Quality Control

Lessons Learned from Deployment and Evaluation of GTFS-realtime Feeds

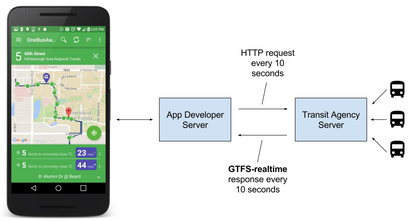
Sean J. Barbeau, Ph.D. – CUTR @ University of South Florida

Background

- Real-time transit info has many benefits, including shorter perceived and actual wait times[1], lower learning curve for new riders[2], increased ridership[3][7], and increased feeling of safety at night[5][6], and better perception of the agency[8].
- Nearly ¾ of users in one study said they relied solely on real-time info instead of schedule[6]
- However, accuracy is important – 9% of surveyed riders said they took the bus less often due to real-time errors

Open transit data

- General Transit Feed Specification (GTFS) has become de facto format for open schedule data, shared by over 1,500 agencies worldwide[9]
- GTFS-realtime feeds, which power mobile transit apps, are becoming more widely available, with over 50 agencies sharing data[10]



- GTFS-realtime v1.0 has lacked well-specified requirements[11] and validation tools
- This results in confusion and disagreements between transit agencies, Automatic Vehicle Location (AVL) vendors, and application developers as to what data should actually appear in a GTFS-rt feed
  - Increases the time, effort, and cost to deploy a new GTFS-rt feed
  - Errors in real-time info affect riders and operations

GTFS-realtime example

- 3 feed types – **TripUpdates** contain delay information, **VehiclePositions** show where vehicle is, and **Alerts** are human-readable impacts on service

trip_id	arrival_time	departure_time	stop_id	stop_sequence
2777	5:52:00	5:52:00	4301	1
2777	5:52:34	5:52:34	3471	2
2777	5:53:46	5:53:46	4456	3
2777	5:54:27	5:54:27	592	4
2777	5:55:11	5:55:11	593	5
2777	5:55:20	5:55:20	4457	6
2777	5:55:40	5:55:40	595	7
2777	5:56:34	5:56:34	596	8

GTFS stop_times.txt

```
trip_update {
  trip {
    trip_id: "2777"
  }
  stop_time_update {
    stop_sequence: 3
    arrival {
      delay: 60  // 60 seconds late
    }
    stop_id: "4456"
  }
```

GTFS-realtime TripUpdate

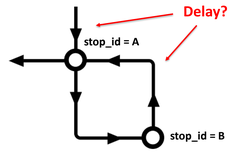
GTFS-realtime v1.0 problems

- Too many optional fields – 56 out of 63 (89%)
  - Quirk of Protocol Buffer docs (for details see http://bit.ly/gtfs-realtime-2)
- Missing fields lead to poor quality data

```
header {
 gtfs_realtime_version: "1.0"
}
entity {
 id: "d131dd02"
 vehicle {
  position {
    latitude: 28.04265
    longitude: -82.45945
  }
...
```

No timestamp for VehiclePosition

```
trip {
 trip_id: "277725"
}
stop_time_update {
 arrival {
  delay: 900 // 15 minutes
 }
 stop_id: "A"
}
```



No stop_sequence – where is 15 min delay?

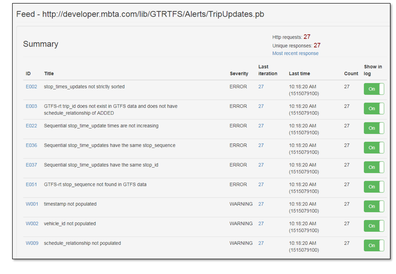
GTFS-realtime v2.0

- Defines new **transit-specific** field requirements:
  - *Required*
  - *Optional*
  - *Conditionally required* (see *Description* field for when this field is required)

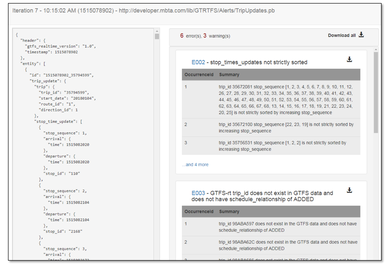
Field Name	Type	Required	Cardinality	Description
stop_sequence	uint32	Conditionally required	One	Must be the same as in stop_times.txt in the corresponding GTFS feed. Either stop_sequence or stop_id must be provided within a StopTimeUpdate - both fields cannot be empty. stop_sequence is required for trips that visit the same stop_id more than once (e.g., a loop) to disambiguate which stop the prediction is for.

GTFS-realtime validator

- Developed using experience from OneBusAway mobile app deployment at new transit agencies using GTFS-realtime
- Open-source - https://github.com/CUTR-at-USF/gtfs-realtime-validator
- Alpha version hosted at http://transittools.forest.usf.edu



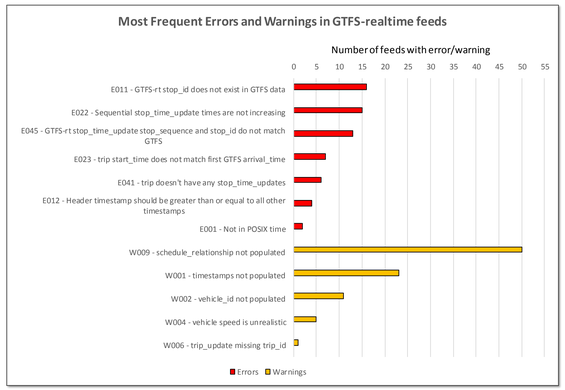
Shows summary of all errors/warnings



Shows message with all error occurrences

Evaluation of industry feeds

- Created open-source tool to batch validate 78 out of 130 GTFS-realtime feeds catalogued on TransitFeeds.com
  - https://github.com/CUTR-at-USF/transit-feed-quality-calculator
- 69% (54) feeds had errors, and 74% (58) had warnings
- Recommend that all agencies produce GTFS-realtime v2.0 feeds, validate feeds prior to accepting feed from vendor



Acknowledgements

Funded by the National Center for Transit Research and the Florida Department of Transportation (FDOT), National Institute for Transportation and Communities (NITC), and Google Summer of Code. The opinions, findings and conclusions expressed in this publication are those of the author(s) who are solely responsible for the facts and accuracy of information and not necessarily those of the Florida Department of Transportation or the U.S. Department of Transportation. The author would also like to thank FDOT District 7, Hillsborough Area Regional Transit and Pinellas Suncoast Transit Authority, as well as Cambridge Systematics that supports OneBusAway for HART, Clever Devices, PSTA's AVL vendor, Cagri Cetin for his invaluable work troubleshooting feeds, Mohan Gandhi Achchakkagari, Nipuna Gunathilake, and Surya Vamshi Kandukoori for the software development work on the GTFS-realtime Validator and data collection and graphs for errors in production GTFS-rt feeds.

[1] Kari Edison Watkins, Brian Ferris, Alan Borning, G. Scott Rutherford, and David Layton (2011), "Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders," *Transportation Research Part A: Policy and Practice,* Vol. 45 pp. 839-848.
[2] C. Cluett, S. Bregman, and J. Richman (2003). "Customer Preferences for Transit ATIS," Federal Transit Administration. Available at http://ntl.bts.gov/lib/jpodocs/repts_te/13935/13935.pdf#sthash.jwn5Oltr.dpuf
[3] Lei Tang and Piyushimita Thakuriah (2012), "Ridership effects of real-time bus information system: A case study in the City of Chicago," *Transportation Research Part C: Emerging Technologies,* Vol. 22 pp. 146-161.
[4] Aaron Steinfeld and John Zimmerman, "Interviews with transit riders in San Francisco and Seattle," ed, 2010.
[5] Brian Ferris, Kari Watkins, and Alan Borning (2010), "OneBusAway: results from providing real-time arrival information for public transit," in Proceedings of the 28th International CHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA, pp. 1807-1816.
[6] A. Gooze, K. Watkins, and A. Borning (2013), "Benefits of Real-Time Information and the Impacts of Data Accuracy on the Rider Experience," in Transportation Research Board 92nd Annual Meeting, Washington, D.C., January 13, 2013.
[7] Brakewood, Macfarlane and Watkins (2015). The Impact of Real-Time Information on Bus Ridership in New York City. Transportation Research Part C: Emerging Technologies, Volume 53, pp. 59-7
[8] C. Brakewood, S. Barbeau, and K. Watkins (2014), "An experiment evaluating the impacts of real-time transit information on bus riders in Tampa, Florida," Transportation Research Part A: Policy and Practice, Vol. 69 pp. 409-422.
[9] MapZen. "TransitLand - An Open Project - For Data Providers." Accessed July 31, 2017 from https://transit.land/an-open-project/
[10] TransitFeeds.com "Search Results." Accessed January 3, 2018 from http://transitfeeds.com/search?q=gtfsrt
[11] GTFS-realtime Google Group. "Proposal: Make FeedHeader.timestamp a required field." Accessed January 2015 from https://groups.google.com/forum/#!msg/gtfs-realtime/wm3W7QJEZ9Y/DLyWKkknJyoJ

CUTR CENTER FOR URBAN TRANSPORTATION RESEARCH
USF UNIVERSITY OF SOUTH FLORIDA
NITC NATIONAL INSTITUTE for TRANSPORTATION and COMMUNITIES

barbeau@cutr.usf.edu

Paper #18-05585