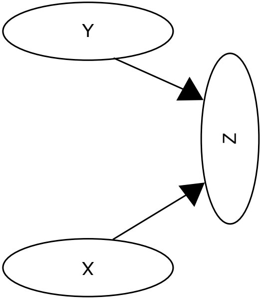
P(future traffic state | current traffic state)

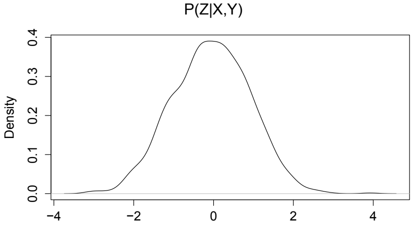
Structure of Bayesian network model



Parameters learning and inference

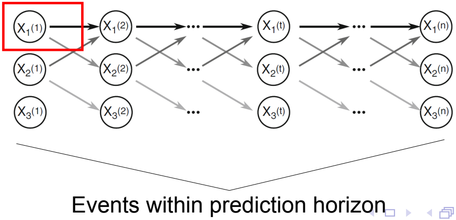
- Network structure implication - a local distribution has only a small number parameters to estimate
- Local distributions are univariate normal random variables
- Local distributions are conditioned on incoming events
- Local distributions are in fact linear models in which the parents play the role of explanatory variables





Implementation setup

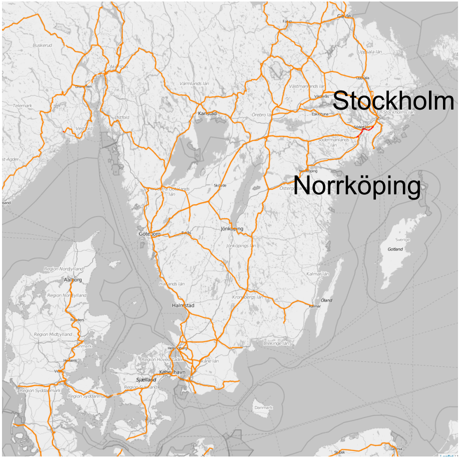
- Network structure is built based on the route plan and schedule assumed to be known during the prediction horizon
- The structure of the network is determined assuming that the train routes and orders are known
- Parameters of the network computed from the historical database (training set)
- Every train event message represents the new information that is propagated through the graph using statistical inference algorithms
- The prediction horizon moves, new trains are added to the model



Outline

Case study

- Case study from a busy corridor between Stockholm and Norrköping in Sweden
- 180 km long double-track line with mixed traffic
- 27 stations and junctions, 10 of which accommodate scheduled stops of passenger and freight trains
- Approximately 300 hundred trains per day traverse the corridor (fully or partially).



Data set description

- The database contains the scheduled and realised times for departures, arrivals and through runs for all trains and stations
- All event times are rounded to full minutes
- Train position and delay update given only at stations with a frequency of 2 minutes on average
- Two months (1 January to 28 February 2015) of historical traffic realisation data has been made available by the Swedish infrastructure manager Trafikverket.
- Randomly selected weekday from the analysis used as a test day to evaluate the performance of the Bayesian network model (test set)
- Remaining data are used for calibration of network parameters (training set)

Experimental setup

- Prediction horizon is 60 minutes
- A real-time environment for model validation was created by sweeping through the chronologically sorted messages in the test set
- Every train event message represents the new information received by the monitoring system
- The graph size does not increase linearly in time as only the nodes representing the last events of all trains are kept in the model due to the inherent Markov property of Bayesian networks

Model structure, size and parameters

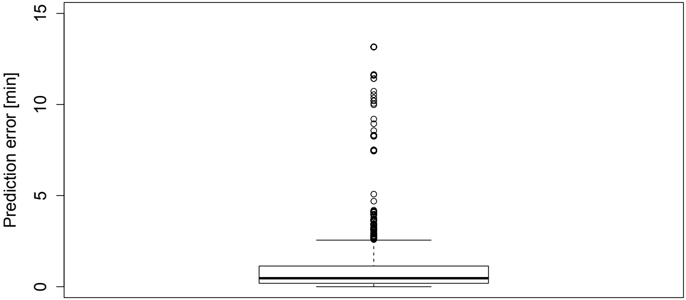
- The average network size is 137,12 nodes and 262.43 arcs

Arc type	Coefficient	RSE	p value	R^2
Headway	0,42	0,33	***	0,52
Running	0,95	0,06	***	0,90
Dwell	0,81	0,09	***	0,79

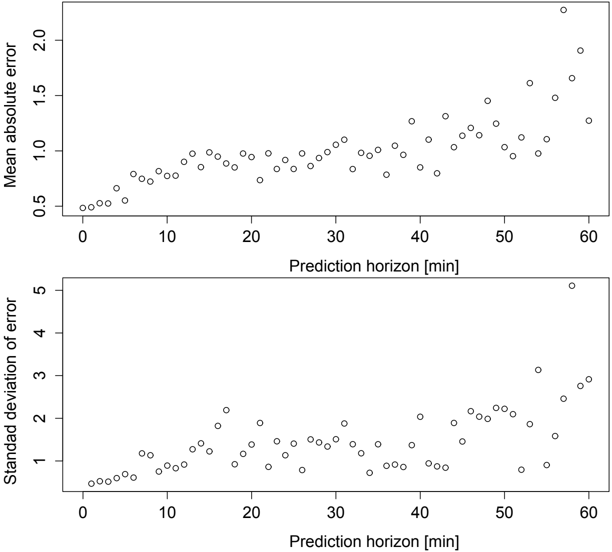
Table: Average values of predictive power of local models

Model performance on the test set

- Model applied on the peak hours (6:30-9:00 and 16:30-19:00) of the test day
- In total the inference algorithm is executed 563 times
- The predicted values are compared against the realised event times

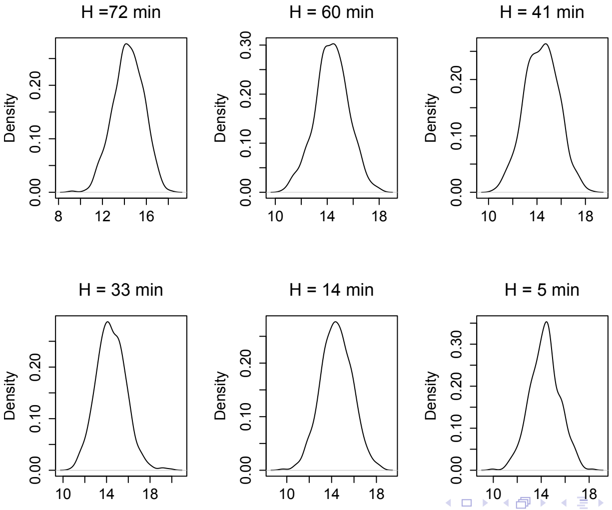


Model performance on the test set



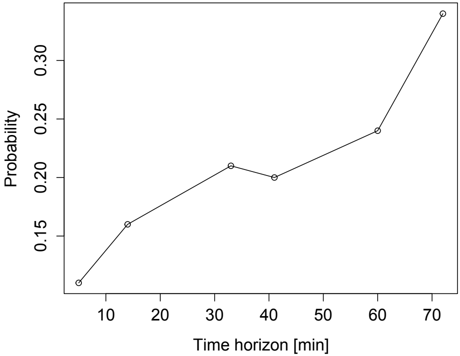
Dynamic updates of the marginal distribution

An example of how the distribution of arrival time of a train to the final station evolves over time in six discrete steps



Dynamic updates of probability

An example of how the probability that the arrival time will be larger than 16 min changes in six discrete steps

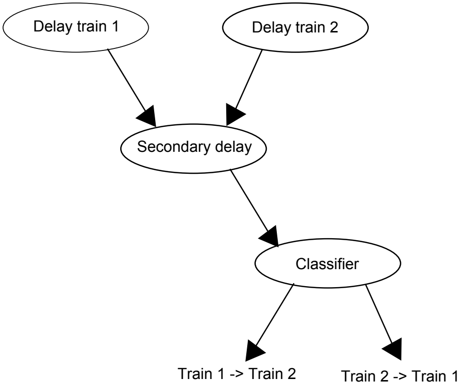


Outline

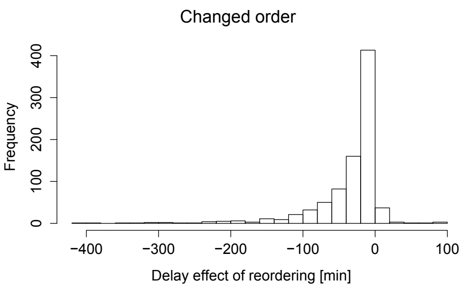
Stochastic modelling of traffic control decisions

Bayesian classification approach

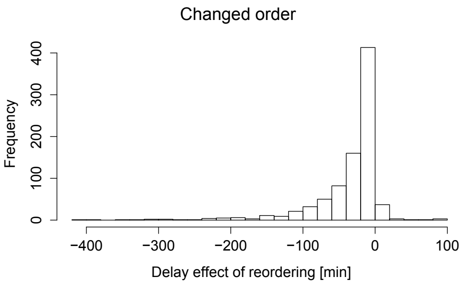
- How to overcome the assumption that traffic train sequences (orders) are known?

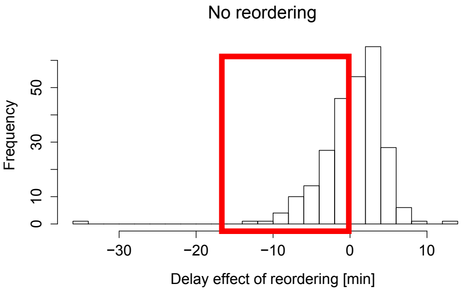


- Analysis of traffic control actions when secondary delays occur
- One point of conflict analysed
- 1110 delays that may cause secondary delays considered
- 77 % of cases resulted in reordering

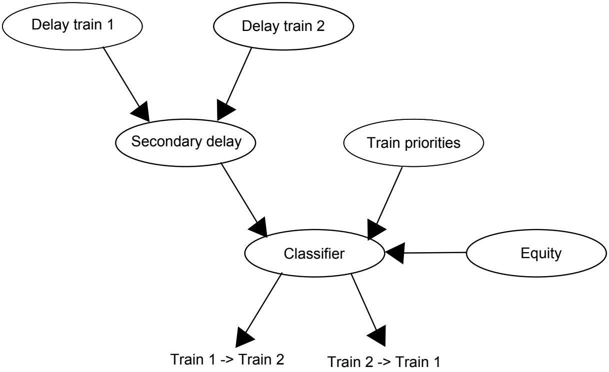


- Analysis of traffic control actions when secondary delays occur
- One point of conflict analysed
- 1110 delays that may cause secondary delays considered
- 77 % of cases resulted in reordering





- How to overcome the assumption that traffic train sequences (orders) are known?



Outline

Summary and conclusions

- Analysis of train delays and their evolution in real time with respect to the dynamic stochastic phenomena
- The goal was to determine the impact of real-time information on reducing uncertainty
- Bayesian networks turned out to be an appropriate method to concisely represent the complex interdependencies between train events
- The model was evaluated in a simulated real time environment and the computational results indicate that the predictions are reliable for horizons of up to 30 minutes
- Potential applications include integration in robust rescheduling framework, online traffic control an delay management models and passenger information systems
- A strong assumption of the model is that there is a perfect knowledge of train orders and routes within the prediction horizon - overcoming this will be in the focus of future research in this direction

Thank you for your attention