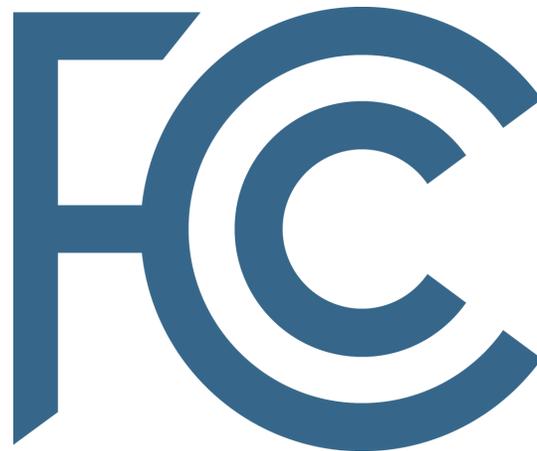# Measuring Broadband America Program

2018 Mobile Measurement

Open Data Release

Technical Description Addendum

Office of Engineering and Technology

I.   Description

This 2018 Mobile Measurement Open Data Release Technical Description Addendum ("Addendum") supplements the Measuring Broadband America Program ("MBA Program") '2016 Mobile Data Release Technical Description that provides detailed background technical information on the process by which the Mobile MBA Program's FCC Speed Test App ("the App" or "client")[1] collects crowd-sourced data from volunteers' mobile broadband Internet access services (BIAS), analysis methodology and privacy analysis techniques relevant for 2016 Open Data Release. This Addendum describes the application of the process developed in 2016 for the processing of data collected from July 1, 2018 to October 1, 2018 to make the time, location or other information less specific; limit the combinations of information; or deleting information.

---

[1] The FCC Speed Test App is available for both Android devices from Google Play, https://play.google.com/store/apps/details?id=com.samknows.fcc&hl=en, and for iOS devices from the Apple App Store, https://itunes.apple.com/us/app/fcc-speed-test/id794322383?mt=8.

APPENDIX A: PRIVACY ANALYSIS OF LOW FREQUENCY DATA

Row level results for this data export were selected by a cross-table primary key, *submission_id*, that included:

- A combination of SIM and network operator code and names to denote a connection to a US carrier by
    - A sim operator code matching a properly formatted mobile country code for the United States and no network_operator_code; or
    - A properly formatted network operator code (NOC) matching SIM mobile country code (MCC) and NULL NOC; a partial NOC matching a MCC and a match for a major four carrier name or a valid MCC SIM; and
- Time between '2016-07-01 00:00:06' and '2018-10-01 00:00:00'

In addition any submission_id with a valid cma_id location is included in the scope of data to be exported.

Three methods are used to identify low-frequency combinations of data in tables and across tables that will be targets for analysis and processing. First all submission_ids found in the curr_submission table that are unique by hour, model, os_version, sim_operator_code and submission_type are flagged. Both the same and different models may have different os_versions so checks for os_version must be done separately from model. In addition, while init and scheduled or manual tests may occur together or apart they must be treated in the strictest sense as a single low-frequency batch of such tests could imply a single handset produced the handsets. Second all submission_ids identified by a cross-table query of all potentially sensitive values collated to identify any unique instances of a combination of the following fields: submission_type, date truncated by hour, model, os_version,cma_id,network_operator_name, network_operator_code, sim_operator_name, sim_operator_code. Finally, unique occurrences of a test by hour in the active metric tables is identified and added to the list of submission_ids to be processed. These low-frequency samples are excluded from the set of "clean" and subject to analysis and processing.

Low-frequency samples are successively aggregated at levels that preserve the most temporal, location or other features. Coarsening samples that may have passed earlier aggregations in order to pool larger numbers of samples in time or space was also considered. Successive aggregations all preserved research value of the underlying active test measurements and other supporting data while sacrificing more sensitive features such as the device model.

Unrelated to low-frequency sample analysis, the data set was also analyzed for risks implied by l-diversity. Location aspects of the data are a potential quasi-identifier and in each CMA geography the population, population density, number of households and businesses were reviewed. In addition, the cross-table primary key, *submission_id* created by the SamKnows ingress software was replaced with a randomized arbitrary bigint list.

Table Name	Table References	

exportids		- Table to define the scope of measurements for the export identifies a combination of sim and network operator code and names to denote a connection to a US carrier by - A sim operator code matching a properly formatted mobile country code for the United States and no network_operator_code; or - A properly formatted network operator code matching SIM MCC and NULL NOC; a partial NOC matching a MCC and a match for a major four carrier name or a valid MCC SIM; and
		`create table exportids as  (select submission_id  from curr_networkdata  a where`
		`a.localdtime between '2013-11-01 00:00:00' and '2016-07-01 00:00:00' and (`
		`(sim_operator_code ~ '(31[106])(...)' and network_operator_code ~ '(^$)|(null)' ) or`
		`(network_operator_code ~ '(31[016])(...)' or`
		`   (network_operator_code ~ '(31[016])' and`
		`     (sim_operator_code ~ '(31[016])(...)' or  (network_operator_name ~`
		`'(^[\s]*at[\s]*[&]*[\s]*t[\s]*)|(^[^|\s]*t[\s\-]*[\s]*mobile)|(verizon)|(sprint[\s]*$)' )  )  )       )) ) ;`
		
		`SELECT 31351972`
		
		`--OLD --- SELECT 71172918`

exportids2	curr_location,
	exportids
		Set of all measurements within scope of the data export
		
		`create table exportids2 as (select (submission_id) from ((select distinct submission_id from ` **`curr_location where cma_id is not null`**`)  union (select `**`distinct submission_id from exportids`**`) ) as c);`
		
		`SELECT 11560305`
		
		`mba=# select count (distinct submission_id) from exportids2 ;`
		`  count`
		`----------`
		` 11560305`

exportids_clean	exportids2, pr_exclude	**Set of all measurements within scope of the data export that in combination did not yield records that were identified in the privacy analysis for review**
		
		`create table `**`exportids_clean`**` as (select distinct e.submission_id from exportids2 e left `**`pr_exclude p on e.submission_id = p.submission_id where p.submission_id is NULL);`**
		
		`select count ( submission_id) from exportids_clean ;`
		`  count`
		`---------`
		` 5779821`
		`(1 row)`
		
		`mba=# select count (distinct submission_id) from exportids_clean ;`
		`  count`
		`---------`
		` 5779821`
		`(1 row)`

pr_exclude	Prsubmissionuniq_v2ids, pr9v2ids	Set of low frequency measurements the focus of the privacy analysis. Measurements in combination of Device

		Model, Operating System version, SIM and Network Operator Code and Name, Test Submission Type and Date and Time to the Hour yield low frequency of test execution. Base table for defining the submission_ids to be processed before export. ``` mba=# create table pr_exclude as (select submission_id, date from ((select s.submission_id, date_trunc as date from prsubmissionuniq_v2ids s) union (select s.submission_id , date from pr9v2ids s)) as foo); ``` ``` SELECT 6827028 ```
`prsubmissionuniq_v2ids`	`Prsubmissionuniq_v2,` curr_submission	Set of measurements that are unique by all relevant features in the submission table for the measurements within scope of the data export create table prsubmissionuniq_v2ids as (select s.submission_id , p.* from prsubmissionuniq_v2 p inner join curr_submission s on p.date_trunc = date_trunc('hour',s.localdtime) and p.model = s.model and p.sim_operator_code =s.sim_operator_code and p.os_version = s.os_version and p.submission_type = s.submission_type); SELECT 4248097 mba=# select count (distinct submission_id) from prsubmissionuniq_v2ids; count --------- 4248097
`prsubmissionuniq_v2`	**exportids2,** **curr_submission**	Group by aggregation query to identify what features yield unique test executions by all relevant features in the submission table for

| | | the measurements within scope of the data export, including Device Model, Operating System version, SIM Operator Code, Test Submission Type and Date and Time to the Hour<br><br>```
create table prsubmissionuniq_v2 as  (select
count (*) as tally , date_trunc('hour' ,
s.localdtime)         ,   s.model        ,
s.sim_operator_code , os_version,
submission_type  from exportids2 e inner
join   curr_submission s on e.submission_id
= s.submission_id group by
date_trunc('hour' , s.localdtime)         ,
s.model         ,  s.sim_operator_code ,
os_version , submission_type having
count(*) =1 ) ;
```<br><br>SELECT 3201553<br><br>select count(*) from prsubmissionuniq_v2 ;<br>count<br>---------<br>3201553 |
| `pr9v2ids` | `pr9v2,`<br>`curr_submission,`<br>`curr_networkdata,`<br>`curr_location` | Set of measurements that are unique by all relevant features in the submission, networkdata, and location tables that are most relevant for privacy review for the measurements within scope of the data export including including Device Model, Operating System version, SIM and Network Operator Code and Name, Test Submission Type, CMA location ID, and Date and Time to the Hour<br><br>```
mba=# create table pr9v2ids as (select
s.submission_ID, s.submission_type, p.date
, p.model, p.os_version, p.cma_id,
p.network_operator_name,
p.network_operator_code,
p.sim_operator_name, p.sim_operator_code
from pr9v2 p inner join curr_submission s on
p.date = date_trunc ('hour', s.localdtime)
and p.submission_type = s.submission_type
and p.model = s.model and
``` |

| | | ```
p.os_version=s.os_version and
p.sim_operator_code = s.sim_operator_code
inner join curr_networkdata n on
s.submission_id = n.submission_id and
p.network_operator_code =
n.network_operator_code and
p.network_operator_name =
n.network_operator_name and
p.sim_operator_name = n.sim_operator_name
inner join curr_location l on
s.submission_id = l.submission_id and
p.cma_id = l.cma_id );

SELECT 7567382
``` |
| `pr9v2` | `exportids2, curr_submission, curr_location, curr_networkdata` | Group by aggregation query to identify the features yielding a low frequency test execution by all relevant features in the submission table for the measurements within scope of the data export

mba=# create table pr9v2 as (select count(distinct s.submission_id) , submission_type, date_trunc('hour', s.localdtime) as date ,   s.model, s.os_version, l.cma_id, n.network_operator_name, n.network_operator_code, n.sim_operator_name, n.sim_operator_code from **exportids2** e inner join **curr_submission** s on s.submission_id = e.submission_id inner join **curr_location** l on e.submission_id = l.submission_id inner join **curr_networkdata** n on e.submission_id = n.submission_id group by date_trunc('hour', s.localdtime), s.model, l.cma_id, n.network_operator_name, n.network_operator_code, n.sim_operator_name, n.sim_operator_code, submission_type , os_version having count (*) < 3) ;
SELECT 3963676 |

`pr9v2_exclude_ids_DISTINCT`	`PR9V2_EXCLUDE_IDS`	Table of aggregation of features for set measurements that are unique by all relevant features in the submission, networkdata, and location tables. Source table for analysis of the low frequency measurements to identify new aggregations on Device Model, Operating System version, SIM and Network Operator Code and Name, Test Submission Type, CMA location ID, and Date and Time to produce aggregations of k-anonymity of two or greater. This table is a "distinct" operation to remove redundant submission_ids.This table is the base for the first round of aggregations and the list of all pr_exclude IDs.
		
		`create table `**`pr9v2_exclude_ids_DISTINCT`**` as ((SELECT * FROM `**`PR9V2_EXCLUDE_IDS`**` ) UNION (SELECT * FROM `**`PR9V2_EXCLUDE_IDS`**`) );`
		
		`SELECT 4546816`
		
		`-- old analysis --SELECT 2947142`

`pr9v2_exclude_ids`	`pr9v2_exclude,`
	`curr_submission,`
	`curr_location,`
	`curr_networkdata`
		Source table for pr9v2_exclude_ids_DISTINCT identifying aggregation characteristics in **pr9v2_exclude** that match the relevant features in the cross-table join for the relevant tables.
		
		`create table pr9v2_exclude_ids as (select s.submission_ID, s.submission_type, p.date , p.model, p.os_version, p.cma_id, p.network_operator_name, p.network_operator_code, p.sim_operator_name, p.sim_operator_code from pr9v2_exclude p inner join curr_submission s on p.date = date_trunc ('hour', s.localtime) and p.submission_type = s.submission_type and p.model = s.model and`

| | | ```
p.os_version=s.os_version and
p.sim_operator_code = s.sim_operator_code
inner join curr_networkdata n on
s.submission_id = n.submission_id and
p.network_operator_code =
n.network_operator_code and
p.network_operator_name =
n.network_operator_name and
p.sim_operator_name = n.sim_operator_name
inner join curr_location l on
s.submission_id = l.submission_id and
p.cma_id = l.cma_id );

SELECT 29582157



--- old query --- SELECT 4750791


--OLD QUERY -- SELECT 17622902
``` |
| **pr9v2_exclude** | **Pr_exclude, curr_submission, curr_locationcurr_networkdata** | Table of aggregation of features for set measurements that are unique by all relevant features in the submission, networkdata, and location tables.  Source table for analysis of the low frequency measurements to identify new aggregations on Device Model, Operating System version, SIM and Network Operator Code and Name, Test Submission Type, CMA location ID, and Date and Time<br><br>```
create table pr9v2_exclude as (select
count(distinct s.submission_id)  ,
submission_type, date_trunc( 'hour',
s.localdtime) as date ,   s.model,
s.os_version, l.cma_id,
n.network_operator_name,
n.network_operator_code,
n.sim_operator_name, n.sim_operator_code
from pr_exclude e inner join curr_submission
s on s.submission_id = e.submission_id
inner join  curr_location l on
e.submission_id = l.submission_id inner
join curr_networkdata n on e.submission_id =
n.submission_id  group by date_trunc(
'hour', s.localdtime),  s.model, l.cma_id,
n.network_operator_name,
n.network_operator_code,
n.sim_operator_name, n.sim_operator_code,
submission_type  , os_version ;
``` |

| | | ```
SELECT 4750791

--- Old Analysis SELECT 4616355
``` |
| **pr10v2ids_exclude** | **Pr10v2_exclude, pr9v2_exclude_ids_DISTINCT** | This table holds 2.2 million aggregated low-frequency scheduled tests samples grouped by Year/Month/Period, model, os version, CMA and a whitespace and case insensitive matching of SIM and network operator names and codes. 2.2 million test samples achieve a k-anonymity of two or greater with this aggregation. This aggregation does not preserve either the hour or day of the month but does provide the period, which is valuable for scheduled test analysis.

create table **pr10v2ids_exclude** as (select s.submission_ID, s.submission_type, p.date , p.period, p.model, p.os_version, p.cma_id, p.network_operator_name, p.network_operator_code, p.sim_operator_name, p.sim_operator_code from **pr10v2_exclude** p inner join **pr9v2_exclude_ids_DISTINCT** s on p.date = date_trunc ('month', s.date) and p.period = retr_period(s.date) and p.submission_type = s.submission_type and p.model = s.model and p.os_version = s.os_version and p.sim_operator_code = upper(regexp_replace(s.sim_operator_code, '\s*', '', 'g'))   and p.network_operator_code = upper(regexp_replace(s.network_operator_cod e, '\s*', '', 'g'))   and p.network_operator_name = upper(regexp_replace(s.network_operator_nam e, '\s*', '', 'g'))   and p.sim_operator_name = upper(regexp_replace(s.sim_operator_name, '\s*', '', 'g'))   and p.cma_id = s.cma_id);
--
--OLD 2233546

mba=# select count(distinct submission_id) from pr10v2ids_exclude;;
  count

 2440472 |
| **pr10v2_exclude** | **pr9v2_exclude_ids_DISTINCT** | This table aggregates low-frequency scheduled tests samples by |

		Year/Month/Period, model, os version, CMA and a whitespace and case insensitive matching of SIM and network operator names and codes. 2.5 million test samples achieve a k-anonymity of two or greater with this aggregation. ``` create table pr10v2_exclude as (SELECT count(submission_id) as tally , submission_type , date_trunc('month', date) as date , retr_period(date) as period , model, os_version, cma_id , upper(regexp_replace(network_operator_name, '\s*', '', 'g')) as network_operator_name, upper(regexp_replace(sim_operator_name, '\s*', '', 'g')) as sim_operator_name , upper(regexp_replace(network_operator_code, '\s*', '', 'g')) as network_operator_code, upper(regexp_replace(sim_operator_code, '\s*', '', 'g')) as sim_operator_code from pr9v2_exclude_ids_DISTINCT where submission_type = 'scheduled_tests' group by upper(regexp_replace(network_operator_name, '\s*', '', 'g')) , upper(regexp_replace(sim_operator_name, '\s*', '', 'g')), upper(regexp_replace(network_operator_code, '\s*', '', 'g')), upper(regexp_replace(sim_operator_code, '\s*', '', 'g')) , cma_id , model, os_version , date_trunc('month', date), retr_period(date) , submission_type having count(*) > 1) ; -- (257009 rows) ```
pr10v2ids_exclude_man	**Pr10v2_exclude_man, pr9v2_exclude_ids_DISTINCT**	This table holds 32605 aggregated low-frequency manual test samples grouped by Year/Quarter/Hour, model, os version, CMA and a whitespace and case insensitive matching of SIM and network operator names and codes. 32605 test samples achieve a k-anonymity of two or greater with this aggregation. This aggregation does not preserve either the year or day of the month but does provide

		the hour, which is valuable for manual test analysis. ``` create table pr10v2ids_exclude_man as (select s.submission_ID, s.submission_type, p.year , p.quarter, p.hour, p.model, p.os_version, p.cma_id, p.network_operator_name, p.network_operator_code, p.sim_operator_name, p.sim_operator_code from pr10v2_exclude_man p inner join pr9v2_exclude_ids_DISTINCT s on p.year = extract (year from s.date) and p.quarter = retr_quarter (s.date) and p.hour = extract(hour from s.date) and p.submission_type = s.submission_type and p.model = s.model and p.os_version = s.os_version and p.sim_operator_code = upper(regexp_replace(s.sim_operator_code, '\s*', '', 'g')) and p.network_operator_code = upper(regexp_replace(s.network_operator_cod e, '\s*', '', 'g')) and p.network_operator_name = upper(regexp_replace(s.network_operator_nam e, '\s*', '', 'g')) and p.sim_operator_name = upper(regexp_replace(s.sim_operator_name, '\s*', '', 'g')) and p.cma_id = s.cma_id); --OLD SELECT 32605 ``` mba=# select count(distinct submission_id) from pr10v2ids_exclude_man ;; count ------- 52706 (1 row)
pr10v2_exclude_man	**pr9v2_exclude_ids_DISTI NCT**	This table aggregates low-frequency manual tests samples by Year/Quarter/Period, model, os version, CMA and a whitespace and case insensitive matching of SIM and network operator names and codes. 2.5 million test samples achieve a k-anonymity of two or greater with this aggregation.

		```create table pr10v2_exclude_man as (SELECT count(submission_id) as tally , submission_type , extract (year from date) as year, retr_quarter(date) as quarter, extract (hour from date) as hour, model, os_version, cma_id , upper(regexp_replace(network_operator_name, '\s*', '', 'g')) as network_operator_name, upper(regexp_replace(sim_operator_name, '\s*', '', 'g')) as sim_operator_name , upper(regexp_replace(network_operator_code, '\s*', '', 'g')) as network_operator_code, upper(regexp_replace(sim_operator_code, '\s*', '', 'g')) as sim_operator_code from pr9v2_exclude_ids_DISTINCT where submission_type ~* 'manual' group by upper(regexp_replace(network_operator_name, '\s*', '', 'g')) , upper(regexp_replace(sim_operator_name, '\s*', '', 'g')), upper(regexp_replace(network_operator_code, '\s*', '', 'g')), upper(regexp_replace(sim_operator_code, '\s*', '', 'g')) , cma_id , model, os_version , extract (year from date), retr_quarter(date) , extract(hour from date) , submission_type having count(*) > 1) ;```
		
		``` sum```
		```-------```
		```--OLD 32605```
		```(1 row)```
		
		```mba=# select count(*) from pr10v2_exclude_man ;;```
		``` count```
		```-------```
		``` 20182```

`pr11v2ids_EXCLUDE`	`pr11v2ids_EXCLUDE_sch`, `pr10v2ids_exclude_man`	This table holds the remaining samples after the first aggregation of manual and scheduled tests.
		
		```create table pr11v2ids_EXCLUDE as (select a.* from pr11v2ids_EXCLUDE_sch a left join pr10v2ids_exclude_man b on a.submission_id = b.submission_id where b.submission_id is NULL)```
		```mmba-# ;```
		```-- OLD SELECT 675585```

| | | ```
mba=# select count(distinct submission_id )
from pr11v2ids_EXCLUDE  ;;
  count
---------
 2044688
(1 row)
``` |
| **pr11v2ids_EXCLUDE_sch** | pr9v2_exclude_ids_DIST INCT, **pr10v2ids_exclude** | ```
create table pr11v2ids_EXCLUDE_sch as (
select a.* from pr9v2_exclude_ids_DISTINCT a
left join pr10v2ids_exclude b on
a.submission_id = b.submission_id where
b.submission_id is NULL )

 -- OLD count  706088
```<br><br>mba=# create table pr11v2ids_EXCLUDE_sch as ( select a.* from pr9v2_exclude_ids_DISTINCT a left join pr10v2ids_exclude b on a.submission_id = b.submission_id where b.submission_id is NULL )<br>mba-#<br>mba-# ;<br>SELECT 2101573<br><br><br>mba=# create table pr11v2ids_EXCLUDE_sch as ( select a.* from pr9v2_exclude_ids_DISTINCT a left join pr10v2ids_exclude b on a.submission_id = b.submission_id where b.submission_id is NULL )<br>;<br>SELECT 2101573<br><br>mba=# select count(distinct submission_id ) from pr11v2ids_EXCLUDE_sch ;;<br>  count<br>---------<br> 2097394 |

`pr11v2ids_exclude_init_YQDP_nomodel`	`pr11v2_exclude_init_YQDP_nomodel,` `pr11v2ids_EXCLUDE`	This table holds 164919 aggregated low-frequency init test samples grouped by Year/Quarter/Day/Period, os version, CMA and a whitespace and case insensitive matching of SIM and network operator names and codes. Test samples achieve a k-anonymity of two or greater with this aggregation. This aggregation does not preserve the month or device model or manufacturer but does provide the operating system version and day and hour, which are valuable temporal variables for init test analysis.

```
--- OLD SELECT 164919

mba=# create table
pr11v2ids_exclude_init_YQDP_nomodel as
(select s.submission_ID, s.submission_type,
p.year , p.quarter, p.day, p.period,
p.os_version, p.cma_id,
p.network_operator_name,
p.network_operator_code,
p.sim_operator_name, p.sim_operator_code
from pr11v2_exclude_init_YQDP_nomodel p
inner join pr11v2ids_EXCLUDE  s on p.year =
extract (year from  s.date) and p.quarter =
extract (quarter from s.date) and p.day =
extract (day from  s.date) and p.period =
retr_period(s.date) and p.submission_type =
s.submission_type and p.os_version =
s.os_version and p.sim_operator_code =
upper(regexp_replace(s.sim_operator_code,
'\s*', '', 'g'))   and
p.network_operator_code =
upper(regexp_replace(s.network_operator_cod
e, '\s*', '', 'g'))   and
p.network_operator_name =
upper(regexp_replace(s.network_operator_nam
e, '\s*', '', 'g'))    and
p.sim_operator_name =
upper(regexp_replace(s.sim_operator_name,
```

		```'\s*', '', 'g')) and p.cma_id = s.cma_id);``` `SELECT 854898`
`pr11v2_exclude_init_YQDP` `_nomodel`	`pr11v2ids_EXCLUDE`	This table aggregates low-frequency init tests samples by Year/Quarter/Day/Period, os version, CMA and a whitespace and case insensitive matching of SIM and network operator names and codes.. `-- OLD SELECT 63152` mba=# create table pr11v2_exclude_init_YQDP_nomodel as (SELECT count(submission_id) as tally , submission_type , extract (year from date) as year, extract (quarter from date) as quarter, extract (day from date) as day , retr_period (date) as period, os_version, cma_id , upper(regexp_replace(network_operator_name, '\s*', '', 'g')) as network_operator_name, upper(regexp_replace(sim_operator_name, '\s*', '', 'g')) as sim_operator_name , upper(regexp_replace(network_operator_code, '\s*', '', 'g')) as network_operator_code, upper(regexp_replace(sim_operator_code, '\s*', '', 'g')) as sim_operator_code from pr11v2ids_EXCLUDE where submission_type ~* 'init_test' group by upper(regexp_replace(network_operator_name, '\s*', '', 'g')) , upper(regexp_replace(sim_operator_name, '\s*', '', 'g')), upper(regexp_replace(network_operator_code, '\s*', '', 'g')),

		upper(regexp_replace(sim_operator_code , '\s*', '', 'g')) , cma_id , os_version , extract (year from date), extract (quarter from date) , extract (day from date), retr_period (date) , submission_type having count(*) > 1) ; --- SELECT 290376

`pr12v2ids_EXCLUDE`	`pr11v2ids_EXCLUDE_sch,`
	`pr10v2ids_exclude_man`
		The remaining 510440 samples after the first aggregations of init, scheduled and manual tests are further aggregated to increase the k value in various aggregations.
		
		`create table pr12v2ids_EXCLUDE as ( select a.* from pr11v2ids_EXCLUDE a left join pr11v2ids_exclude_init_YQDP_nomodel b on a.submission_id = b.submission_id where b.submission_id is NULL )`
		`-- OLD SELECT 510440`
		
		`mba=# create table pr12v2ids_EXCLUDE as ( select a.* from pr11v2ids_EXCLUDE a left join pr11v2ids_exclude_init_YQDP_nomodel b on a.submission_id = b.submission_id where b.submission_id is NULL );`
		`SELECT 1192837`

`pr13ids`	`Pr13_exclude,` `pr12v2ids_EXCLUDE`	488372 remaining samples are aggregated to include Year/Quarter, CMA location, and the network operator code and name. While the temporal and device related information is unavailable for analysis, the remaining samples location and network operator information by year and quarter are valuable for analysis of the active metric test results as well as network related signal strength, bearer channel and other features.

		``
create table pr13ids as (select
s.submission_ID, s.submission_type, p.year
, p.quarter,   p.cma_id,
p.network_operator_name,
p.network_operator_code   from pr13 p inner
join pr12v2ids_EXCLUDE s on p.year =
extract (year from s.date) and p.quarter =
extract (quarter from s.date) and
p.submission_type = s.submission_type and
p.network_operator_code =
upper(regexp_replace(s.network_operator_cod
e, '\s*', '', 'g'))   and
p.network_operator_name =
upper(regexp_replace(s.network_operator_nam
e, '\s*', '', 'g'))   and p.cma_id =
s.cma_id);
-- OLD SELECT 488372
```

SELECT 1153753 |
| **pr13** | **pr12v2ids_EXCLUDE** | ```
-- OLD SELECT 41655

mba=# create table pr13  as (SELECT
count(submission_id) as tally ,
submission_type , extract (year from date)
as year, extract ( quarter from date) as
quarter , cma_id ,
upper(regexp_replace(network_operator_name,
'\s*', '', 'g'))  as network_operator_name,
upper(regexp_replace(network_operator_code,
'\s*', '', 'g')) as network_operator_code
from pr12v2ids_EXCLUDE group by
upper(regexp_replace(network_operator_name,
'\s*', '', 'g')) ,
upper(regexp_replace(network_operator_code,
'\s*', '', 'g')),  cma_id , extract (year
from date), extract ( quarter from date) ,
submission_type having count(*) > 1) ;
SELECT 77176
``` |

Summary of Analysis of
Potential Low-Frequency Outliers in Active Test Results by Hour Block

Summarized below are results of two sets of analysis to identify whether any active metric results appear alone in any hour block in either the export cleared or export processing sets of submission_ids. No records among those flagged for export processing were discovered in isolation by hour. However, some small number of records were discovered to appear in the set of results cleared for export. Those records are removed from the general export and included in the set of results flagged for privacy processing.

Export Filters applied:
- where submission_id not in (select submission_id from pr_exclude_active_get2)
- where submission_id not in (select submission_id from pr_exclude_active_post2)
- where submission_id not in (select submission_id from pr_exclude_active_udp2)

Analysis:

```
create table pr_exclude_active_get as (select distinct submission_id from (
select e.submission_id from  pr_exclude e inner join  curr_httpget c  on
c.submission_id = e.submission_id  where date_trunc('hour', localtime) in
(select date from (select count( distinct e.submission_id) ,
 date_trunc('hour', localtime) as date from pr_exclude e inner join
 curr_httpget c  on c.submission_id = e.submission_id  group by
date_trunc('hour', localtime) having count( distinct e.submission_id) =1 ) as
foo) ) as foo);
--SELECT 176
mmba=# select count(distinct E.submission_id) from pr_exclude_active_get e
inner join pr_EXCLUDE s on e.submission_id = s.submission_id;
count
-------
  307
(1 row)


create table pr_exclude_active_get2 as (select distinct submission_id from (
select e.submission_id from  curr_submission_primarykey_clean e inner join
 curr_httpget c  on c.submission_id = e.submission_id  where date_trunc('hour',
localtime) in (select date from (select count( distinct e.submission_id) ,
 date_trunc('hour', localtime) as date from pr_exclude e inner join
 curr_httpget c  on c.submission_id = e.submission_id  group by
date_trunc('hour', localtime) having count( distinct e.submission_id) =1 ) as
foo) ) as foo);
SELECT 5490


select count(distinct E.submission_id) from pr_exclude_active_get e inner join
pr_EXCLUDE s on e.submission_id = s.submission_id;
307
mmba=# select count(distinct E.submission_id) from pr_exclude_active_get2 e
inner join pr_EXCLUDE s on e.submission_id = s.submission_id;
count
-------
```

```
       0
(1 row)



create table pr_exclude_active_post as (select distinct submission_id from (
select e.submission_id from  pr_exclude e inner join  curr_httppost c  on
c.submission_id = e.submission_id  where date_trunc('hour', localdtime) in
(select date from (select count( distinct e.submission_id) ,
 date_trunc('hour', localdtime) as date from pr_exclude e inner join
 curr_httppost c  on c.submission_id = e.submission_id  group by
date_trunc('hour', localdtime) having count( distinct e.submission_id) =1 ) as
foo) ) as foo);
--SELECT 635

mmba=# select count(distinct E.submission_id) from pr_exclude_active_POST e
inner join pr_EXCLUDE s on e.submission_id = s.submission_id;
count
-------
635
(1 row)



create table pr_exclude_active_post2 as (select distinct submission_id from (
select e.submission_id from  curr_submission_primarykey_clean e inner join
 curr_httppost c  on c.submission_id = e.submission_id  where
date_trunc('hour', localdtime) in (select date from (select count( distinct
e.submission_id) ,  date_trunc('hour', localdtime) as date from pr_exclude e
inner join  curr_httppost c  on c.submission_id = e.submission_id  group by
date_trunc('hour', localdtime) having count( distinct e.submission_id) =1 ) as
foo) ) as foo);
SELECT 10594

select count(distinct E.submission_id) from pr_exclude_active_POST2 e inner
join pr_EXCLUDE s on e.submission_id = s.submission_id;
count
-------
       0
(1 row)



create table pr_exclude_active_udp as (select distinct submission_id from (
select e.submission_id from  pr_exclude e inner join  curr_udplatency c  on
c.submission_id = e.submission_id  where date_trunc('hour', localdtime) in
(select date from (select count( distinct e.submission_id) ,
 date_trunc('hour', localdtime) as date from pr_exclude e inner join
 curr_udplatency  c  on c.submission_id = e.submission_id  group by
date_trunc('hour', localdtime) having count( distinct e.submission_id) =1 ) as
foo) ) as foo);
--SELECT 664
```

```
mmba=# select count(distinct E.submission_id) from pr_exclude_active_UDP e
inner join pr_EXCLUDE s on e.submission_id = s.submission_id;
count
-------
664
(1 row)

create table pr_exclude_active_udp2 as (select distinct submission_id from (
select e.submission_id from  curr_submission_primarykey_clean e inner join
 curr_udplatency c  on c.submission_id = e.submission_id  where
date_trunc('hour', localtime) in (select date from (select count( distinct
e.submission_id) ,  date_trunc('hour', localtime) as date from pr_exclude e
inner join  curr_udplatency  c  on c.submission_id = e.submission_id  group by
date_trunc('hour', localtime) having count( distinct e.submission_id) =1 ) as
foo) ) as foo);
SELECT 10901

select count(distinct E.submission_id) from pr_exclude_active_UDP2 e inner join
pr_EXCLUDE s on e.submission_id = s.submission_id;
count
-------
     0
(1 row)
```

APPENDIX B: REFERENCE DOCUMENT

Mobile Broadband Performance Application Privacy Notice and Terms of Use

(Version 2.0 Updated October 2016)

What Information We Collect
The information the FCC Speed Test App ("Application") collects is limited to information used to measure your mobile broadband service. Your location and time when a test is run, and IP address when data is uploaded to our servers is collected and used in processing your results. We do not collect other personally identifiable information, such as name, phone number, or identifiers associated with your device. We collect five kinds of information to measure the performance of your mobile broadband service.

- Location

- ○ When we measure your mobile broadband performance, we identify the location and unique ID of the cell tower that is providing that service, and the GPS location of your device when the measurements were made.
- ● Time of Data Collection
  - ○ We record the times when the Application performs measurements of your mobile broadband performance. This is necessary because some measurements function by recording a start and end time.
- ● Device Type and Operating System Version
  - ○ When the Application starts, we record the manufacturer and model of device you are using, and the version of the operating system installed on your device.
- ● Cellular Performance and Characteristics
  - ○ We collect information about your cellular device performance, such as your service provider, the strength of the radio signal, and what type of connection and service you have.
- ● Broadband Performance
  - ○ The Application tests your mobile broadband upload and download rate, round trip latency, and packet loss measured between your device and our servers managed by the FCC, the FCC contractor, and other project partners.

This data is stored on your device in the Application's sandbox (a protected private area for the Application) for you to review, and periodically is uploaded to our servers. You may delete older data periodically from your device. You can find more detail in the FCC's technical summary of this program. https://www.fcc.gov/measuring-broadband-america/mobile/technical-summary.html

Data Sharing

The data we collect is helpful for consumers, the FCC, and others interested in understanding mobile broadband in the United States. We share your mobile broadband performance information in four ways.

- ● **Internally**

We may share your mobile broadband performance information within the FCC to write our Measuring Broadband America Report on Mobile Performance and to support our development of broadband policy.

- ● **Third-Party Partners Who Help Us Run Our Test and Support Our Program**

We share your mobile broadband performance information with our contractor who manages the measurement program, with partners who support the program and architecture, and with partners supporting our analysis or contributing peer reviews. We share data subject to agreements that include strict requirements to safeguard the data, not disclose the data, and use data only in furtherance of the goals of the program.

The FCC has contracted with a broadband measurement company, SamKnows, to help manage the program and conduct its technical functions. SamKnows performs functions such as writing and maintaining the software that performs the test measurements, responding to customer service requests, managing the Application and the execution of tests, collecting and protecting data gathered in the trial, and providing analysis and support for the preparation of FCC reports. The terms of our contract with SamKnows include obligations to protect the data and your privacy.

Research platforms and your carrier may support various aspects of the testing and architecture for the program. We may share your detailed mobile performance data with these partners supporting our data processing and analysis for the specific purpose of identifying patterns in the data and recommending ways of addressing any concerns for your privacy and anonymity. The FCC directs researchers, its contractor, and your carrier working under non-disclosure agreements (NDA) to analyze the collected data for privacy risks or concerns and to plan for future measurement efforts.

Researchers and other partners supporting this analysis are skilled in the field of statistics and computer science work to identify patterns that a third party might exploit in ways to compromise your privacy, for example, by comparing the location information with other sources of information. If we find concerns, we will take steps to address the risks by making the time, location, or other information less specific; limiting the combinations of information we release; or deleting information. These partners are required to sign an NDA and must agree to data-handling practices that safeguard the data. The NDA prohibits any public disclosure of the data except in the form of a report to the FCC, limits both the access duration and specific research purposes of data, and requires disclosures if the data is ever compromised in any way.

- **Pursuant to legitimate requests by law enforcement or where otherwise required by law or regulation**

  - We may share mobile broadband performance data collected by this Application subject to legitimate requests by law enforcement or where otherwise required by law or regulation.

- **De-Identified Data Available to the Public**

We will publish de-identified data in summary statistics, coarsened data, and other minimally necessary datasets that support our review of the statistics only if we determine that the data poses very low risks to your privacy. These include:

  - **Summary Data**

We will publish averages and other summary statistics that describe important features of mobile broadband performance. These summary statistics pose very low risks to your privacy.

  - **Coarsened Data**

We will publish only de-identified data in a coarsened form that supports the averages and other statistics that we publish in our reports. Prior to release of this data, the FCC and researchers skilled in the field of statistics and computer science will identify patterns in the raw data that could be exploited to compromise your privacy. If we find concerns, we will take steps to address the risks by making the time, location or other information less specific; limiting the combinations of information we release; or deleting information.

For example, the analysis we perform would identify and address risks from a third party combining the location information in the data we collect with other sources of information to try to identify your device in our database. We will publicly release the results of this analysis to ensure the quality of our review,

and will also describe the measures we took to protect your anonymity. We will release information about the time or location of measurements only after reviewing the risks and making any changes to the precise data necessary to minimize the risk of individual volunteers being identified from the de-identified data.

Data Consumption

By default, the Application is configured to use no more than 100 megabytes (MB) of data each month for automated testing. Any additional tests you initiate will also consume between 100 kilobytes (kB) for the smallest packet loss test and 40 MB for a complete suite of tests. At any time, you can change the amount of data the Application uses by increasing or reducing the "monthly data cap" in the settings menu. You should ensure that your monthly broadband usage, including the data used by the Application and Service during the Program, does not exceed the data allowance included in your mobile broadband package. Your carrier may charge you for any data used in excess of your package's data allowance.

Stopping Your Participation
If at any time you want to stop participating in this program, simply delete the Application and the accompanying data, and the Application will not collect further data. Any data collected prior to the deletion of the Application will remain in the data sets subject to the above terms and protections.

- **Other terms of use of the Application**
  - Detailed Testing Architecture and Technical Summary (https://www.fcc.gov/measuring-broadband-america/mobile/technical-summary.html)
  - Program Stakeholders' Code of Conduct (https://www.fcc.gov/measuring-broadband-america/mobile/code-of-conduct.html)
  - GNU Licensing Agreement (http://www.gnu.org/licenses/gpl-2.0.html)
  - This PRA Collection Authorized By OMB Control No. 3060-1139 (Expires 05/31/2017). Paperwork Reduction Act (PRA) of 1995 Notice (https://www.fcc.gov/measuring-broadband-america/mobile/pra-privacy-act-notice.html)
  - Mobile Provider Disclaimers (https://www.fcc.gov/measuring-broadband-america/mobile/mobile-provider-disclaimers.html)
  - Application Terms and Conditions and Privacy Notice (https://www.fcc.gov/measuring-broadband-america/mobile/mobile-terms-privacy-notice.html)
  - Mobile broadband providers endorse some applications certified for use on their network. This application may not have be formally certified by your provider or carry a formal endorsement from your mobile broadband network