# Telnyx CDR : Demisfying Mispriced Calls

## Problem Statement :

The billing of calls on Telnyx and vendor systems can result in discrepancies due to errors on either side, such as errors in billed duration, billed LRN, rounding in billed amount, or the use of the wrong rate deck. This can lead to mis-priced calls and billing discrepancies that need to be investigated by comparing CDRs on both sides.

For example, a call may be billed with the correct duration but the wrong rate, resulting in differences in the billed amount between Telnyx and the vendor. These discrepancies can result in revenue losses for Telnyx and impact customer satisfaction.

To address this issue, a program is needed that can read in Telnyx CDRs, vendor CDRs, and auxiliary data files to identify the root cause of mis-priced calls. The program should be able to compare Telnyx CDRs with vendor CDRs and auxiliary data files to find discrepancies in pricing, call duration, or other relevant metrics. Statistical methods or machine learning algorithms can be used to identify patterns and outliers in the data, and help pinpoint the root cause of the mis-priced calls, whether it's a technical issue or a human error.

## Assumptions

1. Each call can be identified by the combination of caller, callee, and timestamp. We ignored duplicate ids which is roughly 0.0175%
2. Timestamps are recorded to a reasonable proximity. We set all timestamps to the nearest tens second place.
3. Errors only occur in the aforementioned buckets, such as errors in billed duration, billed LRN, rounding in billed amount, or the use of the wrong rate deck.
4. Telnyx's CDRs format does not change overtime.
5. CDRs can have multiple errors in each record.

## Exploratory Data Analysis

| Metrics | Telnyx | Vendor |
|---|---|---|
| Total CDRs | 9.983 M | 9.998 M |
| Callee Count | 4.424 M | 4.427 M |
| Caller Count | 1577 | 1579 |
| Callee LRNs | 93131 | 19905 |

Preprocessing the data was done in separate segments:

1. Some of the IDs needed slicing to standardize the format in both tables.
2. Generated a unique ID for a CRD based on the combination of caller ID, callee ID, and timestamp. The generate_id() function concatenates the values into a single string, which is then hashed using the SHA-256 algorithm resulting in a unique combination of values passed into the function.
3. Mapped with the area codes to find out whether the call is Inter/Intra State.
4. 0.0175% & 0.0068% CDRs from Vendor's and Telnyx's have duplicate IDs. Ignored these as they are less significant for the investigation.

## Errors

Different Type of Errors in CDRs:

| Error Type | Count Errors |
|---|---|
| LRN Error | 5473703 |
| Rounding Error | 5065820 |
| Wrong Rate Error | 1176032 |
| Duration Error | 256033 |

## LRN Errors:

Approximately 56.11% LRNs in the Telnyx Table has error in Callee LRN

| | caller | callee | callee_LRN | timestamp | billedduration | billed_rate | billed_cost | id |
|---|---|---|---|---|---|---|---|---|
| 3 | 9785108014 | 9542510030 | 786930 | 2017-09-11 14:30:01 | 6 | 0.00354 | 0.0004 | d3849d92155d4eab |

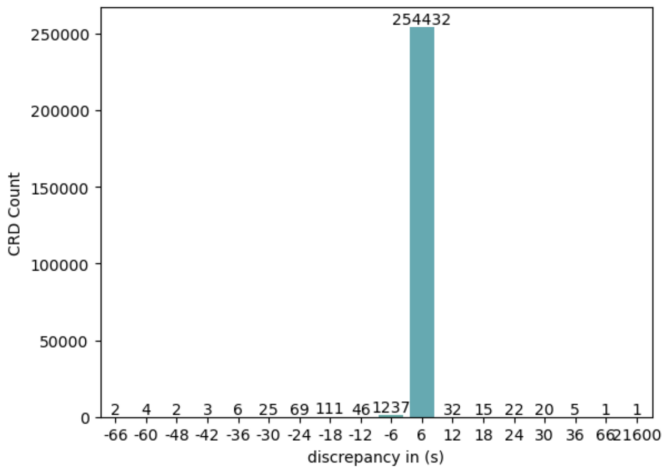
See that in the Telnyx table, callee LRN parsings are wrong. It needs to be corrected.

## Rounding Errors:

Out of 9.983M Telnyx CRDs, ~190K have rounding errors which is 1.9% accumulating bill difference of 18.93$. On the other hand, Out of 9.998M Vendor CRDs, ~4954K have rounding errors which is 49.55% accumulating difference of 495.44$.
Thereby, Our analysis shows that there were technical issues with the vendor's rounding system that resulted in incorrect pricing calculations.

## Bill Duration Errors:

Among 9.75M CRDs, Bill Duration Discrepancies was found in 255K events which is 2.62% of the merged unique CRDs. Most of the discrepancies associated with bill duration are with 6s difference for 254K events. In the erroneous CRDs, 35.32% have missing Callee Area Code & Max Billed Duration Discrepancies are from Inter State Calls.




Thereby, our analysis suspects that there might be some kind of latency issue for which our calls have such small bill duration errors.

**Wrong Rate Errors:**

This error is responsible for 12.06% of all errors (1176K) in the dataset. Within this category of errors, it seems that the majority of errors (~70%) occurred for inter-state calls, with 823K calls affected. This suggests that there may be some issues with how Telnyx is selecting rate decks for inter-state calls.

Further analysis could be done to identify specific patterns or trends within the inter-state calls that were affected by wrong deck errors ,if specific rate plans were provided. This could help to pinpoint the root cause of the errors and identify specific steps that Telnyx could take to prevent them from happening in the future. Additionally, it may be useful for Telnyx to review its rate deck selection process and ensure that it is using the appropriate deck for each call based on the destination and other relevant factors.

**Error Combinations :**

| Error Combinations | Error Counts |
|---|---|
| ('LRN Error',) | 2321.48K |
| ('LRN Error', 'Rounding Error') | 2282.14K |
| ('Rounding Error',) | 1714.49K |
| ('LRN Error', 'Wrong Rate Error', 'Rounding Error') | 516.05K |
| ('Wrong Rate Error', 'Rounding Error') | 298.25K |
| ('LRN Error', 'Wrong Rate Error') | 191.70K |
| ('LRN Error', 'Duration Error', 'Rounding Error') | 142.24K |
| ('Wrong Rate Error',) | 138.15K |
| ('Duration Error', 'Rounding Error') | 81.90K |
| ('LRN Error', 'Duration Error', 'Wrong Rate Error', 'Rounding Error') | 19.51K |
| ('Duration Error', 'Wrong Rate Error', 'Rounding Error') | 11.24K |
| ('LRN Error', 'Duration Error', 'Wrong Rate Error') | 0.58K |
| ('Duration Error', 'Wrong Rate Error') | 0.56K |
| ('Duration Error',) | 0.00K |

The table shows the frequency of different error combinations in the Telnyx Call Detail Records (CDRs). The most common error combination is LRN Error and Rounding Error, followed by LRN Error and Wrong Rate Error. It is notable that there are a significant number of CDRs with all three errors (LRN Error, Wrong Rate Error, and Rounding Error), suggesting that these errors may be related in some way.