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
04/07/2019	2.2	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 events by a processor for charging and intelligence purposes.

CDR events should be fed to the processor over the network either in real time or by parsing CDR data files - usually at scale of 1-2 GB - produced by the cellular switches. CDR data describe the services consumed by subscribers as recorded by the mobile network infrastructure such as voice calls and SMS messages.

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.

Users will be able to query the data using a UI client that will communicate via network with the processor.

2 Definitions, symbols and abbreviations

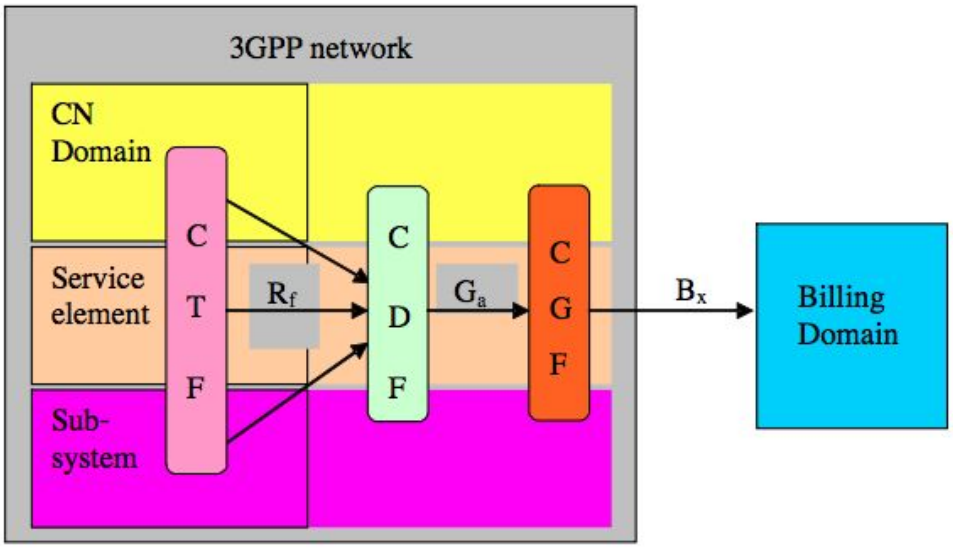
2.1 Definitions

- **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, the 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).

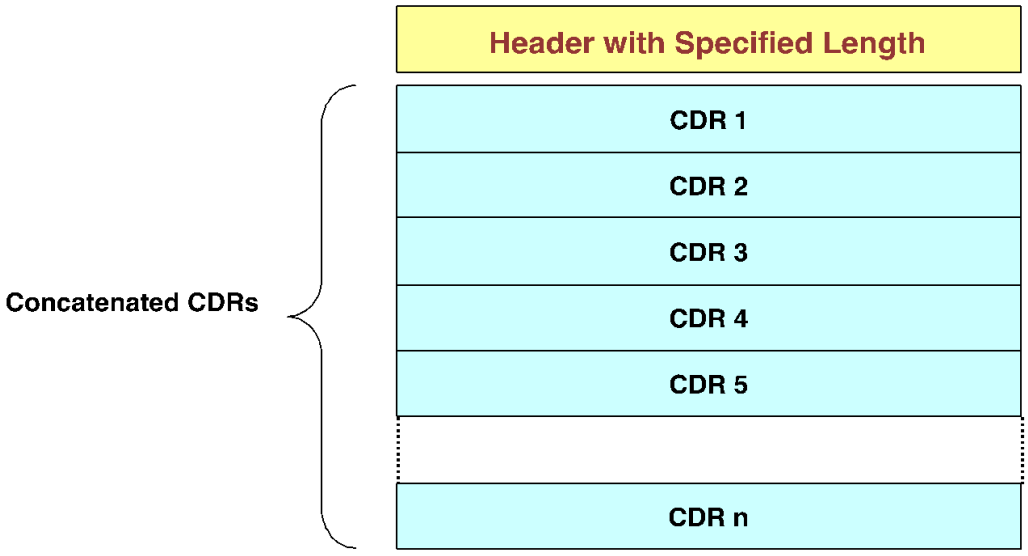
4.2 Offline Charging

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 "|" character

- Sequence Number
  Running number unique to the file.

- Subscriber 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 be using EU 2 digits standard.
     See the following link for Israeli operators:
     https://en.wikipedia.org/wiki/Mobile_Network_Codes_in_ITU_region_4xx_(Asia)#Israel_-_IL

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

  Examples:

  1. **IMSI**: 425020528409010
     MCC: 425 4: Asia, 25: Israel
     MNC: 02 Cellcom
     MSIN: 0528409010

2. **IMSI**: 262040162782277
   MCC: 262 which is Germany
   MNC: 04 which is Vodafone
   MSIN: 0162782277

- Subscriber 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
- Subscriber MSISDN
  Mobile number of subscriber composed of country code (1 -3 digits prefix) and a phone number. The MSISDN contains a maximum of 15 digits.

- 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

- Second Party IMSI
  Dialled party IMSI for outbound (MO) or the caller IMSI if this is an inbound call. This will be empty for Data calls.

- Second Party MSISDN
  Second party msisdn identified the destination for outbound or the source for inbound. This will be empty for Data calls.

5 Processing Functions

The processor of the CDR files will be fed CDR data by an external system delivering data files to a specific configurable directory. New files will be delivered approximately 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 in total
- Incoming voice calls duration in total
- Total Data transferred in KB
- Total Data received in KB
- Total SMS sent
- Total SMS received
- Per second party identified by MSISDN: total voice calls duration and SMS exchanged.

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 of 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 REST based API using HTTP get requests.

See REST tutorial at https://restfulapi.net/

6.1 Required Functionalities

The system will support the following commands, no authentication is required. All requests will use HTTP Get method and responses will be JSON formatted with requested data or with an error indication.
Specifying the format of the requests and responses is part of the implementation design documents.

See 6.1.1 for a suggested format.

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 or mobile number.

request: http://server:8080/query/msisdn/MSISDN
response: aggregate usage for given MSISDN or an error object.

Sample:
request: http://server:8080/query/msisdn/9720528409042
response: {

```
"msisdn": "9720528409042"
"voice-out": 61523,
"voice-in": 1343,
"data-out": 9273,
"data-in": 8215,
"sms-out": 54,
"sms-in": 21
}
```

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.

request: http://server:8080/query/operator/MCCMNC
response: aggregate usage for given operator or an error object.

Sample:
request: http://server:8080/query/operator/4025
response: {

```
        "voice-out": 61523827346,
        "voice-in": 34551343,
        "sms-out": 8253,
        "sms-in": 7129
    }
```

6.1.3 Links Query

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

request: http://server:8080/query/link/first-party/second-party
sample: http://server:8080/query/link/9720528409042/496221540

Response is JSON object with fields describing the attributes of the link between the two parties: voice calls and sms volume exchanged.

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 on shutdown. The system will load this data on startup.

This feature should be implemented in phase 2.

8 Additional Requirements

8.1 Deployment

The component of the system should be deployable to separate tiers.

a. Client UI - potentially ~100 end points
    i. Console based client
    ii. Optionally a web UI
b. Processor Server - One machine
c. CDR harvesters/Readers - potentially ~10 machines.

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.

A simulator process should be provided for testing purposes.

8.3 Configuration

The system should be driven by configuration file/files.
You should describe the format of the required configuration files as part of the documentation of the project.

sample configuration for server:

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

8.4 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