GTFS-realtime v2.0

Sean J. Barbeau, Ph.D.
Principal Mobile Software Architect for R&D
Center for Urban Transportation Research
University of South Florida
Why real-time transit info?

• Real-time information (RTI) has **many benefits**
  – Shorter perceived wait time [1][8]
  – Shorter actual wait time [1]
  – Lowers learning curve for new riders [2]
  – Increased feeling of safety (e.g., at night) [5][6]
  – Improved perception of agency [8]
  – Increased ridership [3][7]

• **Improve the rider experience** without increasing vehicle frequency

• **Emerging analytics tools** are powered by this data

---

• **GTFS-realtime** is becoming de facto standard
  – **Over 50 agencies** now have GTFS-realtime feeds!
Quality is important!

- In one study, **9% of riders** said they took the bus **less often** due to errors in RTI[9]

- GTFS-realtime v2.0 will help agencies produce better quality RTI

---

WHAT’S WRONG WITH GTFS-REALTIME V1.0?
Problem with GTFS-realtime v1.0

• A LOT of *optional* fields – 56 out of 63 (89%)
  – Quirk of Protocol Buffer docs (for details see http://bit.ly/gtfs-realtime-2)

• Leads to sub-optimal feeds
  – Poor data quality
  – Bad rider experience
  – Inaccurate analytics – garbage in, garbage out
Example 1 – Vehicle Position

• All timestamps are optional
• When was position calculated?

```json
header {
  gtfs_realtime_version: "1.0"
}
entity {
  id: "d131dd02"
  vehicle {
    position {
      latitude: 28.04265
      longitude: -82.45945
    }
  }
}
```
Example 2 – Loop route

- stop_sequence field is optional
- Is 15 minute delay before or after stop_id B?

```javascript
trip {
    trip_id: "277725"
}
stop_time_update {
    arrival {
        delay: 900 // 15 minutes
    }
    stop_id: "A"
}
```
Industry-wide GTFS-realtime v1.0 Validation Results

- ~75% of feeds we analyzed had errors
Clarifying what’s really optional

GTFS-REALTIME V2.0
GTFS-realtime v2.0

• Defines new transit-specific requirements

• Each field is labeled as either:
  – Required
  – Optional
  – Conditionally required
    • See Description field for when this field is required
message StopTimeUpdate

Realtime update for arrival and/or departure events for a given stop on a trip. Please also refer to the general discussion of stop time updates in the TripDescriptor and trip updates entities documentation.

Updates can be supplied for both past and future events. The producer is allowed, although not required, to drop past events. The update is linked to a specific stop either through stop_sequence or stop_id, so one of these fields must necessarily be set. If the same stop_id is visited more than once in a trip, then stop_sequence should be provided in all StopTimeUpdates for that stop_id on that trip.

Fields

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Type</th>
<th>Required</th>
<th>Cardinality</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>stop_sequence</td>
<td>uint32</td>
<td>Conditionally required</td>
<td>One</td>
<td>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.</td>
</tr>
<tr>
<td>stop_id</td>
<td>string</td>
<td>Conditionally required</td>
<td>One</td>
<td>Must be the same as in stops.txt in the corresponding GTFS feed. Either stop_sequence or stop_id must be provided within a StopTimeUpdate - both fields cannot be empty.</td>
</tr>
</tbody>
</table>
What’s new in GTFS-realtime v2.0
Better quality real-time information

Read more...

Quick analysis of feeds

GTFS-REALTIME VALIDATOR
GTFS-realtime Validator

• Test your v1 and v2 feeds for errors
  – Live demo - http://transittools.forest.usf.edu/

Feed - http://developer.mbta.com/lib/GRTFSS/Alerts/TripUpdates.pb

Summary

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Severity</th>
<th>Last iteration</th>
<th>Last time</th>
<th>Count</th>
<th>Show in log</th>
</tr>
</thead>
<tbody>
<tr>
<td>E002</td>
<td>Unsorted stop_sequence</td>
<td>ERROR</td>
<td>2</td>
<td>05:49:25 PM (1494366565)</td>
<td>2</td>
<td>On</td>
</tr>
<tr>
<td>E022</td>
<td>Sequential trip stop_time_update times are not increasing</td>
<td>ERROR</td>
<td>3</td>
<td>05:49:34 PM (1494366574)</td>
<td>3</td>
<td>On</td>
</tr>
<tr>
<td>W001</td>
<td>Timestamp not populated</td>
<td>WARNING</td>
<td>3</td>
<td>05:49:34 PM (1494366574)</td>
<td>3</td>
<td>On</td>
</tr>
</tbody>
</table>
GTFS-rt Validator – View data & errors
Evaluation of industry feeds

- Created open-source tool to batch validate 78 out of 130 GTFS-realtime feeds catalogued on TransitFeeds.com

- 69% (54) feeds had errors, and 74% (58) had warnings
Most Frequent Errors and Warnings in GTFS-realtime feeds

<table>
<thead>
<tr>
<th>Error Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E011</td>
<td>GTFS-rt stop_id does not exist in GTFS data</td>
</tr>
<tr>
<td>E022</td>
<td>Sequential stop_time_update times are not increasing</td>
</tr>
<tr>
<td>E045</td>
<td>GTFS-rt stop_time_update stop_sequence and stop_id do not match GTFS</td>
</tr>
<tr>
<td>E023</td>
<td>trip start_time does not match first GTFS arrival_time</td>
</tr>
<tr>
<td>E041</td>
<td>trip doesn’t have any stop_time_updates</td>
</tr>
<tr>
<td>E012</td>
<td>Header timestamp should be greater than or equal to all other timestamps</td>
</tr>
<tr>
<td>E001</td>
<td>Not in POSIX time</td>
</tr>
<tr>
<td>W009</td>
<td>schedule_relationship not populated</td>
</tr>
<tr>
<td>W001</td>
<td>timestamps not populated</td>
</tr>
<tr>
<td>W002</td>
<td>vehicle_id not populated</td>
</tr>
<tr>
<td>W004</td>
<td>vehicle speed is unrealistic</td>
</tr>
<tr>
<td>W006</td>
<td>trip_update missing trip_id</td>
</tr>
</tbody>
</table>

Number of feeds with error/warning

What’s next for Transit Agencies?

• Require that all AVL vendors provide GTFS-realtime v2.0 feeds (especially in RFPs)
  – Github is official new home of GTFS/GTFS-realtime specs - [https://github.com/google/transit](https://github.com/google/transit)

• Run GTFS-realtime validator frequently

• Also require that vendors follow GTFS Best Practices
What’s next for GTFS-rt Community?

• Create GTFS-realtime Best Practices
  – “Warning” from GTFS-realtime validator\(^1\)
  – Proposals without unanimous agreement
  – Other community input

• Clarify more GTFS-realtime gray areas
  – Either new proposals, or in best practices

• Better targeted documentation
  – More focused on use cases and features (e.g., ability to cancel trips)

What’s next for GTFS-rt Community?

- Continue to add new rules to GTFS-realtime Validator
- Hosting GTFS-realtime Validator as a service for agencies and vendors
  - Alpha at http://transittools.forest.usf.edu
- Tackle GTFS-sometimes-ish use cases
  - Different endpoint for major changes to trips
- Develop additional open-source tooling for prediction generation
  - Including data warehousing
  - Leverage TheTransitClock (formerly Transitime)
    - http://thetransitclock.org
Thanks!

Sean J. Barbeau, Ph.D.
barbeau@cutr.usf.edu
813.974.7208

Funding from:

WHY SO MANY OPTIONAL FIELDS?
GTFS-rt uses Protocol Buffer (PB) format

- *gtfs-realtime.proto* PB file defines elements to exchange
Protocol Buffers save space

- Compressed binary is around 6 times smaller than plain text
PB required ≠ transit required

• v1.0 - Optional/required for Cardinality field values were copied from .proto file

```protobuf
message FeedHeader {
  // Version of the feed specification.
  // The current version is 2.0.
  required string gtfs_realtime_version = 2;

  // Determines whether the current fetch is incremental. Currently,
  // DIFFERENTIAL mode is unsupported and behavior is unspecified for feeds
  enum Incrementality {
    FULL_DATASET = 0;
    DIFFERENTIAL = 1;
  }
  optional Incrementality incrementality = 2 [default = FULL_DATASET];

  // This timestamp identifies the moment when the content of this feed has been
```