Evaluating real-time information systems on public transport disturbances

Alessio Daniele Marra*[1], Francesco Corman[2]

[1] Dr., Institute for Transport Planning and Systems, ETH Zurich, Switzerland

[2] Prof. Dr., Institute for Transport Planning and Systems, ETH Zurich, Switzerland

SHORT SUMMARY

Real-Time Information (RTI) systems are a key component of the management process of public transport disturbances. Smartphone applications, in particular, are becoming a popular means of disseminating information to passengers. Despite the widespread usage of RTI systems, little is known on how accurate those systems are, which information they provide, or which disturbances are not reported.
This work proposes a methodological framework and a set of metrics to evaluate a text-based RTI system, comparing the alerts sent to passengers with the actual disturbances in the network, described by Automatic Vehicle Location data (AVL). A case study is conducted on the RTI system of Zurich to evaluate its performance. The results show high precision in providing correct information, despite only a small percentage of disturbances are reported. Finally, this work proposes recommendations on improving the RTI system analyzed.

Keywords: Real Time Information; Disruptions; Public Transport; AVL data

1. INTRODUCTION

Real time information (RTI) on public transport disturbances can significantly improve the travel experience of passengers. In a literature review, (Brakewood & Watkins, 2019) show the primary positive effects are the decreased waiting time, overall travel time, change in route choice and increased satisfaction with the transport system. RTI can be provided to passengers with different means, such as: a screen at a bus stop, voice alerts, or mobile applications. (Harmony & Gayah, 2017) identified that the preferred option for receiving RTI are mobile applications. Nevertheless, there is little knowledge on the effectiveness of mobile applications and social media to share RTI (Hu et al., 2018; Rahman et al., 2019).
Most of the works in literature exploits surveys or simulation to study how RTI influences passengers, in terms of travel behavior, route choice and waiting time (Akhla et al., 2022; Leng & Corman, 2020; Paulsen et al., 2021). However, according to (Papangelis et al., 2016), there is almost no evidence on RTI requirements for passengers during disruptions. In this sense, they recommend the information should be accurate, timely, and directed to passengers' needs, rather than generic.
Evaluating an RTI system is significantly beneficial for a service provider, to identify the main flaws of the system, and its effectiveness compared to other systems. However, despite the acknowledged importance of RTI systems, their performance in practice remains poorly understood, including their level of accuracy, which disturbances are notified, and which ones are ignored. In fact, a valuable RTI system should not only notify large disruptions, but also small disturbances, which may have a large impact on passengers (Marra & Corman, 2020), if they are not correctly informed (Marra & Corman, 2023).
This work proposes a methodological framework to analyze any text-based RTI system, and a set of metrics to evaluate it, in terms of correctness, amount of information provided and timeliness.

The core idea behind is the comparison of alerts sent to passengers with Automatic Vehicle Location data (AVL), describing all disturbances occurred in the network. The proposed methodology can be used to evaluate the performance of an RTI system, quantifying numerically some of its strengths and weaknesses. We apply the framework in a case study, the RTI system of Zurich, informing passengers on a smartphone application. Finally, we offer suggestions for improving the performance of the examined RTI system, which can also be applied to similar systems.

2. METHODS AND DATA

We analyze an RTI system, comparing the alerts sent to passengers with the disturbances occurred in the network. The comparison shows which disturbances are notified and how accurate is the information provided. Figure 1 shows the methodology to analyze the RTI system of Zurich. The same framework can be used to study any RTI system, adapting it to the information provided in the alert. The data and each step are described in the following sections.

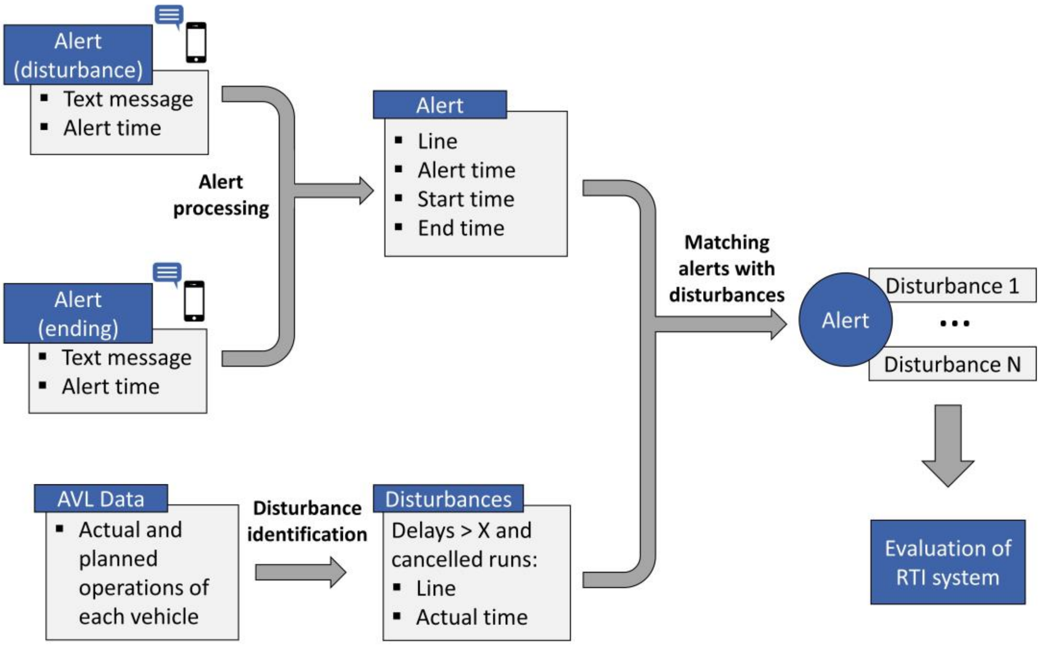

Figure 1: Methodology to analyze Zurich RTI system

Description of the alerts

We study the RTI system of the public transport network of Zurich, Switzerland, named “VBZ-alerto”. The system sends automatic alerts (text messages) on a smartphone app, Telegram, describing disturbances occurred in the city.

We collected 430 alerts between 11.11.2022 and 22.02.2023. Each alert is formed by a text and an “alert-time”, i.e. the time the alert is sent on the app. The alerts can be classified into three types: disturbance alerts (55%), notifying a disturbance; ending alerts (38%), notifying the end of a disturbance, previously notified; other alerts (6%), with unique text.

For a disturbance alert, the text follows the following structure:

“**Lines A, B, C:** description of disruption (unstructured)
dd.mm.yyyy hh:mm – indefinite
detailed description (unstructured)”.

Therefore, each alert contains information on the disturbed lines and the "starting-time", i.e. the time the disruption started according to the operator. The "ending-time" is indefinite for most of the alerts (90%). However, an ending alert starting with "Resolved" notifies the "ending-time" of a previous disturbance alert. 71% of disturbance alerts have an ending alert. Additional information on the disturbance, like stops involved, are not provided in a standard format, and thus cannot be derived automatically from the text.

In the following analyses, we consider only alerts with the ending specified, in the text or by a following alert. An alert involving more than one line is considered as a different alert for each line. In total, 332 alerts are analysed.

Actual disturbances

We identify all disturbances occurred (and not only the ones notified), from long-term AVL data of Zurich public transport. The AVL data contain the actual and planned times of each vehicle at each stop in Zurich. Therefore, they describe what actually occurred in the network. According to (Zhang et al., 2022), AVL data are the preferred source to detect disturbances, compared to passenger data, incident logs and social media.

The AVL data describe the actual disturbances occurred each day, in the form of delays and cancelled runs. Larger disturbances, such as an entire line cancelled, are described by a set of smaller disturbances, such as a series of cancelled runs. Therefore, in this work, we consider as a disturbance any delay >**D** and any cancelled run. We choose D=8 min, as it is the average headway in Zurich. Each disturbance has an "actual-time", indicating the time the disturbance occurred. For each run, we consider only the first disturbance. Namely, if a vehicle is delayed at several stops, the first delay is considered as the begin of the disturbance.

Finally, we also analyze lines not functioning for longer time, i.e. at least **N** consecutive cancelled runs (N=3 in our experiments).

Matching alerts with actual disturbances

The quality of an RTI system can be evaluated, comparing the set of alerts sent, with the actual disturbances (described in the AVL data). In fact, matching the alerts with the disturbances occurred in the network, allows to identify which disturbances are notified and which not.

We consider an alert matching a disturbance, if the following conditions hold:

- The disturbance and the alert concern the same line;
- The actual-time of the disturbance lies between the starting-time and ending-time of the alert, including a buffer **B** (10 min);

Each alert may match multiple disturbances, since, for instance, an alert notifying a cancelled line for an hour corresponds to multiple cancelled runs.

Evaluating the RTI system

We evaluate an RTI system, based on how good the alerts inform about the disturbances. We propose a set of metrics to evaluate the correctness, amount of information provided, and timeliness of the RTI system. The correctness can be measured as for a classification model, identifying which operations are considered disturbances, and therefore should be alerted. The performance can be evaluated as in Table 1. The disturbances matched by an alert are indeed informed to the passengers, and represent the True Positives of the system. All disturbances not matched by an alert are the False Negatives. Disturbances alerted but not available in the data (i.e. disturbances not actually occurred, or errors in the AVL data) are the False Positives. All normal operations (without disturbances and not alerted) are the True Negatives.

Table 1: Correctness of an alert system

	Alert sent	**No alert sent**
Disturbance in the AVL data	Actual disturbance notified [True Positive]	Actual disturbance not notified [False Negative]
No disturbance in the AVL data	Disturbance notified but not in the data [False Positive]	Normal operations [True Negative]

We measure the performance of the system in terms of precision and recall, as follows:

$$\boldsymbol{Precision} = True\ Positive / (True\ Positive\ +\ False\ Positive) \tag{1}$$
$$\boldsymbol{Recall} = True\ Positive / (True\ Positive\ +\ False\ Negative) \tag{2}$$

Precision and recall measure respectively the correctness of an RTI system and the percentage of disturbances notified to passengers. High precision means the alerts provide correct information to passengers, while low precision means the alerts provide wrong information. High recall indicates most of the disturbances are alerted, while low recall indicates the passengers are not informed about most of the disturbances. We remark each alert matches multiple disturbances; therefore, the precision should be computed as number of correct alerts, while the recall as number of alerted disturbances.
We evaluate the performance also in terms of timeliness of the alerts. For each alert, three time-related information are available on the start of the disturbance: “alert-time”, when the alert is sent; “starting-time”, when the alert says the disturbance started; “actual-time”, when the disturbance actually started, according to the data, i.e. the time of the first disturbance matching the alert. We evaluate the timeliness, based on three metrics:

$$\boldsymbol{Promptness}\ =\ alert\ time -\ starting\ time \tag{3}$$
$$\boldsymbol{Latency}\ =\ actual\ time -\ starting\ time \tag{4}$$
$$\boldsymbol{Reactivity}\ =\ alert\ time -\ actual\ time \tag{5}$$

The promptness represents how late the alert is sent to passengers, compared to the beginning of the disturbance, according to the operator. The latency represents how late the disturbance started, compared to when the alert says it started. The reactivity represents how late the alert is sent, compared to the actual beginning of the disturbance.

3. RESULTS AND DISCUSSION

During the study period, we analyzed 332 alerts, for an average of 2.29 lines per day alerted. In contrast, the actual disturbances are much more, with a total of 52097 disturbed runs, and an average per day of 401 disturbed runs and 48.5 disturbed lines. This large difference is expected, since we are considering both small and large disturbances.
Regarding the correctness of the RTI system of Zurich, we observed a precision of 98% and a recall of 12%. The very high precision shows that the alerts provide correct information, and that when an alert is sent to a passenger, the disturbance can be observed in the AVL data and therefore occurred in reality (if there are no errors in the AVL data). Only 6 alerts did not find a correspondence in the AVL data, which may be due to a wrong alert-time in the text of the alert, a measurement error in the AVL data, or a disturbance with a delay shorter than 8 min. The low recall shows that passengers are informed only of 12% of disturbances (48 runs on average per day). This is

detrimental for passengers, since they are not informed of most of disturbances they encounter during their trip. However, we remark that it is unrealistic for an RTI system to have a very high recall, since it is not possible to inform passengers of all disturbances. In fact, sending just a single alert per disturbed line is equal to an average of 48.5 alerts per day, which may be overwhelming for a passenger.

Among the notified disturbances, 91% are cancelled runs, while 9% are delays. Instead, among the not notified disturbances, 78% are cancelled runs, while 22% are delays. Therefore, passengers are more informed when a line is not running, than when it is delayed.

We also analyzed long-term cancelled lines, i.e. when at least 3 consecutive runs are cancelled. In this case, the recall is higher (30%), showing long-term cancellations are more frequently notified than single cancellations (12%).

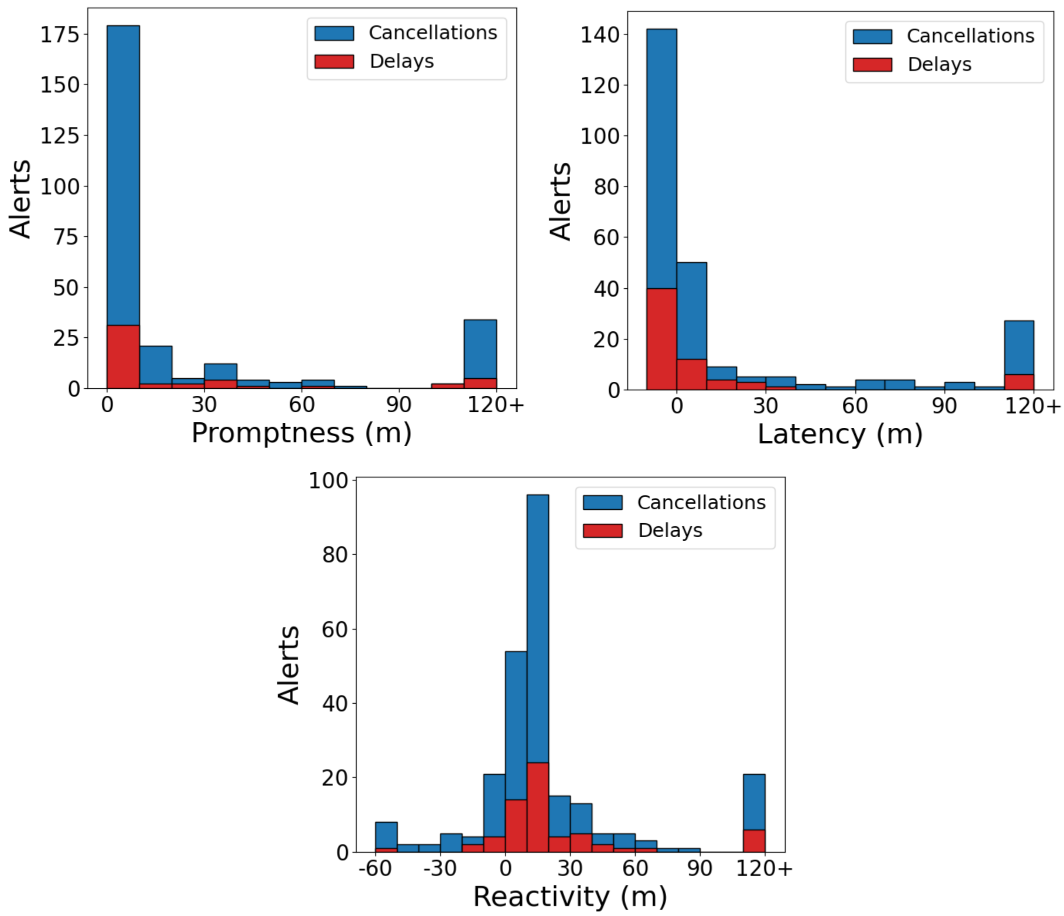

Figure 2: Promptness, Latency and Reactivity for the observed alerts

Figure 2 shows the timeliness of an RTI system based on three metrics, as defined in Section 3. For most of the alerts (73%) the promptness is below 10 minutes. This shows that most of the time the operators inform passengers in less than 10 minutes from the beginning of the disturbance, as written in the alert. This time may be adequate for long disturbances, while not for shorter ones (e.g. small delays), since informing passengers after 10 or more minutes may not be useful for them. For most of the alerts, the latency is between -10 and 10 minutes (77%), showing the information provided on the beginning of the disturbance is correct. However, for many alerts the latency is large (120+ minutes). This is often the case when the actual start is not known, thus it is reported in the early morning (e.g. 06:00). Finally, the reactivity has a wider range, showing the alerts are often sent much before or after the actual beginning of the disturbance.

The analyses and results shown suggest important recommendations to improve the quality of Zurich RTI system, which applies also to similar RTI systems. The disturbances occurring daily in the network are too many to be notified to all passengers indiscriminately. Therefore, personalized alerts are recommended to inform about all occurred disturbances (increasing the recall), without overwhelming the passengers with too many alerts. The Zurich RTI system provides correct information, in terms of disturbed line and time, for most of the alerts. However, further information on the involved stops, vehicles or type of disturbance are not always provided. This makes unclear to passengers if the disturbance is affecting them (maybe a passenger is at a non-disturbed stop). In this sense, enhancing the information quality can improve the passenger's reaction to disturbances and the overall travel experience. From the analysis of timeliness, we identified that the RTI system provides timely and accurate information on the starting time of a disturbance. However, the gap between the beginning of a disturbance and the notification to passengers can be reduced, especially for small disturbances, whose notification should be faster to be effective.

4. CONCLUSIONS

RTI systems are widely acknowledged as key contributors to passengers' travel experience, in case of public transport disturbances. Despite their recognized value, little research has investigated how to evaluate an RSI system, how accurate is the information provided and which disturbances are (or not) notified. This work answers these questions proposing a methodology to evaluate a text-based RTI system. The core idea is the comparison of alerts with AVL data, to identify which disturbances are notified to passengers and which not. Afterwards, a set of metrics is defined to assess the correctness and timeliness of the system. These metrics help to identify the drawbacks of an RTI system and the directions of improvement. We applied the proposed methods in a real test case in Zurich. The results identify that the alerts are highly precise and punctual, but they cover only a small fraction (12%) of all disturbances in the network.

For future work, the analyses can be extended in several directions. The disturbances can be divided into different categories (e.g. small delays, large disruptions). Analyzing how frequently those categories are notified may highlight which disturbances are prioritized and which are dismissed. Furthermore, the importance of different alerts can be estimated based on the effects of the disturbance on passengers, or the travel time saved thanks to the alert.

Finally, we remark this is an on-going study, and the dataset is increasing every day. Therefore, in the future we plan to study a much larger dataset, allowing more detailed analyses.

REFERENCES

Akhla, A. A. N., Ling, T. C., Shibghatullah, A. S., Mon, C. S., Cherukuri, A. K., Yen, C. L., & Yi, L. C. (2022). Impact of Real-Time Information for Travellers: A Systematic Review. *Journal of Information & Knowledge Management*, 2250065.

Brakewood, C., & Watkins, K. (2019). A literature review of the passenger benefits of real-time transit information. *Transport Reviews*, *39*(3), 327–356.

Harmony, X. J., & Gayah, V. V. (2017). Evaluation of Real-Time Transit Information Systems: An information demand and supply approach. *International Journal of Transportation Science and Technology*, *6*(1), 86–98.

Hu, M. W., Huang, W. K., & Chen, Y. (2018). Evaluating the Effects of Alleviating Urban Traffic Congestion Using Real-Time Traffic Information on Mobile Devices. In K. W. Chau, I. Y. S. Chan, W. Lu, & C. Webster (Eds.), *Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate* (pp. 689–700). Springer.

Leng, N., & Corman, F. (2020). The role of information availability to passengers in public transport disruptions: An agent-based simulation approach. *Transportation Research Part A: Policy and Practice*, *133*, 214–236.

Marra, A. D., & Corman, F. (2020). From Delay to Disruption: Impact of Service Degradation on Public Transport Networks. *Transportation Research Record*, *2674*(10), 886–897.

Marra, A. D., & Corman, F. (2023). How different network disturbances affect route choice of public transport passengers. A descriptive study based on tracking. *Expert Systems with Applications*, *213*, 119083.

Papangelis, K., Velaga, N. R., Ashmore, F., Sripada, S., Nelson, J. D., & Beecroft, M. (2016). Exploring the rural passenger experience, information needs and decision making during public transport disruption. *Research in Transportation Business & Management*, *18*, 57–69.

Paulsen, M., Rasmussen, T. K., & Nielsen, O. A. (2021). Impacts of real-time information levels in public transport: A large-scale case study using an adaptive passenger path choice model. *Transportation Research Part A: Policy and Practice*, *148*, 155–182.

Rahman, R., Roy, K. C., Abdel-Aty, M., & Hasan, S. (2019). Sharing Real-Time Traffic Information With Travelers Using Twitter: An Analysis of Effectiveness and Information Content. *Frontiers in Built Environment*, *5*.

Zhang, N., Graham, D. J., Bansal, P., & Hörcher, D. (2022). Detecting metro service disruptions via large-scale vehicle location data. *Transportation Research Part C: Emerging Technologies*, *144*, 103880.