UNIFIED EVALUATION METHOD FOR TRAFFIC CONTROL ALGORITHMS

Robbin Blokpoel[1], Jaap Vreeswijk[2], Daniel Krajzewicz[3], Tina Kless[4]

1. Senior Researcher, Imtech Traffic & Infra,
Basicweg 16, 3821 BR Amersfoort, The Netherlands,
+31 334541731, robbin.blokpoel@imtech.com
2. Product Manager Research, Imtech Traffic & Infra, The Netherlands
3. Scientific Coworker, German Aerospace Centre, Germany
4. Graduate Student, Technical University of Dresden, Germany

ABSTRACT

Traffic lights play a major role in traffic management. It is known how to compute the delay-minimizing optimal programs for single intersections or networks of them, if a static flow is assumed. But traffic flow changes on different time scales and many adaptive algorithms were developed, all of which adapt the green time durations, phase orders, and cycle lengths to the current traffic state. Evaluation of these more advanced traffic light control algorithms is more complex. The perfect control strategy is not known, so a comparison with a ground-truth is not possible. In literature many different metrics are used to evaluate the algorithms, but none of them are complete and they are difficult to compare with each other. This paper will derive a unified evaluation method for traffic control algorithms. As a first step a complete set of metrics with minimal overlap is determined. This set will include users' perception, which is currently not used in literature to evaluate control algorithms, but shows significant added value. The second and last step is combining the metrics into one final criterion for evaluation by using a policy based scheme.

Keywords: traffic control, evaluation, floating car data.

INTRODUCTION

Traffic lights play a major role in traffic management. It is known how to compute the delay-minimizing optimal programs for single intersections or networks of them, if a static flow is assumed. But traffic flow changes on different time scales and many adaptive algorithms were developed, all of which adapt the green time durations, phase orders, and cycle lengths to the current traffic state. These adaptions are triggered by input data stemming from traffic surveillance by static field sensors or model predictions and comprise for example flow rate, gap time and queue length. Traditional traffic surveillance techniques, e.g. loop detectors, are expensive and may require temporal road closure for installation and maintenance. Emerging cooperative techniques [1] like vehicle-to-infrastructure communication may increase the knowledge about the traffic situation as well and open new channels for delivering information.

Most cooperative systems, however, require large penetration rates in order to assure their functionality, making the first steps towards their deployment unattractive. The Colombo project [2] works on overcoming this hurdle by delivering a set of modern, self-organizing traffic management algorithms designed for being applicable even at low penetration rates, ensuring their usability from the very first deployment days on. COLOMBO will focus on two traffic management topics: traffic surveillance and advanced traffic light control algorithms.

Evaluation of traffic surveillance in simulation is straight-forward, the algorithms can simply be compared with the ground-truth that can be acquired directly from the vehicles in the simulation. However, for advanced traffic light control this is more complex. The perfect control strategy is not known, so a comparison with a ground-truth is not possible. Several measurements can be acquired, travel time, number of stops and delay time compared to free flow can all be determined overall, per route or per class (e.g. car, truck, pedestrian, public transport, etc.). These measurements should be acquired accurately and in an unambiguous, reproducible way as described in [3]. Emissions are included, but recently the project has improved the emissions modelling in the traffic simulator SUMO [4] with the addition of (Passenger car and Heavy Duty Emissions Model (PHEM) [5] data.

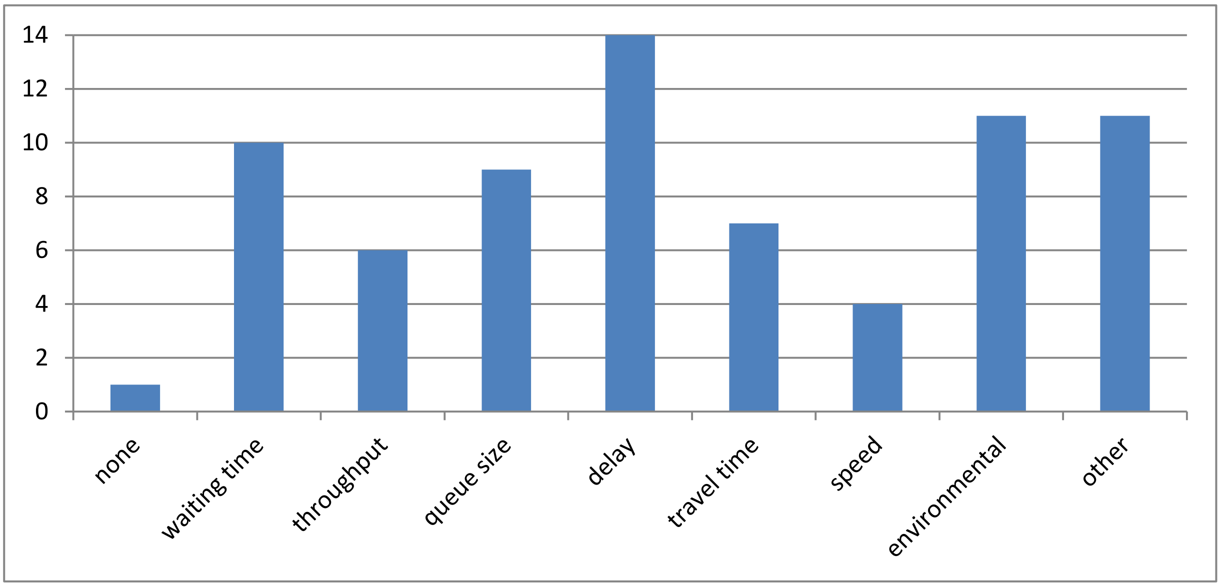

Figure 1: Occurrences of metrics for evaluation of traffic control (values taken from [6])

In [6] a literature review has been carried out over 40 publications and it was analysed which metrics were used for evaluating the traffic control strategy. The resulting graph is shown above. As can be seen from this figure, the most frequent metric is delay time, but still only used in 35% of the publications. This means that it will be difficult to compare the traffic control strategies of those different papers to each other and demonstrates a need to have a unified evaluation method that does make this possible.

Using all these measurements and summing them up will not necessarily indicate how well a traffic control strategy performs. For instance, which control strategy is better? One that has an average delay time of 90s, while the maximum delay time is 500s or a second strategy that has an average of 100s but a maximum of only 120s? Similarly, also the ratios between emissions and delay time have to be determined and this basically holds for all measurements. This means that uni- and multivariate statistics have to be applied onto the measurements for the evaluation.

Arriving at one number is important, not only to have a final “grade” of a traffic control strategy, but also to be suitable for an automatic tuning and configuration algorithm [7] to use it. For automatic configuration one criterion is required, as with multiple criteria an algorithm cannot determine which test run was best, unless all criteria are best for one scenario.

SELECTION OF EVALUATION METRICS

Several metrics from [3] will be compared and choices are made which metrics will form a complete set with minimal overlap to evaluate traffic control algorithms. Travel time and delay time for instance have a large overlap, because the travel time equals delay time plus free flow travel time. Like described in the introduction, also the drivers perception to the delay time will be taken into account using the work of [8]. This implies that not only the average delay values will be taken into account, but also the statistical distribution.

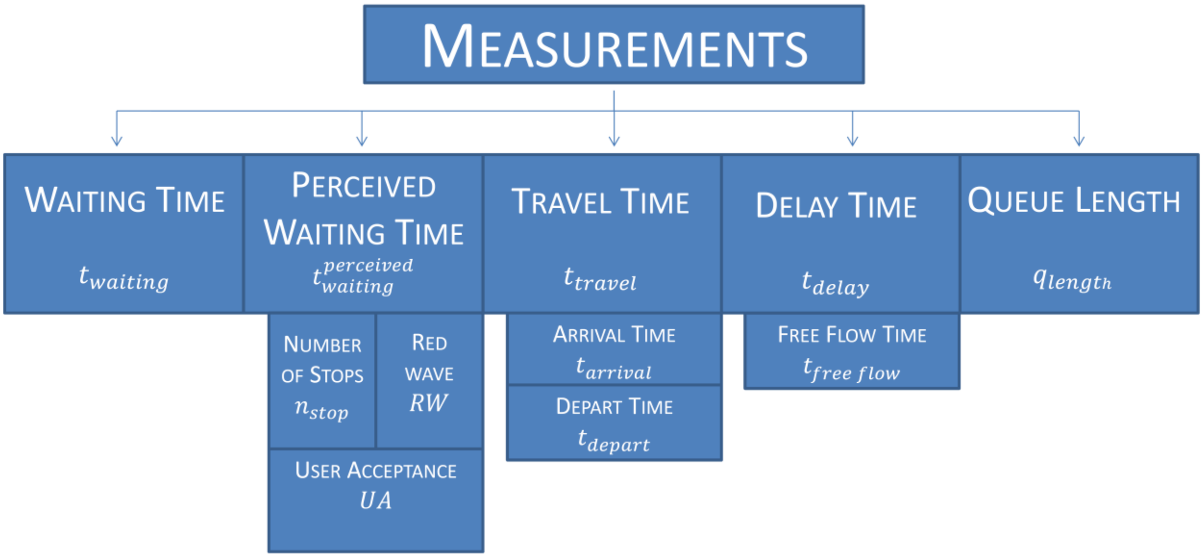

Figure 2: Basic measurements and the needed metrics

Waiting Time

One of the most essential measures is the waiting time ($\boldsymbol{t_{waiting}}$). It is often used as a measure of performance (MOP) [8]. Furthermore, it is the most important indicator of service in the Netherlands [8]. For this reason the waiting time is a fundamental factor for the evaluation of traffic control algorithms. Waiting time is defined as the time that the vehicle is stopped. Many road users are driving at a low speed towards the stop line or end of the queue, while not really standing still yet. Therefore, the waiting time considers cars, which drive with a speed lower than 5 km/h and are not further away from the stop line or the previous car than 5 m. That way the waiting time is measurable.

Perceived waiting time

But what is not measurable is the perceived waiting time. Every road user experiences his waiting time differently than the actual waiting time. But for evaluating the waiting time and for measuring the acceptance of a traffic light it is very important to approach the perceived waiting time. Usually the perception of traffic lights is not included in the process of developing a traffic light controller. For this reason [8] analysed the topic of perception of waiting time at signalized intersections. In the study the perceived waiting time ($\boldsymbol{t_{waiting}^{perceived}}$) is a function of the actual waiting time. Additional factors that influence the drivers experienced waiting time are the number of stops and the presence of a red wave. The number of stops ($\mathbf{n_{stop}}$) is defined as the number of times a car has stopped in the same queue. A red wave (RW=1) or no red wave (RW=0) depends if the car has to stop at two or more consecutive intersections or not.

$$t_{waiting}^{perceived} = \beta_0 + \beta_1 \cdot RW + (\beta_2 + \beta_3 \cdot n_{stop} + \beta_4 \cdot RW) \cdot t_{waiting} + \beta_5 \cdot t_{waiting}{}^2 \quad (1)$$

β_0	13,859
β_1	17,254
β_2	0,661
β_3	- 0,233
β_4	- 0,432
β_5	0,006

Table 1: static parameters for the perceived waiting time [8]

Furthermore, the study focused on the relationship between perceived waiting time and the user acceptance. The user acceptance (UA) of signalized intersections is considered as a function of perceived waiting time. The UA can adopt real values from 0 to 1. The measure is expressing the average presumption of how the road user is going to accept the waiting time [8].

$$UA = \frac{1}{1 + e^{\beta_0 + \beta_1 \cdot PWT}} \quad (2)$$

β_0	- 3,650
β_1	0,055

Table 2: static parameters for the user acceptance [8]

A common behaviour of road users is that they are going to slow down, if they are reaching a stop line or the end of a queue, but the speed will still be over 5 km/h. This way the time lost will not influence the waiting time. For this reason just considering waiting time would make the results of a traffic light look too positive [3]. Therefore, other measures should also be involved in the evaluation algorithm in order to make sure the result is as complete as possible.

<u>Travel Time</u>

The travel time (t_{travel}) is in the literature a very often used measure. But also it is a measure with ambiguous definitions [3]. The definition used in this context will just consider vehicles that have entered and left the simulation area. If vehicles would be considered that did not leave the network yet, the travel time for these vehicles would be appear shorter just because they did not finish their route yet. Therefore, travel time is the time passed between entering and leaving the network, not considering unfinished trips.

$$t_{travel} = t_{arrival} - t_{depart} \tag{3}$$

Delay Time

To get a means for travel time evaluation, the delay time is a good measurement. Sometimes the delay time is also called loss time. The delay is defined as the time that is lost between two intersections or the beginning and the end of the network compared to the time that is needed in a free flow situation. Free flow means that no interactions with other vehicles are disturbing the journey and the driver is obeying the maximum speed limit [3]. The intervals between two intersections should start and end just after the stop line in order to get a realistic intersection performance [3].

$$t_{delay} = t_{travel} - t_{free\ flow} \tag{4}$$

Queue Length

A queue is defined as the sum of all waiting vehicles that arrive at one signal group in front of a traffic light. The length (q_{length}) indicates the space the queuing cars are using before an intersection in meters. It is counted on a per vehicle basis upon arrival at the queue. For all previously described measurements the average value, the maximum, the minimum or/and the standard deviation could be important. But not for every vehicle class all measurement types are significant (see Table 3).

For pedestrians and bikes it is for example irrelevant how long the queue is. Especially because bikes and pedestrians are not lining up in a queue like cars do. Most important for these vehicle classes is the travel time, the delay time and the waiting time. The travel time is important to calculate the delay time. For delay time an average value and the standard deviation should be considered. Regarding the waiting time a maximum waiting time should be measured and compared to a previously determined maximum. The reason for this is that pedestrians and bikes (although maybe with a different threshold) get impatient after a certain amount of time and they will cross the intersection, despite a red light. That could lead to dangerous situations and should be prevented.

Regarding the evaluation of passenger cars, the measures travel time, delay time, waiting time and queue length are fundamental. The perceived waiting time and the needed data like the number of stops and the existence of a red wave are also relevant. The delay time for cars should be measured as an average value and as the standard deviation. Also the average value and the standard deviation of the waiting time and the perceived waiting time are going to be analysed. For the queue length its maximum is most important as it can block the entry to other queues.

For public transport vehicles the travel time and the waiting time is important to measure. Most important in this regard is the delay time, which is measured with its average value, standard deviation and maximal value. Everything is measured per signal group, to make a more precise analysis of the results possible. This way the configuration for one specific problem at a certain signal group can be improved when necessary.

Vehicle Class	**Measurement**	**Type**
Pedestrians	Travel Time	Average
	Delay Time	Average
		Standard deviation
	Waiting Time	Maximum
Bicycles	Travel Time	Average
	Delay Time	Average
		Standard deviation
	Waiting Time	Maximum
Car	Travel Time	Average
	Delay Time	Average
		Standard deviation
	Waiting Time	Average
		Standard deviation
	Perceived Waiting Time	Average
	Queue Length	Maximum
Public Transport	Travel Time	Average
	Delay Time	Average
		Standard deviation
		Maximum
	Waiting Time	Average

Table 3: important measures and types of metrics grouped by vehicle class

COMBINING DIFFERENT EVALUATION METRICS

The selected metrics still have to be combined into a final result. How to get to that result depends on the policies that the road authority specifies for the network to apply the traffic control algorithm to. The problem of combining different policies into one strategy has been solved in [9]. However, the final set of metrics differs from the set in [9] and the extension of that work will be presented here. The major challenge is how to combine the driver's perception of delay time and throughput with the other metrics.

An algorithm is an explicit well-defined instruction code to solve a problem. In this case the problem is an optimization problem. The aim is to combine the demands of all road users and at the same time the system should get a certain value as a single evaluation result. This result makes it possible to compare and evaluate the performance of different control strategies.

Every measurement needs to get a final grade in order to compare and rate it. This grade depends on recommended restrictions from research and the policies that the road authority specifies for the network. The maximum waiting time for pedestrians for example should not be higher than 90 seconds, because otherwise impatient pedestrians could cross the road.

1	***2***	***3***	***4***	***5***
insufficient	sufficient	satisfactory	good	very good

Table 4: grade scale

The grade scale is presented in table 4, in which insufficient means that the measured value for the specific metric is higher than the maximum or lower than minimum value. Very good means that the recommended value is even outreached. After all metrics received a grade the metrics have to be weighed regarding their importance (equations 5, 6 and 7).

$$\boldsymbol{E_m = \frac{1}{\sum_{m_x} w_{m_x}} \sum_{m_x=1} g_{m_x} \cdot w_{m_x}} \quad (5)$$

$$\boldsymbol{E_v = \frac{1}{m} \sum_{m=1}^{m} E_m} \quad (6)$$

$$\boldsymbol{E_{sg} = \frac{1}{v} \sum_{v=1}^{v} E_v} \quad (7)$$

v	vehicle class
m	measurement
m_x	measurement type
g	grade
w	weight
sg	signal group
E	evaluation value

Weight	Description
0	Not important
1	Interesting, but not significant
2	Relevant
3	Important
4	Very important

Table 5: weighing scale

The weight of zero should not be used, because all irrelevant metrics are already filtered. The value of a weight parameter is going to be proposed here, but of course it is changeable for networks with special policies. This way a reasonable weighing scale is suggested, but at the same time the system does not get static.

Vehicle Class	Measurement	Type	Proposed Weight
Pedestrians	Travel Time	Average	1
	Delay Time	Average	2
		Standard deviation	2
	Waiting Time	Maximum	4
Bicycles	Travel Time	Average	1
	Delay Time	Average	3
		Standard deviation	3
	Waiting Time	Maximum	4
Car	Travel Time	Average	1
	Delay Time	Average	2
		Standard deviation	3
	Waiting Time	Average	2
		Standard deviation	3
	Difference Waiting Time and Perceived Waiting Time	Average	3
	Queue Length	Maximum	4
Public Transport	Travel Time	Average	1
	Delay Time	Average	4
		Standard deviation	4
		Maximum	4
	Waiting Time	Average	2

Table 6: weighing of measures and types of metrics grouped by vehicle class

In the following figure a flow diagram is shown, that visualizes the behavior of the algorithm.

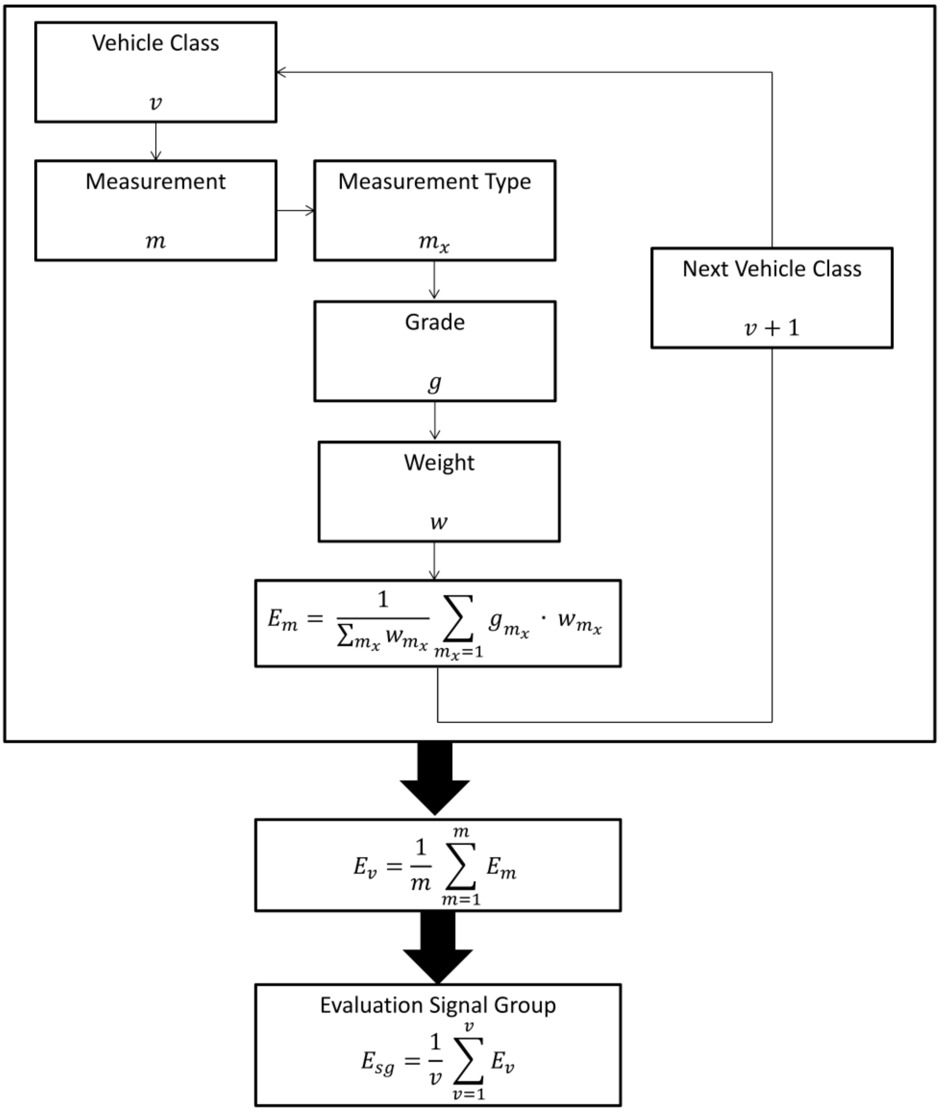

Figure 3: Structure chart of the algorithm

The grade scale shown in the previous (Table 4) applies also for the evaluation value E_{sg}.

CONCLUSION

The goal of this paper was to use a selection of measurements that are relevant to evaluate a traffic control. Furthermore, these measurements should be combined with the help of an algorithm to a final number. That way traffic control strategies can be evaluated and compared. But also the measurements have to be rated to a final grade. This paper showed which metrics are important for the different vehicle class evaluations. It also gives a solution how the evaluation metrics should be combined into one final result. This result is as claimed measurable, assessable and comparable.

REFERENCES

1. Koenders, E., in ‘t Velt, R., *Cooperative technology deployed*, ITS Europe, Lyon, France, 2011.
2. Krajzewicz, D., et al., *COLOMBO: Investigating the Potential of V2X for Traffic Management Purposes assuming low penetration Rates*, ITS Europe congress, Dublin, Ireland, 4th of June 2013.
3. Blokpoel, R., Krajzewicz, D., Nippold, R., *Unambiguous metrics for evaluation of traffic networks*, 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), 19-22 Sep. 2010, Madeira, Portugal.
4. Krajzewicz, D., Bonert, M., Wagner, P., *The Open Source Traffic Simulation Package SUMO*, RoboCup 2006 Infrastructure Simulation Competition, Germany, Bremen.
5. Rexeis, M., Hausberger, S., et al., “*Heavy duty vehicle emissions*,” University of Technology, Graz, Final Report of WP 400 in ARTEMIS, 2005.
6. Krajzewicz, D., et al., Towards a unified evaluation of traffic light algorithms, Transport Research Arena, Paris, France, 14th of April 2014.
7. Stützle, T., López-Ibáñez, M., *Automatic (offline) Configuration of Algorithms, Genetic and Evolutionary Computation Conference*, GECCO '13, New York, 2013.
8. Bijl, B. van der, Vreeswijk, J. D., Bie, J. & Berkum, E. C. van, *Car drivers' perception and acceptance of waiting time at signalized intersectons*. 14th IEEE ITSC Conference, Washington DC, 5-7 October 2011.
9. Van Vliet, K., Turksma, S., *ImFlow: Policy-based adaptive urban traffic control First field experience*, ITS Europe, Dublin, Ireland, 2013.