GreedyGame : Session Calculation

Problem Statement:

1. Calculate the number of sessions (valid and total) of a game and the average session time(only valid) from the given dataset and explain your calculation strategy.
2. Provide insights about users (if any)
3. Point out discrepancies in the data and suggestive work around (if any)

Description:

At GreedyGame, one of the user behaviour metrics is the Gaming Session. These are calculated based on the events that are sent out by our Mobile SDK. A gaming session happens when a user plays a game on a device (identified by 'ai5' field). The following are some attributes of a session.

1. A session typically starts with a GGSTART event
2. A session typically ends with a GGSTOP event.
3. If there is a time difference of more than 30 secs between a GGSTOP and GGSTART, they are considered to be different sessions.
4. There can be multiple GGSTART and GGSTOP calls in a single session.
5. If a session is more than 60 seconds long, it is classified as a valid session.
6. If a session is less than 1 second, it should be ignored
7. Incase of multiple GGSTART and GGSTOP calls, the exact time of the session should be taken for establishing session validity (not the difference between the first GGSTART and the last GGSTOP call)

To understand this functionality better, consider the following sequence of calls for a particular ai5.

Event	Timestamp
ggstart	2016-05-09 2:00:00
ggstop	2016-05-09 2:09:13
ggstart	2016-05-09 2:09:14
ggstop	2016-05-09 2:17:57
ggstart	2016-05-09 3:00:00
ggstop	2016-05-09 3:02:46

If you look at the data above, this represents two sessions (first one from 2:00:00 to 2:17:57, with a total time of 17:56 and the second one from 3:00:00 to 3:02:46 with a total time of 2:46)

Because of the nature of mobile data, there are bound to be data losses in the entries. The algorithm should take this into account when figuring out sessions and work with a minimum loss strategy.
A brief description of some of the fields of the dump is given below:

Field	Description
ai5	One-way hash of a particular device identifier
sdkv	Mobile SDK version
event	Type of event on the mobile device (GGSTART or GGSTOP)
game_id	Unique Identifier for a particular game
timestamp	Time when the request arrived on server (GMT)
ts	Epoch timestamp value of the device

You are free to choose a language to tackle the problem.
There is no direct time limit for the problem but it is highly recommended that you revert as soon as possible with the solution.
Do think about scaling of your approach to a traffic 100x of the data set provided.

Submissions:
Upload your code to a code versioning system (like GitHub)
Send out a detailed mail or attached document explaining your approach