Pattern-based short-term prediction of urban congestion propagation and automatic response

Friedrich Maier, ***Robert Braun***, ***Fritz Busch***, *Technische Universität München*,
Paul Mathias, *Siemens AG, I&S ITS.*

This paper presents a method for the online prediction of urban congestion patterns including their spatio-temporal propagation based on historic traffic state data. Traffic state data for each link and time interval within the Berlin street network comes from a dynamic route choice and traffic assignment model.
From extensive historic traffic state data, congestion patterns are generated and classified in an appropriate manner.
Based on this analysis, a method was developed to predict the propagation of congestion within the network based on pattern recognition.
Significant parts of the network-wide prognosis are selected and sent as messages to the operator of the traffic management centre. A further step identifies actuators at in- and outflow areas of current and predicted congestion in order to increase the outflow from and decrease the inflow to the congested area. Messages for variable message signs are generated automatically and displayed to the operator with other appropriate measures.
The work presented was carried out within the German research project *iQ mobility*, which was funded by the initiative *Verkehrsmanagement 2010* (Traffic Management 2010).

1. INTRODUCTION

Traffic congestion in urban and rural areas has negative impacts on economic efficiency, the environment, traffic flow and traffic safety.

As well as knowledge of the actual network-wide traffic state, a quick and significant spatio-temporal prediction is a precondition for focused and timely initiation of suitable traffic management measures by an operator or an automatic procedure in order to reduce the negative effects.

Numerous approaches for congestion detection exist for applications on freeways, using traffic flow theories and therewith the relationship between macroscopic traffic parameters such as traffic volume, traffic density and time mean speed displayed by several versions of the fundamental diagram. Section-related approaches are partly adjusted and extended for sections with traffic signal control.

Urban networks are characterised by short links and by junctions that are frequently signal-controlled and are therefore especially challenging due to the resulting complex and dynamic traffic patterns. Local network properties do have an important influence on congestion propagation.

One objective of the research project *iQ mobility* was to develop a tool to predict the propagation of congestion within an urban street network. To achieve this goal a statistical approach was selected that was based on a history of temporally-discrete traffic parameters for all links of a network. Figure 1 illustrates a method overview.

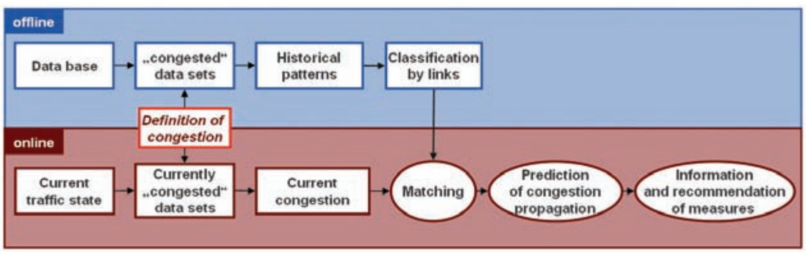
A large set of data collected by the traffic management centre in Berlin (VMZ Berlin) was made available. It consists of a road network with about 9000 links with detected traffic data, modelled traffic state data and traffic messages.

The model MONET/VISUM-Online is used in Berlin. It is based on the method of traffic volume propagation by VORTISCH (2005). It describes the traffic state for each link in each time interval (15 min). Calculated parameters are traffic volume, travel time, space mean speed and a three-step level of service.

The research project *iQ mobility* was funded by the German Federal Ministry for Economics and Technology within the initiative *Verkehrsmanagement 2010* (traffic management 2010). The project started in November 2004.

iQ mobility's main objective was to merge multiple existing single solutions into a multi-modal quality management system. Quality indicators were to be measured, calculated, eval-

Figure 1: Method overview



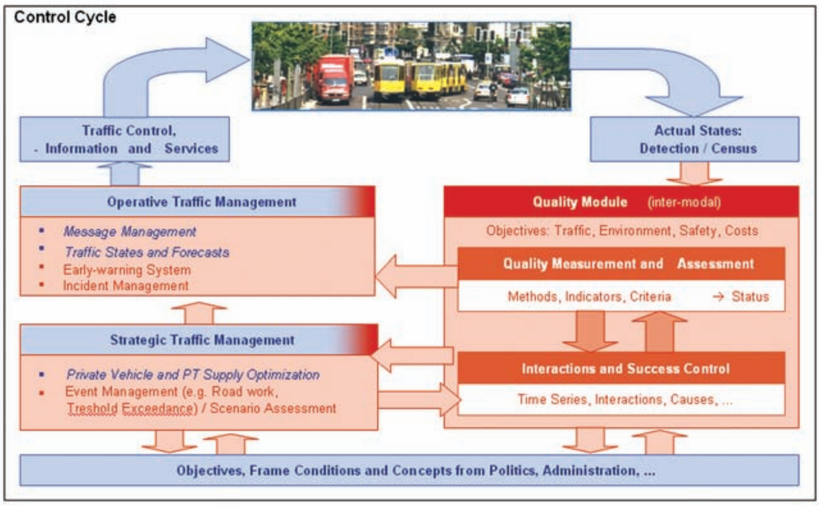


Figure 2: The iQ mobility-process of quality management for road traffic

uated and analysed in different target areas in the project. They were to be integrated in traffic control as well. Central target areas were the traffic situation on roads, the traffic related environmental pollution, traffic safety and costs (Projektteam iQ mobility (2007)).

Figure 2 illustrates the process of quality management with the multi-modal quality module and the strategic and operative traffic management.

The presented methods are components of the operative traffic management. They were tested at the *iQ mobility* test-site in Berlin.

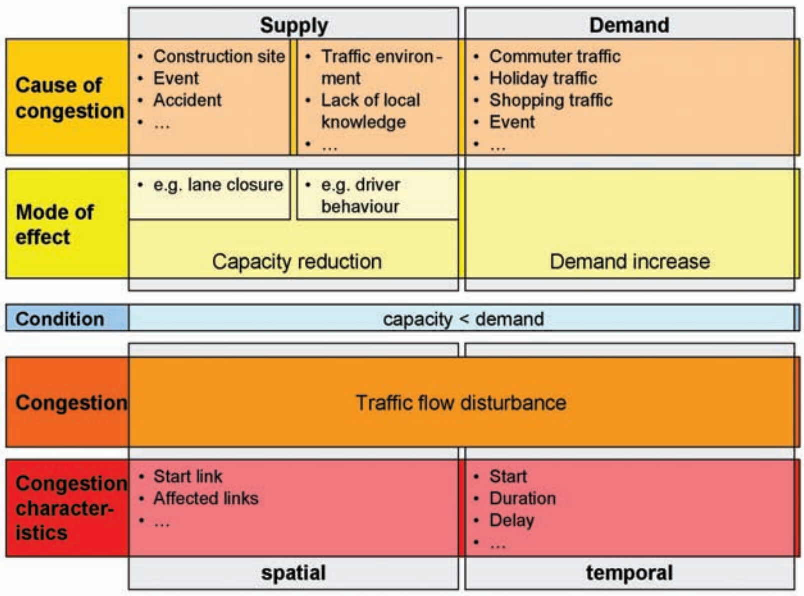
2. TERMS AND DEFINITIONS

Figure 3 gives an overview and the context of relevant terms concerning congestion.

Congestion causes with their *mode of effect* lead to *congestion* on a section of the network where the traffic demand exceeds the capacity.

The cause of congestion can arise either on the supply or on the demand side. We distinguish predictable and unpredictable causes of congestion. Predictable causes on the capacity side are construction sites and events on the road as well as, to a certain extent, road weather conditions. Accidents, however, are a typical example for unpredictable congestion causes.

Figure3: Context of used terms



Supply-side congestion causes reduce the capacity of a section of road. Congestion can be caused by restrictions in the existing infrastructure (eg reduction of lanes, narrowing of or turns on the carriageway, bumps or humps) or by changing driver behaviour. The impact, ie the change in capacity, is a quantifiable parameter. However, it often can only be estimated very roughly.

Demand-side causes of congestion are related to an increase in traffic volume on a section. Congestion causes on the demand side are usually regular occurrences like commuting, holiday or shopping traffic as well as scheduled events of all kinds.

Congestion needs to be described both spatially and temporally. The location of the congestion can vary along with the timeline.

Congestion appears if demand exceeds capacity. The problem with this definition is that the demand cannot be measured directly. Only the existing traffic volume can be measured at a determined cross section, where at most just the current capacity can be attained. Therefore, a different indicator must be used to identify congestion.

3. SPECIFICATION OF A CONGESTION INDICATOR AND DESCRIPTION OF CONGESTION PATTERNS

Reading out those datasets from the historic traffic state data, which describe time interval related *congested links* is the precondition to the determination of a *congestion indicator* and of an appropriate threshold. The readout data can be assembled into congestion patterns if the appropriate links are spatial and the time intervals are temporally adjoining.

Traffic congestion on motorways has already been investigated extensively. Many different definitions of congestion based on characteristic values can be found in the literature. Speed is most often used as a congestion indicator, as well as traffic density on occasion. In so doing the traffic state is often subdivided into three to five levels.

The following four parameters were analysed according to their suitability as congestion indicators for the urban context:

- The absolute difference of link-specific current space mean speed and free flow speed

$$v - v_0 [km/h]$$

- The quotient of link-specific actual space mean speed and free flow speed

$$\frac{v}{v_0}[-]$$

- Delay

$$\Delta T = \frac{s}{v} - \frac{s}{v_0}[s]$$

- Level of service [-]

v… link-specific actual space mean speed: MONET/VISUM-Online calculates the link- and interval-specific space mean speed using a function of resistance depending on traffic volume q.

v_0… link-specific free flow speed

s… link length

The following criteria were used as background for the selection of a congestion indicator and a threshold:

- Distribution of link lengths: it is assumed that the link length has no influence on its congestion sensitivity.

Evaluations using different congestion indicators could lead to different results for links with different lengths. With a good definition of congestion using characteristic values the length distribution of the links in the network is similar to the link length distribution resulting from the data sets that are categorised as describing congestion.

- Combination of congestion patterns and historic traffic messages: if numerous congestion patterns can be assigned to archived traffic messages, an appropriate value-based definition of congestion can be assumed.
- The characteristic of congestion patterns resulting from the definition of congestion based on a specific value provides a clue for the estimation of the suitability of the definition of congestion has to be supposed..
- In Brilon and Schnabel (2003) and in HCM (2000) specifications of value based definitions of congestion can be found.

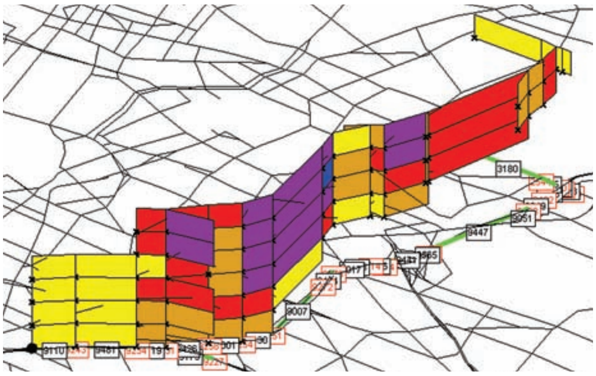


Figure 4: Congestion pattern, time proceeding bottom-up

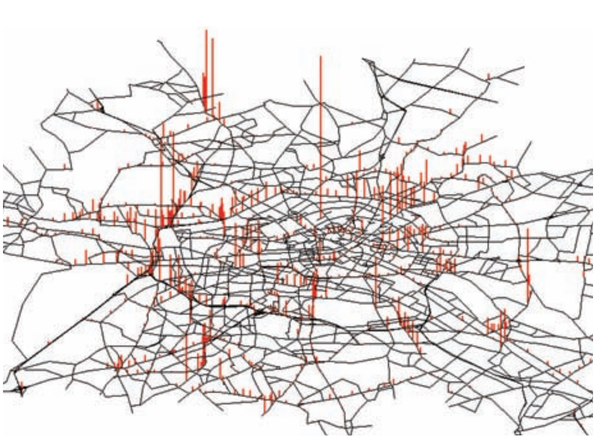


Figure 5: Frequency of congestion of each link in the Berlin road network in May 2005

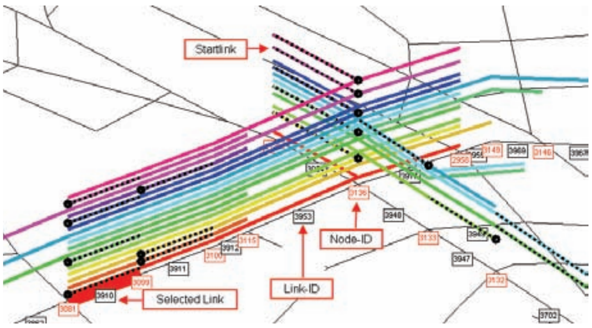


Figure 6: Example of a link class

Definition of congestion	Entry rate [%]	Number of congestion patterns	Rate [%] with connection to traffic message	Congestion with 1 link [%]	Congestion with ≥ 4 links [%]	Congestion duration ≤0.5h [%]
(1) $v/v_0 \leq 0.3$	0.07	1928	54.1	70.7	7.1	67.9
(2) $v/v_0 \leq 0.4$	0.20	3404	41.2	54.1	18.9	63.6
(3) $v/v_0 \leq 0.5$	0.58	5691	27.3	45.2	27.2	57.8
(4) $\Delta T > 60s$	0.19	5658	21.9	80.8	2.0	55.3
(5) $\Delta T > 45s$	0.59	8533	17.7	76.6	3.1	54.6
(6) LOS=2	0.28	2922	50.7	74.9	5.1	53.2

Table 1: Comparison of different definitions of congestion

Results for the various different value-based definitions of congestion are contained in Table 1, using data from May 2005. The entry rate in this table describes the number of data sets that document congestion according to the used definition, related to the number of all data sets.

The studies have shown that the quotient of link- and time-specific space mean speed and link-specific free flow speed is mostly in accordance with the indication of congestion using a value of 0.4 as a threshold.

Congestion is present if the characteristic value falls below the threshold. A warning level is reached between the limits 0.4 and 0.5.

Figure 4 shows the spatio-temporal distribution of a resulting congestion pattern in the network. The patterns of congestion can be described by features like start link, start time, total length and total time loss.

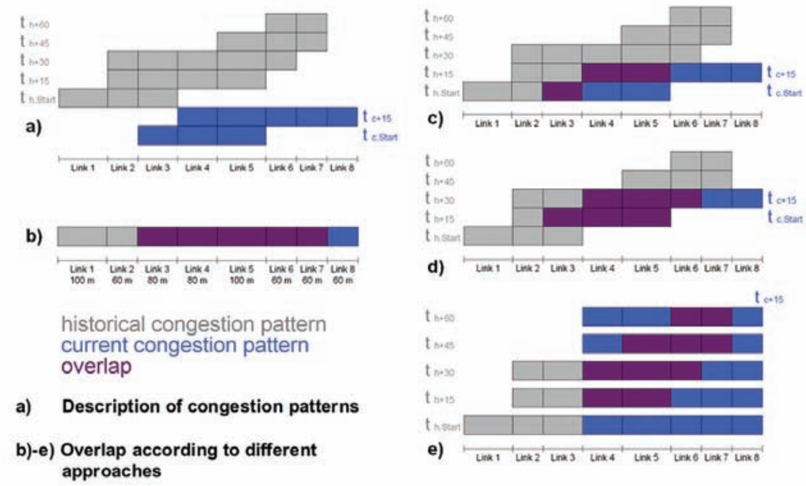
Figure 5 displays the spatial distribution of data sets that achieve the definition of congestion or the warning level and gives an estimation of the main bottlenecks in the Berlin road network.

4. PREDICTION OF URBAN CONGESTION PROPAGATION

The categorisation of historic congestion patterns in suitable classes is fundamental for the short-term prediction of congestion propagation to enable a quick selection of relevant historic congestion patterns.

The classification by congestion features like time of day, total delay or total length is not promising, because references to the local network design are still missing. For example, two cases of congestion in the morning in different sections of the network are not necessarily similar in any way. If

Figure 7: Overlap of two congestion patterns according to different approaches



other criteria are used an analogous situation will result.

This is why a classification by links was carried out: each link represents one class, to which all congestion patterns that affected this link are assigned. Thus, each pattern of congestion belongs to as many classes as links it affects. If one or more connected links are currently congested, historic patterns that have taken place at these links can be queried (Figure 6).

If congestion is observed in the current time interval according to the valid value-based definition of congestion, all historic patterns of congestion are queried which correspond to the currently-appearing pattern.

Patterns that are more similar to the actual congestion are more appropriate for use for the short-term prediction of congestion propagation.

Therefore, it is necessary to evaluate the similarity between the current and each historic congestion pattern in order to select and to appropriately weight historic patterns for the prediction. This is carried out with a pattern recognition method according to the following criteria (compare Figure 7):

- Spatial features, for the complete historic congestion pattern – Figure7b)
- Spatio-temporal features, for the complete historic congestion pattern, without or with offset concerning the first interval of the patterns – Figures 7c) and d)
- Spatio-temporal features, regarding the quality of traffic flow, for the complete historic congestion pattern
- Spatial features, regarding the quality of traffic flow, for each time interval of the historic congestion pattern – Figure 7e), not including the quality of traffic flow

A value called *overlap* quantifies the similarity of the two compared congestion patterns. The spatial overlap according to Figure 7b) results from

$$OL = \frac{\left(\sum_{i=1}^{i=n_1 \cap 2} l_i\right)^2}{\left(\sum_{j=1}^{j=n_1} l_j\right)\cdot\left(\sum_{k=1}^{k=n_2} l_k\right)} \qquad (1)$$

OL… overlap
l… link length
n… number of links affected by the congestion pattern
i,j,k … indexes for the congested links

The supplemented equation considering temporal features as well – according to figure 7c), d) – is

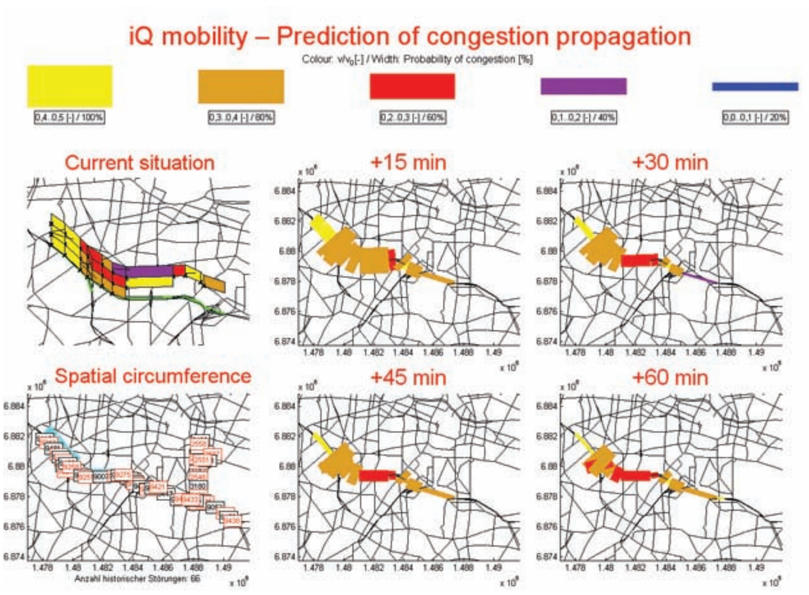


Figure 8: Example for the prediction of congestion propagation

$$OL = \frac{\left(\sum_{p=1}^{p=Z_{1\cap 2}} \sum_{i=1}^{i=n_{1\cap 2}} l_{i,p}\right)^2}{\left(\sum_{q=1}^{q=Z_{1\cap 2}} \sum_{j=1}^{j=n_{1,q}} l_{j,q}\right)\cdot\left(\sum_{r=1}^{r=Z_{1\cap 2}} \sum_{k=1}^{k=n_{2,r}} l_{k,r}\right)} \qquad (2)$$

Z… number of time intervals
p,q,r… indexes for the time intervals

To realise the approach of Figure 7d) equation (2) has to be used for all possible offsets in order to assess the optimal offset. According to Figure 7e) equation (1) is used for the spatial comparison of the actual congestion and historic congestions for all intervals of the historic congestion pattern.

Up to now the equations do not include quality of traffic flow aspects, which can be displayed by the quotient of travel speed and free speed.

They can be incorporated into the equation in the form of a factor for each summand in the numerator, which is close to 1 for similar traffic states and close to 0 for completely different traffic states.

The result of all approaches is an overlap for each chosen historic pattern of congestion between 0 (no similarity) and 1 (absolute similarity) which grows for a stronger similarity between the historic and the current pattern of congestion.

The overlap is calculated for each historic congestion pattern that has at least one common link with an actual case of congestion. This is a meaningful value that can be used in deciding whether the historic pattern should be used for the propagation prediction of the current case of congestion and subsequently for the weighting of the historic pattern within the prediction method. Historic congestion patterns can be excluded from the prediction if their overlap with the current pattern is considerably lower than that of other historic patterns or if the number of historic congestion patterns with a higher value reaches a given threshold.

Hence, the prediction of congestion propagation for the actual congestion is determined by a weighted sum of the chosen historic congestion patterns. Some historic congestion patterns indicate that a certain link will be congested in one of the subsequent time intervals, while others propose just the opposite. Hence the result of the sum is a probability of congestion for each link and each of the time intervals thereafter.

To generate information about the characteristics of the congestion on a certain link in a future time interval the value 'predicted quotient between space mean speed and free flow speed' is calculated using the historic patterns that predict congestion for that link in the referred time interval and their overlaps are used for weighting. The result is a specific value for the predicted quality of traffic flow. The value may be included in the determination of a congestion level.

The probability of congestion and predicted quotient between space mean speed and free flow speed are calculated for each link and each time interval using

$$POC = \frac{\sum_{i=1}^{n} c \cdot OL_i}{\sum_{i=1}^{n} OL_i} \cdot 100 \qquad (3)$$

and

$$QSF = \frac{\sum_{i=1}^{n} qsf \cdot c \cdot OL_i}{\sum_{i=1}^{n} c \cdot OL_i} \qquad (4)$$

POC… probability of congestion [%]
c… congestion (binary; not congested = 0, congested = 1)
OL… overlap
n… number of historic congestion patterns chosen by

the classification by links
i… index for the historic congestion patterns
QSF… predicted quotient between space mean speed and free flow speed
qsf… quotient between space mean speed and free flow speed in the historic congestion pat-tern

The number of predicted time intervals is not limited by the method, but the further in the future the predicted time interval is, the more diffused the results will be. This is why an appropriate limitation is recommended.

The result of the prediction of congestion propagation can be visualised (Figure 8). The width of the bars represents congestion probability; the colour depends on the quotient between space mean speed and free flow speed ranging from yellow via red to purple.

5. ONLINE-PREDICTION OF CONGESTION PROPAGATION IN A NETWORK

Until now a method for the prediction of congestion propagation was shown for a single congestion case. The method was implemented prototypically within the project *iQ mobility* for a network-wide online application in Berlin.

The necessary input data is the network-wide modelled traffic situation provided by the traffic management centre in Berlin.

The traffic situation is checked for network areas with traffic states according to the value-based definition of congestion or according to the warning level. The propagation of the current congestion is predicted for these areas.

It is possible that one or more links are affected by propagation predictions of two current congestion cases, so that several predictions for the link need to be combined. The single predictions contain the probability of congestion and the quotient between space mean speed and free flow speed for each link and time interval. The combined values are calculated by

$$POC_{fused} = 100 \cdot (1 - \prod_{i=1}^{n} (1 - \frac{POC_i}{100})) \qquad (5)$$

$$QSF_{fused} = \frac{\sum_{i=1}^{n} QSF_i \cdot LOC_i}{\sum_{i=1}^{n} LOC_i} \qquad (6)$$

n… number of valid predictions for the multi-predicted link
i… index for the valid predictions

Final result is a table containing the interval-specific values POC and QSF for the links that might be congested during the nearest future time intervals.

The prototype uses historic traffic data from October 2005 to October 2006. It is tested in the *iQ mobility* test site in the centre of Berlin.

6. WARNING MESSAGES AND RECOMMENDATIONS OF MEASURES

The method to predict congestion propagation generates warning messages that enable the operator in a traffic management centre to take timely measures to manage the traffic and thus reduce negative consequences of congestion. The objective of the short-term prediction is to identify the network sections with the heaviest predicted changes concerning the traffic state. The predicted changes of the traffic state are quantified for the different network sections in order to report the most relevant ones to the operator. The quantification process accounts for the lengths of congested links, the probability of congestion according to the prediction and the changes in the quotients of space mean speed and free flow speed for two time intervals. A high positive value represents strong congestion propagation; a significantly negative value denotes that congestion may likely decline.

For practical reasons, it is fixed that maximum messages per time interval are generated concerning congestion propagation and a maximum of one case of degenerating congestion is reported to the operator. The messages include all links that are or will be affected as well as the congestion probability for two time intervals and a five-stage 'level of congestion' indication. The level of congestion is derived from the predicted quotient of space mean speed and free flow speed.

The operator will be informed about ongoing and predicted congestion in the network and appropriate measures will be suggested to him. These measures have two goals:

- Limit the inflow to a congested network section
- Increase the outflow from a congested network section

Besides traffic information via radio and internet, the actuators illustrated in Figure 9 can be used to achieve these goals.

Traffic signals that can increase the outflow from a section are to be found at the front-end of the congestion or within the congested area. Obviously, an increased outflow cannot possibly be achieved for some supply-side congestion causes (such as accidents, for example).

Actuators to help reduce the inflow to congested network sections are to be located on relevant routes leading to these sections. Therefore, the traffic volumes in these sections have to be tracked upstream against driving direction until they are below a threshold volume and the passed actuators are to be diagnosed. Particularly for small networks, the upstream tracking can be stored statically. Dynamic tracking can be carried out by using routes from traffic assignments or by using approximate or measured turning rates.

The prototype system for the test site in Berlin identifies traffic signals in the outflow area as well as variable traffic signs (eg for variable speed limits), ramp metering signals and

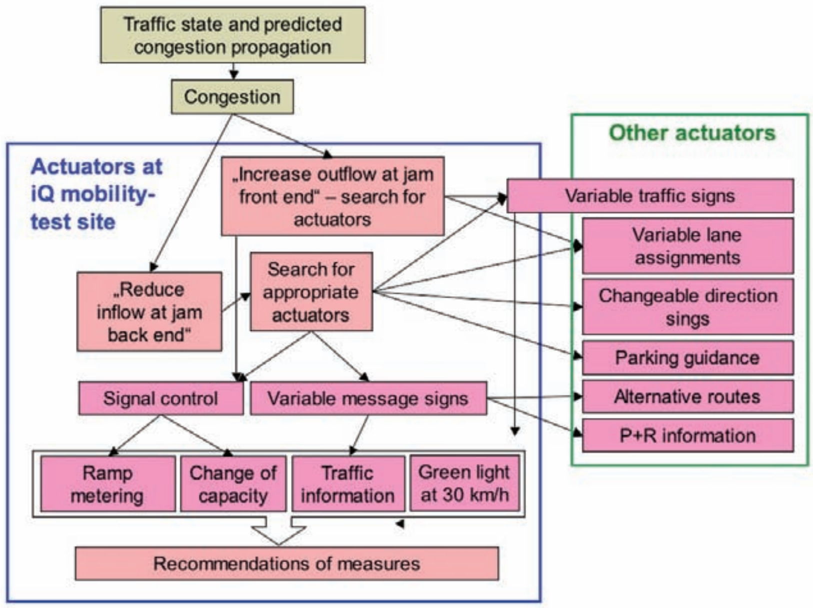


Figure 9: Actuators for traffic management measures in the case of congestion

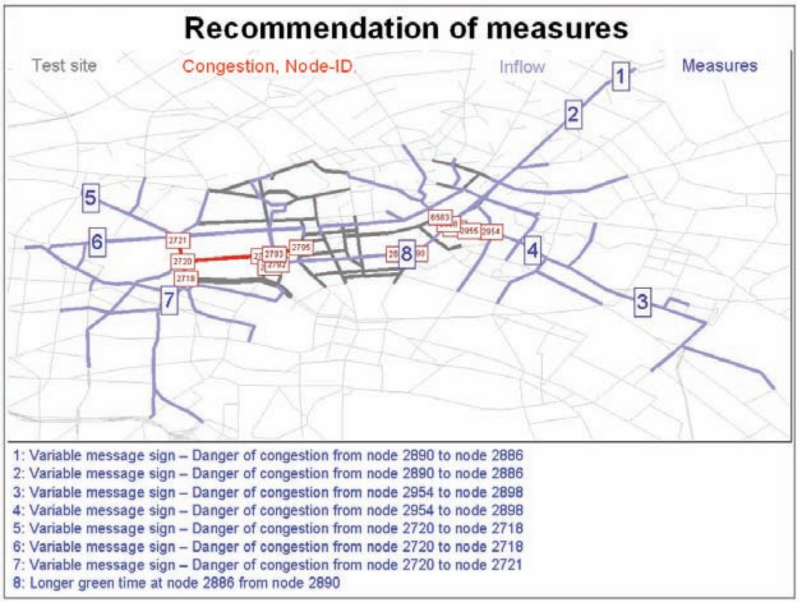


Figure 10: Example of automatically generated recommendations for measures

variable message signs situated in the inflow area to congested network sections. Additionally, it generates recommendations for text messages (eg congestion warnings) on the variable message signs. The results of this process are illustrated graphically (Figure 10).

7. SUMMARY AND OUTLOOK

This paper presents a method for the online prediction of urban congestion propagation. The basis for the predictions are numerous congestion patterns, which are generated offline from historic traffic state data using a suitable value-based definition of congestion. If a network section is currently congested or predicted to be congested, several historic congestion patterns that spatially overlap the current or predicted congestion are compared to the current situation and provided with a value called overlap that implies the similarity between current and historic situation. This value is used as weight for the historic situation in the subsequent propagation prediction step. Network sections with particularly meaningful prediction results are selected and used for warning messages that are sent to the operator in a traffic management centre.

A further step localises actuators in the inflow to and in the outflow from current and predicted congestion for reducing inflow and increasing outflow volumes. Messages for variable message signs are generated and displayed together with other actuators.

A prototype of the prediction method has been in operation on the *iQ mobility* test site in the city centre of Berlin since August 2007. The test results, which were obtained exclusively on a visual basis, are consistent in general, but will soon be backed by an in-depth assessment of the results.

Due to the characteristics of the data used, it could not be investigated whether a classification of historic congestion patterns according to their causes and a relevant filter in the range of patterns would have improved the prediction, or if a comparison of historic to current patterns would have resulted in a selection of similar causes.

Furthermore, the switching states of the actuators during historic congestion are not known. As they might have an influence to congestion propagation, they could be relevant for the prediction method. Hence, they should be considered for further development.

The presented methods assist the operator in a traffic management centre in determining at an early state which targeted corrective measures should be taken in the event of congestion. Therefore, the methods contribute to an improvement in urban traffic flow and, additionally, are a good basis for future partly- or fully-automated traffic management decisions.

REFERENCES

Brilon W., Schnabel W. (2003): Bewertung der Qualität des Verkehrsablaufs auf Hauptverkehrsstraßen. Straßenverkehrstechnik 01/2003
HCM (2000): Highway Capacity Manual. TRB, National Research Council, Washington, D.C.
Projektteam iQ mobility (2007): iQ mobility – Integriertes Qualitäts- und Mobilitätsmanagement im straßengebundenen Verkehr der Region Berlin-Brandenburg. Supplement Straßenverkehrstechnik 11/2007
Vortisch, P. (2005): Modellunterstützte Messwertpropagierung zur Verkehrslageschätzung in Stadtstraßennetzen. Dissertation at the Faculty of Civil Engineering, Geo- and Environmental Sciences, Universität Karlsruhe

AUTHORS

Dipl.-Ing. Friedrich Maier
Tel: +49 89 289 28586 – E-Mail: Friedrich.Maier@vt.bv.tum.de
Dipl.-Ing. Robert Braun
Tel: +49 89 289 22446 – E-Mail: Robert.Braun@vt.bv.tum.de
Univ.-Prof. Dr.-Ing. Fritz Busch
Tel: +49 89 289 22438 – E-Mail: Fritz.Busch@vt.bv.tum.de
Technische Universität München, Chair of Traffic Engineering and Control
Arcisstraße 21, D-80333 Munich
Dr. rer. nat. Paul Mathias
Siemens AG, I&S ITS, Hofmannstr. 51, D-81359 Munich
Tel: +49 (89) 722 25731, E-mail: paul.mathias@siemens.com