CDR-INSIGHT PROJECT

5/11/2015	1.1	Muhammad Zahalqa
3/3/2016	1.2	Muhammad Zahalqa
4/5/2016	1.3	Muhammad Zahalqa
20/7/2017	1.4	Muhammad Zahalqa
31/7/2017	1.4.1	Alex Katz - typo cleaning
02/06/2019	2.0	Muhammad Zahalqa
25/06/2019	2.1	Muhammad Zahalqa

1 Scope

The present document is part of a series of documents that specify charging functionality and charging management in GSM/UMTS and EPS networks.

Specifically, this document describes the handling of CDR files by a processor for charging and intelligence purposes. The processing is based large data files containing CDR records describing services consumed by subscribers as recorded by the mobile network infrastructure.

Each CDR record contains descriptions of a usage event such as voice calls and SMS messages with all relevant data.

The processor will aggregate the information presented in these files in order to support producing billing for subscribers and answer intelligence queries about the subscribers..

2 Definitions, symbols and abbreviations

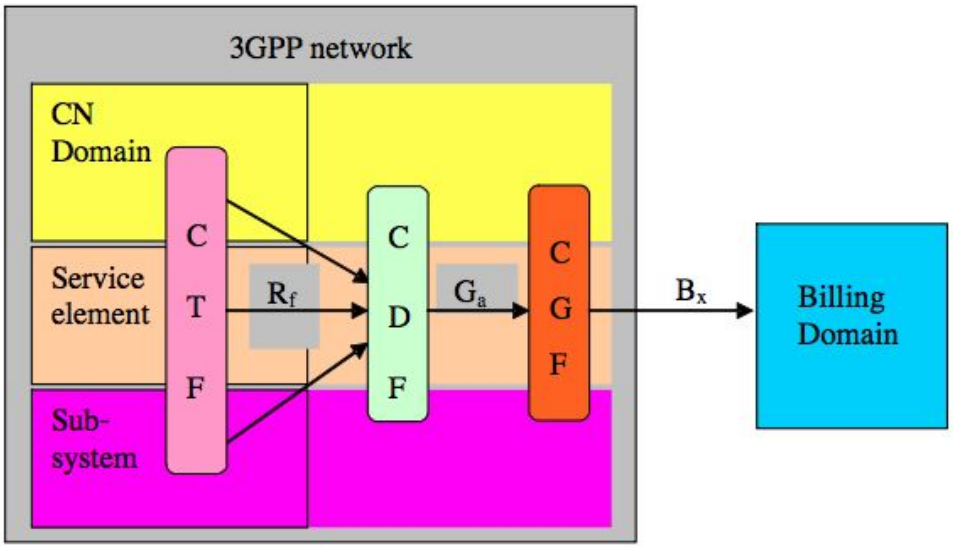
2.1 Definitions

- **2G- / 3G-**: prefixes 2G- and 3G- refer to functionality that supports only GSM or UMTS, respectively, e.g. 2G-SGSN refers only to the GSM functionality of an SGSN.
- **Accounting**: process of apportioning charges between the Home Environment, Serving Network and Subscriber. accounting meter record: record containing one or more counters employed to register the usage of resources en
- **Billing Domain**: part of the operator network, which is outside the core network, which receives and processes CDR files from the core network charging functions. It includes functions that can provide billing mediation and billing or other (e.g. statistical) end applications. It is only applicable to offline charging (see "Online Charging System" for equivalent functionality in online charging).
- **Charging Data Record (CDR)**: formatted collection of information about a chargeable event (e.g. time of call set-up, duration of the call, amount of data transferred, etc) for use in billing and accounting. For each party to be charged for parts of or all charges of a chargeable event a separate CDR shall be generated.
- **GPRS:** packet switched bearer and radio services for GSM and UMTS systems.
- **Subscriber:** entity (associated with one or more users) that is engaged in a Subscription with a service provider. The subscriber is allowed to subscribe and unsubscribe services, to register a user or a list of users authorised to enjoy these services, and also to set the limits relative to the use that associated users make of these services.
- **Successful call:** connection that reaches the communication or data transfer phase e.g. the "answered" state for speech connections. All other connection attempts are regarded as unsuccessful.

2.2 Abbreviations

BD	Billing Domain
CDR	Charging Data Record CG Charging Gateway
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
MMS	Multimedia Messaging Service
MO	Mobile Originated
MOC	MO Call
MSISDN	Mobile Subscriber ISDN Number
MT	Mobile Terminated
MTC	MT Call
SIM	Subscriber Identity Module
SMS	Short Message Service

3 Common charging architecture



CTF: Charging Trigger Function
CDF: Charging Data Function
CGF: Charging Gateway Function
BD: Billing Domain. This may also be a billing system/ billing mediation device.

4 Charging Mechanisms

4.1 Mechanisms

GSM/UMTS/EPC networks provide functions that implement offline and/or online charging mechanisms on the bearer (e.g. EPC), subsystem (e.g. IMS) and service (e.g. MMS) levels. In order to support these charging mechanisms, the network performs real-time monitoring of resource usage on the above three levels in order to detect the relevant chargeable events.

Typical examples of network resource usage are a voice call of certain duration, the transport of a certain volume of data, or the submission of a MM of a certain size. The network resource usage requests may be initiated by the subscriber (MO case) or by the network (MT case).

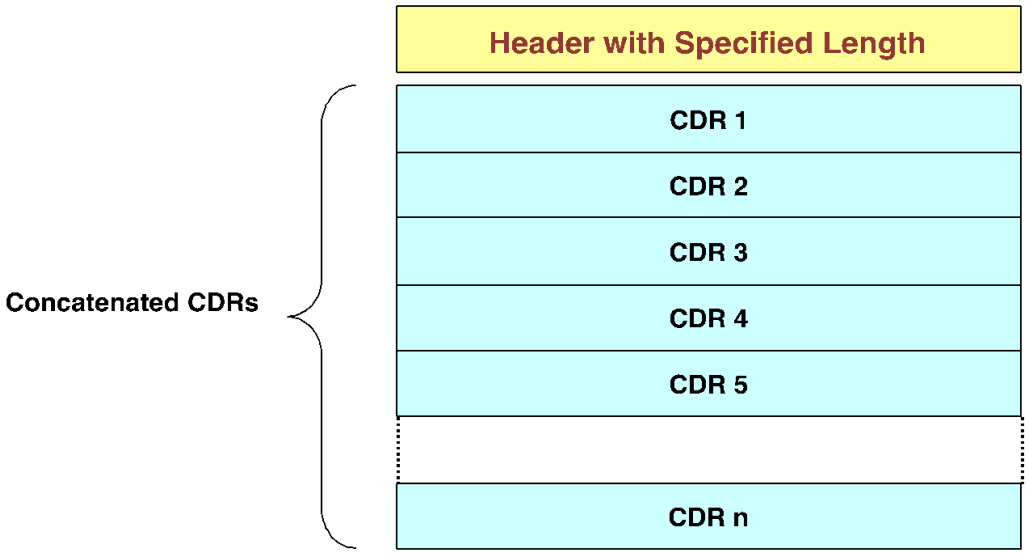
4.2 Offline Charging

Offline charging is a process where charging information for network resource usage is collected concurrently with that resource usage. The charging information is then passed through a chain of logical charging functions.
At the end of this process, CDR files are generated by the network, which are then transferred to the network operator's BD for the purpose of subscriber billing and/or inter-operator accounting (or additional functions, e.g. statistics, at the operator's discretion). The BD typically comprises post-processing systems such as the operator's billing system or billing mediation device.

4.3 CDR File Format

4.3.1 File format principles

The CDR files contain a header section followed by a variable sized CDR data section. The CDR data section contains zero or more concatenated CDR records. Each CDR record is encoded on a single line.



4.3.2 CDR Format

CDR record is composed of the following fields separated by |

- Sequence Number
  Running number unique to the file

- Caller IMSI
  Unique number that identifies every user of a cellular network. The ITU-T recommendation ITU E.212 limits the maximum length of an IMSI to 15 digits. (can be stored as a 64-bit field.)
  The IMSI is composed of:
  1. First 3 digits represent the country: MCC
     https://en.wikipedia.org/wiki/Mobile_country_code

  2. Next 3 or 2 digits represents the mobile network operator: MNC
     USA standard uses 3 digits, EU standard uses 2 digits.
     We will assume it's always 3 digits.
     See the following links for North America Operators:
     Canada and USA

  3. The remaining 10 digits are the mobile subscriber identification number (MSIN) within the network's customer base

  Examples:
  - **IMSI**:313460000000001
    MCC: 313 which is the USA
    MNC: 460 which is Mobi
    MSIN: 000000001
  - **IMSI**:3023204242424242
    MCC: 302 which is Canada

MNC: 460 which is Roger Wireless
MSIN: 4242424242

- Caller IMEI
  The Subscriber Equipment Number field contains the identification of the mobile device.
  Example: 35-209900-176148-1

- Usage Type
  One of
  - MOC outgoing voice call
  - MTC incoming voice call
  - SMS-MO outgoing message
  - SMS-MT incoming message
  - D Data
  - U call not answered
  - B busy call
  - X failed call
- Caller Mobile Number
  15 digits max

- Call Date DD/MM/YYYY

- Call Time HH:MM:SS

- Duration in seconds

- Bytes Received if type is Data

- Bytes Transmitted if type is Data

- Callee MSISDN
  Dialled party MSISDN or the caller MSISDN if this is an inbound call. This will be empty for Data calls.

- Calle Mobile Number
  15 digits max

5 Processing Functions

The processor of the CDR files will be fed CDR files in a specific configurable directory. New files will be delivered every 10 minutes.
These are usually large files (1-5 GB) and are copied to the directory by a remote process as they are continuously collected from the network switches.

Processor will read and process files in order to fulfill the following functions:

5.1 Billing

For each customer, identified by IMSI an aggregate information will be produced that will contain the total of

- Outgoing voice calls duration to a subscriber within the mobile operator
- Incoming voice calls duration to a subscriber within the mobile operator
- Outgoing voice calls duration to a subscriber outside the mobile operator
- Incoming voice calls duration to a subscriber outside the mobile operator
- SMS messages sent within the mobile operator
- SMS messages received within the mobile operator
- SMS messages sent outside the mobile operator
- SMS messages received outside the mobile operator
- Total KB data transfered
- Total KB data received

5.2 Inter Operator Settlement

For each mobile operator we will aggregate all:

- Total incoming voice call duration (seconds)
- Total outgoing voice call duration (seconds)
- Total Incoming SMS messages
- Total Outgoing SMS Messages

5.3 Links Graph Database

For each customer, identified by IMSI build a graph database of all parties that a customer came in contact with. The weight of an edge in the graph connecting two customers is composed of the tuple from total duration of all calls between them and the total number of SMS exchanged.

6 Query Gateway

The system will contain a query gateway that enables interrogation of the system using a URL based scheme.

6.1 Required Functionalities

The system will support the following commands, no authentication is required, which will return the response in a JSON format (to be defined by the implementation).

Sample query format:
http://server:8080/query/path?from=0123456789&to=383838383

6.1.1 MSISDN Retrieve Query

Support retrieving aggregate usage information for one subscriber at a time. The subscriber will be identified by his MSISDN.

6.1.2 Operator Retrieve Query

Support retrieving aggregate usage information for one operator at a time. The operator will be identified by his MCC/MNC.

6.1.3 Links Query

Produce a list of all subscribers who were in direct contact with a given subscriber identified by MSISDN.

6.1.4 Link Path Query

Produce a list of all subscribers along a path between two given subscribers identified by their MSISDN.

7 Storing Data into a Database

The system will store all of it's aggregate and graph data into a mysql database periodically. The period should be configurable.

A snapshot of the aggregate and graph data will also be stored before shutdown of the server.

8 Additional Requirements

8.1 Configuration

The system should be driven by configuration files.

For example: setting directories to be read from, number of threads, etc.
Configuration files should have ini file format:

```
# incoming data listening port
data-port=4040
# web api port
query-web=8080
# query processing concurency level
query-concurrency=4

# database connection
```

8.2 Handle CDR input files

Handling of CDR input files will be done by separate reader processes that will:

- Discover and process files as they appear in the input directory.
- Processed files should be moved to a "done" directory.
- The system should be able to handle more than one file simultaneously. Typically there will be 2-4 files delivered at once with a total of ~100 files per hour.
- Support future change of file format.
- Support concurrent processing of more than one file format
- Support receiving CDR data via means other than files.

8.3 Performance

The system should allow for high performance by utilizing the most out of the current hardware.
Assume we can put the incoming CDR data files on a file storage with high IO capacity.

9 Deliverables

9.1 High Level Design Document

A document in PDF format with high-level block diagram describing the system.

9.2 Detailed Level Design Document

A document in PDF format with

- Detailed level class diagram describing the system modules.
    - Each module should describe the detailed implementation requirements and infrastructure dependencies.
    - Each module should have it's threading requirements described.
- Sequence Diagram describing major interactions in the system.

9.3 Working Project

- A fully commented source code
- Makefile
- Unit tests for applicable infrastructure modules
- Readme file describing how to build and how to operate the system