Towards a unified Evaluation of Traffic Light Algorithms

Daniel Krajzewicz\textsuperscript{a*}, Robbin Blokpoel\textsuperscript{b}, Wolfgang Niebel\textsuperscript{a}, Cornelia Hebenstreit\textsuperscript{c}, Jérôme Härri\textsuperscript{d}, Michela Milano\textsuperscript{e}, Anna-Chiara Bellini\textsuperscript{e}, Thomas Stützle\textsuperscript{f}

\textsuperscript{a}Institute of Transportation Systems, German Aerospace Center, Berlin, Germany
\textsuperscript{b}Imtech Traffic & Infra, Amersfoort, Netherlands
\textsuperscript{c}Technical University Graz, Graz, Austria
\textsuperscript{d}EURECOM, Sophia Antipolis, France
\textsuperscript{e}Università di Bologna, Bologna, Italy
\textsuperscript{f}University libre de Bruxelles, Brussels, Belgium

Abstract

Within the COLOMBO project, modern traffic surveillance and traffic light control algorithms based on data obtained from vehicular communications are developed. An evaluation of existing descriptions of traffic light evaluation show that both, common measurement definitions as well as standardised simulation scenarios are missing. We present a definition of measurement methodology based on earlier projects, namely iTETRIS and FESTA. Additionally, the scenarios under development are presented, including basic, synthetic ones, as well as some that represent real-world networks.

Keywords: traffic lights evaluation, traffic simulation, simulation scenarios, performance metrics.

Résumé

Dans le COLOMBO projet on développe la surveillance du trafic et des algorithmes de commande des feux de signalisation basés sur les données de communications véhiculaires. Une évaluation des descriptions existantes des feux de trafic montrent qu’il manque des définitions communes de mesures et des scénarios de simulation normalisés. Nous présentons une définition de la méthodologie de mesure basée sur des projets antérieurs comme iTETRIS et FESTA. En outre les scénarios en cours de développement sont présentés.

Mots-clé: évaluation des feux tricolores, simulation de trafic, scénarios de simulation, mesures de performance.

\textsuperscript{*} Daniel Krajzewicz. Tel.: +49-30-67055-273; fax: +49-30-67055-291.
E-mail address: daniel.krajzewicz@dlr.de.
1. 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 if a static flow is assumed. But traffic flow changes on different time scales and many 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 flow rate, gap time, etc. Recent research and development includes cooperative vehicles as new data sources. Usually, the performance of new algorithms is evaluated before their field deployment using traffic simulations.

The work presented here is performed within the COLOMBO project (COLOMBO, 2012), co-funded by the European Commission. COLOMBO’s main technical task is to develop cost-efficient traffic management applications for traffic surveillance and traffic control via traffic lights that use information gained from vehicles equipped with vehicular communication (V2X) technology. Even though such technology is going to be deployed from the year 2015 on according to (C2C CC, 2012) the possibility of a slow increase in such vehicles’ number should be considered. This is backed by the average age of the car fleet between 4 years in Luxembourg, 8 years in the EU25, and over 14 year in Greece (EEA, 2011) and the lack of obligatory equipment as expected in the USA. For this reason, the solutions developed within COLOMBO are targeting to operate at low penetration rates of V2X-equipped vehicles. As a coarse working target, 10% are assumed as a “low penetration rate”, with investigations starting from 1%.

COLOMBO develops the solutions based on simulations only. COLOMBO’s overall simulation system (COLOMBO, 2013) joins available, established and domain-specific simulation systems, including the communication simulator ns-3 (ns-3, 2013) and the traffic simulation SUMO (DLR, 2013; Krajzewicz, Erdmann, Behrisch, Bieker, 2012). Both are joined into a single execution system using the “iTETRIS Control System” (iCS; Rondinone, 2013), a middleware for simulating large-scale traffic management applications based on vehicular communications developed within the iTETRIS project (iTETRIS, 2013) that was performed between 2008 and 2011, co-funded by the European Commission. Due to the goal of developing solutions that reduce the environmental impact, the system is extended by PHEM (Hausberger, 2003), a leading tool for computing vehicular emissions. Albeit interfaces to PHEM itself were already implemented, a derived module, named PHEMlight that is directly included in SUMO, will be used. This direct inclusion reduces the complexity of the overall simulation system, avoids the generation of intermediate files, and allows to use further features of PHEM. Besides these components, the COLOMBO simulation system includes a tool for automatic algorithm configuration used for optimizing the behavior of the developed solutions, and means for final output evaluation. Fig. 1 shows the overall simulation system schematically.

![Fig. 1. A schematic view at the COLOMBO overall simulation system.](image-url)
To enable and support this evaluation, the respective evaluation methodology has to be defined beforehand. It comprises the breakdown of the top-level goals, i.e. objectives, into computable measurements, the agreement on evaluation procedure(s), and a set of external simulation scenarios and according configurations. Chapter 2 contains results of a literature review which investigated how and to what extent this preparatory work is described in publications. COLOMBO’s goal is to deliver a “traffic light algorithms evaluation suite” as a generic tool set that can be applied to other traffic light algorithm investigations. Due to its general methodology, investigations by different researchers become less time consuming and better comparable. The approaches about unified measurements are described in Chapter 3, and different simulation scenario configurations follow in Chapter 4. The paper is concluded in Chapter 5.

2. Traffic Lights Evaluation in Literature

One assumption within COLOMBO was that most publications about new traffic light algorithms are hardly comparable due to different scenario set-ups and performance indicators (PIs), also called Measures of Performance (MOP). To validate the ground truth, fifty (50) publications that evaluate new traffic light algorithms were scanned. Ten of those publications have been neglected, as they present an idea only, give overviews about traffic light algorithms, or because they did not evaluate a traffic light algorithm for any other reason. The following information was extracted from each of the remaining 40 publications:

- What kind of a scenario (single intersection, corridor, network) was used and is it properly described, including both the road network (numbers of lanes, allowed directions per lane) and the demand?
- Which measures and PIs were used and are they properly defined?

Additionally, the used traffic simulation and the evaluated algorithm’s paradigm (Fuzzy Logic, Genetic Algorithms, etc.) were noted. These two latter classifications (see Fig. 2) are reproduced as they may give hints about the context the publications have been performed within. Looking at the used traffic simulations, one can find a high number of self-developed systems (10 of 42) while commercial simulation packages, such as Vissim, CORSIM, Paramics, TRANSYT, or AIMSUN (one representation), are used more seldom. The emphasis on certain techniques from computer science, such as Neural Networks, Q-learning, or Genetic Algorithms is also higher than one would expect, so that one could come to the conclusion that most of the evaluated publications stem rather from computer scientists than from traffic engineers. This surely does not reflect the common usage of traffic simulations, as the major use case is fine-tuning and benchmarking of real-world traffic light systems before deploying them. But it fits very well to the scope of COLOMBO: delivering a base system for traffic light algorithms evaluation to a broad, multi-disciplinary community.

Within the forty publications, 44 different metrics have been used to monitor the investigated traffic light algorithm. They can be grouped as shown in Fig. 3 with “waiting time”, “queue size”, “delay”, and “travel time” used most often. Nonetheless, each of these four groups contains specific variations of a basic “metrics”. As an example, “delay” is constituted by the following (dedicated) names used in the evaluated papers: “delay”, “average delay per vehicle per second”, “average delay per vehicle”, “delay at intersection”, “total mean delay”, “mean rate of delay”, “total delay”. It should also be mentioned, that only one of the examined publications gave a formula for the used measures – the others just named or described the used ones.
The spatial 2D-scale of the examined scenarios ranges from a single intersection, over a corridor (several intersections in a line), to complex road networks. In 8 cases, a single intersection investigation serves as an initial proof before applying the algorithm for controlling traffic in a corridor or a network. At all, 11 corridor, 20 network, and 19 single intersections scenarios were used. When looking how well the scenarios are described, one has to realize that 89% of single intersections could be replicated using the information given in the according paper, but only 36% of corridors, and 20% of the networks. A 3D-model for the inclusion of the altitude could not be observed. The descriptions of the demand span between a single peak hour and 16 hours of an average working day.

To carry out valid simulations some work steps have to be followed, including calibration and statistical testing, as described in guidelines like (FHWA, 2004) and (FGSV, 2006). It did not always become clear, whether these prerequisites for scientific studies were fulfilled.

As a summary, one can state that given the publications only, a direct comparison of the performance of the traffic light algorithms presented in the evaluated papers is not possible. A clear definition of measures to use as well as a set of commonly used (“standard”) scenarios should help to compare algorithms and to determine where they perform well.

3. Derivation of Evaluation Measures

COLOMBO re-uses methods for analyzing the performance of ITS solutions developed within prior projects co-funded by the European Commission. FESTA (Field opErational teSt supporT Action). FESTA was concerned with the development of guidelines for field operational tests (FOTs) and put a strong emphasis on the comparability of results from different FOTs (FESTA, 2008). As a framework for data gathering and evaluation, the distinction between hypotheses, performance indicators (PIs), measures, and sensors was made. Here, PIs are the top-level values, chosen and collected to prove or disprove a previously formulated research hypothesis. A set of PIs may be used to support a hypothesis. PIs are defined as “[…] quantitative or qualitative measurements, agreed on beforehand, expressed as a percentage, index, rate or other value, which is monitored at regular or irregular intervals and can be compared with one or more criteria.”, emphasizing that “A denominator is necessary for a PI. A denominator makes a measure comparable (per time interval/per distance/in a certain location/…).” (FESTA, 2008, p. 29). Measures are the values delivered by sensors and are used for computing the PIs, where “Several PIs can use the same measures as input, and the same measures can be derived from different types of sensors.” (FESTA, 2008, p. 30). In contrary to PIs, measures do not require a denominator.

FESTA itself was mainly concerned about the performance of individual vehicles, a different scope than the work topic addressed in COLOMBO where macroscopic traffic performance measures are of interest. As a consequence, the measures and performance indicators used in FESTA cannot be used. Instead, COLOMBO will re-use the “metrics” for benchmarking traffic efficiency developed by the iTETRIS project. The iTETRIS metrics (Blokpoel, Krajzewicz & Nippold, 2010) were set-up with the goal to have an unambiguous and fully described set of measurements based on simulation outputs for benchmarking simulated ITS solutions. After collecting candidates, a first distinction between measurements that describe the performance of traffic and those

![Fig. 3. Occurrences of metrics within the evaluated papers.](image-url)
that describe the infrastructure was made. The latter ones – including measures such as “number of potholes” or “average number of lanes per road” were neglected, as they form the input to the simulation, not her output, and because they are not affected by ITS solutions at all.

The second distinction made in iTETRIS was to distinguish between “intersection metrics” and “network-wide metrics”. This classification was done as not all metrics of the first kind can be appropriately used for the second areal scope. For example, the “queue length” is a common measure during the evaluation of a single (controlled) intersection, but is rather useless for being applied to a complete network. On the other hand, “travel time” is often used to determine the performance of an ITS application within a road network, but is not used for single intersections. For all chosen metrics, their basic measures (comparable to FESTA’s sensor outputs) were defined, first, and in a second step the PIs themselves. The detail grade of the PIs defined in iTETRIS and of those delivered by FESTA is similar. As an example, we would like to present the iTETRIS definition of the network-wide mean (average) velocity measure (from iTETRIS, 2009, with minor corrections):

\[
P_s^{\text{velocity,mean}} = \frac{1}{t_s^{\text{end}} - t_s^{\text{beg}}} \sum_{t = t_s^{\text{beg}}}^{t_s^{\text{end}}} \sum_{v \in n_{v, s, t}} v_{\text{veh}, t}
\]

where:

- \(t_s^{\text{beg}}\): the time at which scenario \(s\) begins (its first simulation step, in [s])
- \(t_s^{\text{end}}\): the time at which scenario \(s\) ends (its last simulation step, in [s])
- \(n_{v, s, t}\): (set of) vehicles within the scenario \(s\) at time \(t\)
- \(v_{\text{veh}, t}\): velocity of vehicle \(v\) at time \(t\) (in [m/s])

Overall, the iTETRIS project delivered 28 definitions for basic measures, 12 intersection metrics and 17 network performance metrics. Within COLOMBO, these metrics will be revisited, extended by further performance indicators, such as level-of-service (LOS), and transferred into FESTA’s nomenclature. Furthermore, the COLOMBO project is working on metrics that describe the performance of vehicular communications.

The large number of different indicators is surely to be scrutinized. It allows a large variety of investigations, but lacks an explicit advice which of them can be used for a certain purpose. For this reason, the correlations between different measures have to be investigated and a most promising sub-set of those has to be chosen.

4. Scenarios

When using traffic simulations for fine-tuning and benchmarking real-world traffic light the intersection or area to equip, to update, or to upgrade is modeled in a simulation, including the new traffic light program. The resulting scenario is specific for the investigated intersection or area. This usage differs from what is tried to be supported by the work on scenarios in COLOMBO: the development of new algorithms for traffic light control. Here, scenarios that allow evaluating the algorithms in a broad range of possible settings are required.

To achieve this goal, a distinction between “synthetic” and “real-world” scenarios is made, both having a certain purpose. “Synthetic” scenarios, ranging from simple cross-shaped single intersections up to complex networks, do not resemble a certain part of a real-world network. Nonetheless, they allow to investigate the performance of the developed traffic light systems in a clearly defined and easily understandable environment. The supported real-world scenarios are meant to confront the investigated algorithm with complex settings that include different modes of transport, complex geometries, and priority restrictions. Both kinds of scenarios are described in detail in the following. They shall be made available to the public at a later step of the project.
4.1. Synthetic Scenarios

Synthetic scenarios may range from single intersections up to complex networks. The geometries of the modeled intersections, the flows across them, the parameters of roads, and, in the case of more complex networks, the distances between the intersections can be explicitly defined making the scenarios easily replicable across different simulation software packages and allowing to rebuild them by naming few information only. The simplifications increase the ability to evaluate the performance of a modeled traffic light in an analytical way. The possibility to increase the complexity by adding further peculiarities, such as dedicated bus lanes, pedestrian crossings, or certain right-of-way rules helps the developer of an algorithm to cope with real-world complexity.

To support the evaluation of a traffic light in a broad range of (still synthetic) scenarios, a small Python library was implemented. It allows generating single intersection, corridor, and network scenarios in an automatic way. Basic synthetic scenarios are set up by defining a symmetric four-legged intersection with static flows. The attributes of the involved roads – their numbers of lanes, allowed velocities, additional lanes at the intersection they yield in, the central reservations, lanes dedicated to a certain transport mode – can be iterated to a large degree, albeit after being discretized (starting positions of additional lanes position, e.g.) and regarding inter-dependencies between the attributes (there cannot be more bus lanes than lanes at all). When building larger road networks, a pre-defined single intersection is used as template. The script allows overwriting certain attributes explicitly or using rules.

Fig. 4. Examples of synthetic road networks as shown by SUMO.

Besides road networks, demand definitions are needed for obtaining runnable traffic simulation scenarios. Most of the evaluated publications (see section 2) used a static flow along the complete simulation time. Some algorithms that adapt to the current traffic demand but need changing vehicle flows to show their capabilities. A work hypothesis is that traffic light algorithms work on different time scales. Common actuated algorithms look at inter-vehicle time gaps which are in the magnitude of seconds. More complex traffic light control algorithms, such as UTOPIA, e.g., use queue lengths in front of the controlled traffic lights as well as their development over time. Their reaction time scale is in the magnitude of minutes. Weekly switch plans, finally, take into account traffic flow differences between different days of a week or daily shifts in the traffic flow, such as the fact that usually, traffic moves into the city in the morning, and leaves the city in the afternoon. They operate in the magnitude of hours. This assumption is investigated within COLOMBO in a two-fold way. On the one hand, given measures of real-world traffic behaviour are evaluated for determining flow change frequencies to understand the dynamics found in the real world. On the other hand, existing traffic light adaptation algorithms will be investigated for their performance under changing flow conditions, where both, the frequency and the amplitude of the changes are varied. To allow such investigations, three parameterised demand models can be used: a static flow, a linear transfer between two flows, and flows modelled using a sine wave composition (cf. Sohr, Wagner, Brockfeld, 2009), including the amounts of turning vehicles and percentages of vehicle types. Figure 5 shows example developments of the demands over time generated using these models.
The described methods for network and demand generation allow setting up a large number of scenarios, including single intersections, corridors, and networks, with an increasing complexity. They allow to determine the performance of the considered traffic light under a high variety of settings. On the other hand, the large number of possible scenarios raises the question whether all of them are meaningful. This will be the addressed in the following parts of the project.

4.2. Real-World Scenarios

Real-world scenarios resemble traffic in a part of the world. Such a representation should include all the influences that form real conditions, including correct road geometry, the positions and definitions of the traffic lights and other signalling structures, the real demand for all types of transport, or prioritisations of certain types of transport. The major problem for academics, especially those not originally located in traffic science (compare section 2), is to gather the data needed to replicate these parts. If given, the adaptation of such data for the used simulation is usually very time-consuming as the data are usually stored in formats that cannot be directly loaded by a traffic simulation. Therefore, the release of publicly available real-world scenarios is assumed to be a great help for the scientific community. COLOMBO will release a set of such scenarios.

Real-world road networks can be hardly described using an abstract notation. Albeit few standardized formats for describing road networks exist, such as OpenDRIVE, RoadXML, or OpenStreetMap, traffic simulations usually use their own format that resembles the application’s peculiar view at road networks. For this reason, the real-world scenarios developed in COLOMBO will be set up as inputs for the used road traffic simulation SUMO, only. The SUMO application suite contains a network conversion module that could be extended to export road networks for their use in other traffic simulations, but this work is not planned for the COLOMBO project.

The same counts for the description of the demand. At the scenario level, each vehicle is described individually, giving its type, route through the road network, and departure time and further optional attributes exist that control how the vehicle shall be inserted into the road network and how they shall leave it. Large SUMO scenarios may cover up to Millions of single vehicle descriptions. It is not possible to aggregate the vehicle routes, e.g., into O/D matrices, at this point, as certain peculiarities of the demand could get lost. For these reason, real-world scenarios will be made available as SUMO-inputs only.
Most of the supplied scenarios will re-use work done in past projects. Within iTETRIS, a set of network scenarios originally located in the Italian city of Bologna was implemented (see Fig. 6), based on inputs from the municipality of Bologna. The original scenarios were given as inputs to the traffic simulation tools Vissim and VISUM and as a set of GIS (geographic information system) input files. The scenario on the left contains the entire inner city ring and all minor roads around it. This ring road has up to 3 lanes for the main flow, but the minor roads also contain narrow one-way streets which are restricted to be used only to public transport. The main roads are sometimes congested and this poses a challenge for traffic light control to solve. The smaller scenario on the right in Fig. 6 is located at the outer part of the city ring. The supplied demand information covers the peak traffic hour between 8:00 am and 9:00 am in both cases. Both networks, however, lack pedestrians and bicycles, which can be an important factor to signal control when they are present. The scenarios were converted into SUMO-formats and are currently under revalidation.

Figure 7 shows the north of the city of Assen (Netherlands). This scenario has large traffic streams going north-south and vice versa depending on the time of the day. This is because this network connects a highway located to the north and the city centre to the south. The area itself is an industrial area, which leads to traffic turning at the intersections as well. The main policy goal for the network is to have a green wave in north-south direction while not causing too much waiting time and long queues for other traffic.

The middle two intersections also have pedestrian and bike crossings that are part of the traffic light control plan. The original plans are semi-dynamic, the main direction has a fixed time green wave, while the other signal groups allow some flexibility according to the current demand. Within the scenario, this is implemented as a completely fixed time plan that approximates the original plan in the best way possible. This may yield in
differences of the clearance and the intergreen times between each signal group pair. This means that the plan, even implemented as completely fixed time, will still be very complex. Therefore, the signal times included in the scenario use an average intergreen time. Later in the project it will be investigated if an interface between SUMO and the original controller can be made, so that the semi-dynamic behaviour can be simulated with the same controller as used in reality.

![Figure 8. Pickering scenario with a single intersection.](image)

Figure 8 shows a scenario from Pickering, North Yorkshire, in the United Kingdom where left-hand driving is ruled. Theoretically it should not make a difference for a traffic light controller whether traffic drives on the right or on the left, but such a scenario with left-hand driving allow verifying this. Another challenge of this network is the roundabout to the right. This intersection is not controlled, but can cause small traffic jams that spill back to the single intersection on the left. Therefore, this scenario is a good test case to see how a controller deals with spillback. On a more detailed level, the intersection also contains pedestrian crossings and has partial conflicts between right turning traffic and oncoming traffic that is going straight. The right turning traffic has to wait for a gap in the oncoming traffic before it can leave the intersection. This can again be challenging to the controller as it causes the saturation flow to vary with the turning percentage.

Further scenarios under preparation were originally set up within the German project "ORINOKO". They are located in the fair area of Nuremberg, Germany (see Fig. 9). Here, traffic light definitions as well as detector measures are available, too. The traffic light definitions include a description of the weekly switch plan implemented in this area. As no flow information was available, the demand had to be generated using the induction loops measures. The original scenarios used a commercial road network from NavTeq. To make the scenarios available to third parties, a network from OpenStreetMap has to be extracted and enhanced.

![Figure 9. Nuremberg scenario from the ORINOKO project.](image)

5. Summary and Outlook

We have shown the work performed within the COLOMBO project on evaluating traffic light algorithms. A brief survey on similar work done by other research groups was presented, showing that clear definitions of
performance indicators and scenarios used for such evaluation should be attempted if the results shall be comparable. We have described how a set of well-defined indicators, initially generated within the iTETRIS project, is extended and projected onto the evaluation schema proposed by the FESTA project. Additionally, a sub-set of the scenarios generated within COLOMBO was presented of which some will be made available to the public at later steps of the project.

The shown scenarios and indicators allow to compare the performance of newly developed traffic light algorithms against fixed cycle controls or previously evaluated algorithms. The implemented library for generating synthetic scenarios – including the needed road networks as well as the demands will allow to perform investigations for a large number of configuration. This should help to determine settings where a developed traffic light systems performs well.

Even if the used performance indicators and the scenarios are well described, it is likely to get different results from different traffic simulation software. Also the type of algorithm embedding matters, as Hardware-, Software-, and Emulator-in-the-loop yield different outcomes. Establishing a common standard for measuring traffic light performance using traffic simulations is therefore beyond the project's scopes - too many different simulation systems would have to be considered and updated, if needed. Nonetheless, we hope to increase the portability of evaluation scenarios and make some first steps towards a unified evaluation of traffic light system, increasing the comparability of such research.

Acknowledgements

We want to thank the European Commission for co-founding the COLOMBO project. Substantial work on the specific topic of the paper was contributed by Stefan Hausberger (TU Graz).

References


CAR 2 CAR Communication Consortium C2C CC (2012), Memorandum of Understanding on Deployment (press release), 2012-10-10


