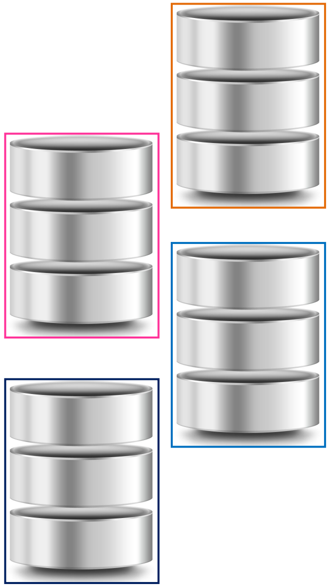
Big Data for MNO

Sept 2017

www.cubro.com

Big Data / Data Enrichment



The illustration shows the silo architecture of the past in which each system produces XDRs, counter and logs and other type of information which is stored in many different places, formats and DBs.

The information can only be correlated by specially designed software and often the raw data is not accessible to the customer.

Each request for a new report generates efforts and cost. As a result, a lot of useful information is not used and often there is a data overlap. This generates additional cost.

Big Data is not just a buzzword. This approach brings real advantage to customers. Many data sources feed one big data storage system. This system is a document storage system, which can handle a very large amount of unstructured data.

- One data storage for all types of data
- One reporting tool
- One point of securing data
- One managed Infrastructure

http://bigdata-madesimple.com/11-interesting-big-data-case-studies-in-telecom/

Big Data - A Treasure for MNO

CUBRO

- The only asset from Facebook and Google is their data!
- As a MNO you are sitting on a big treasure, your data is processed by your network every second.
- You can harvest this fortune, to use the data internally or resell the data to third party.
- Cubro can help you to make money out of your Data.
- Cubro can help to switch your monitoring systems from an OPEX grave to a cash cow.

Big Data is a Treasure for MNO

Usage of Geolocation Data :

Marketing: This data can be used to analyze customer behavior in general; the usage of this data is endless. Here are few examples:

How long customers stay at a location
How often is a customer joining a specific location.
Which transport medium is the customer using to come to a location.

With this information it is possible to optimize the availability of resources.
(Public Transportation, more staff at the parking place, optimise the pricing, etc)

Security: With this data security threats can be detected at a very early stage, because it is possible to see how many people are moving towards a place, and if this is an unusual behavior then Law and Enforcement can react faster.

Find persons when they dial an emergency number, or are in a catastrophic scenario.

Events: Analyzing customer movements at big events to avoid panic and rush situation.

Traffic: With this data traffic steering can be automated by changing traffic light configuration.

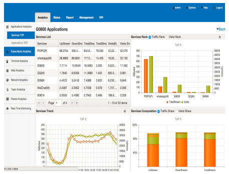
Big Data Integration Mobile Networks

CUBRO

Kibana BI

Revenue assurance, Fraud detection, Network planning, SLA…..

Real-time troubleshooting



Elastic Search

Hadoop Cluster (Long time storage)

Mongo DB (Short term storage)

XDR and KPI fetch (not real-time)

XDR as UDP stream real-time

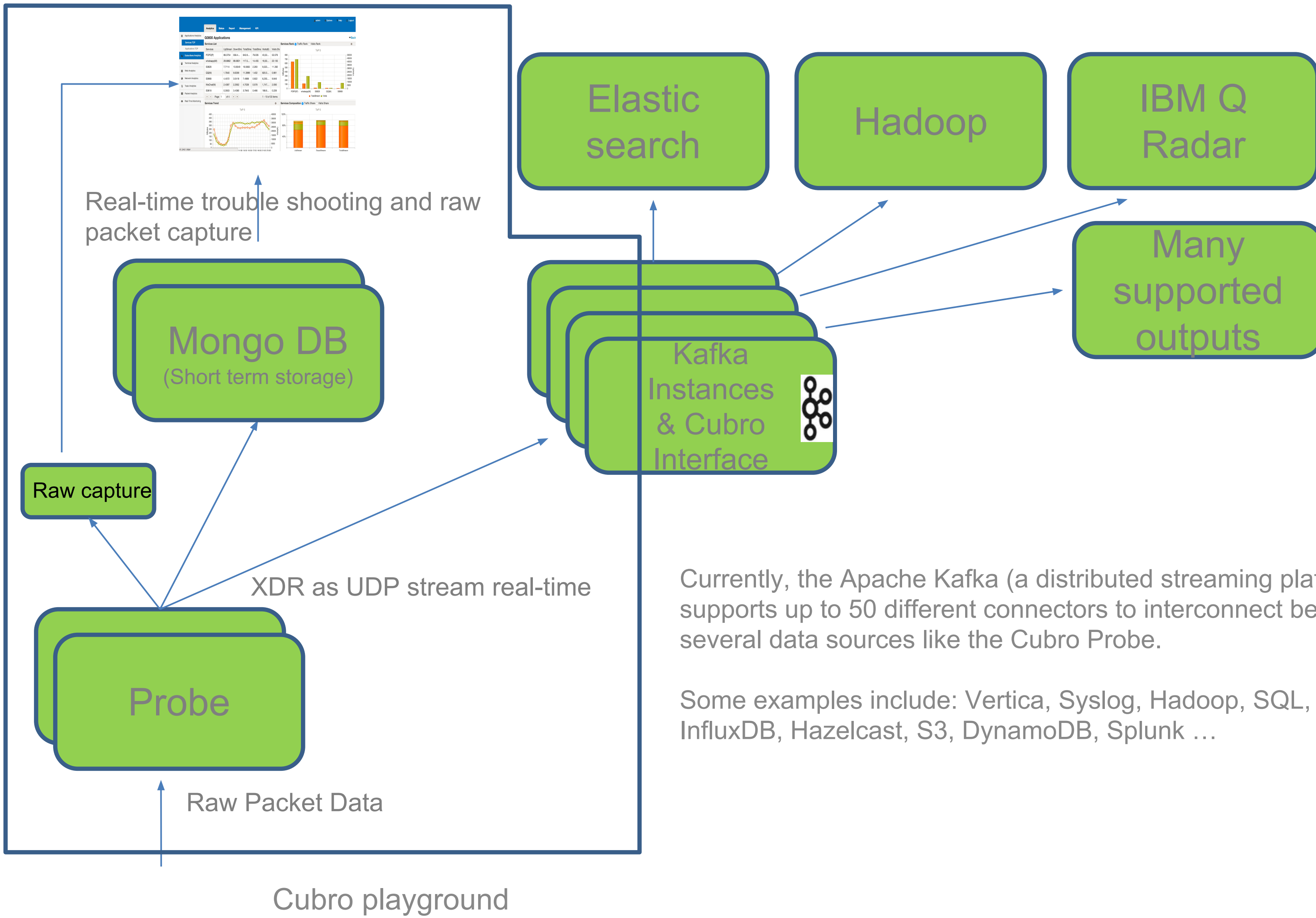
Probe

Raw Packet Data

Mid size provider data calculation is 4,5 million subscribers and 100 Gbps data load

6 – 7 billion CDRs a day, 5 TB storage needed for the CDR per day

Big Data Integration Mobile Networks



Currently, the Apache Kafka (a distributed streaming platform) supports up to 50 different connectors to interconnect between several data sources like the Cubro Probe.

Some examples include: Vertica, Syslog, Hadoop, SQL, Hbase, InfluxDB, Hazelcast, S3, DynamoDB, Splunk …

Current XDRs (extended data record)

Gn signaling record,
GTPv2 signaling record,
S3 signaling record,
DNS signaling record,
User service flow record,
MMS MO signaling record,
MMS MT signaling record,
WAP_CONNECT record,
WAP signaling record,
ONLINE VIDEO record,
FTP record,
RTSP record,
E MAIL record,
VOIP record,
P2P record,
IM record,
S1 signaling record,
S1 EMM signaling record,
S1 ESM signaling record,
S1AP protocol switching record,
S1AP protocol RAB record,
S1AP protocol management record,
S6a record,
S1 SMS record,
S1 CS fallback record,
SGS MM signaling record,
SGS CS signaling record,
X2 interface management record,
X2 interface switching record,
UU signaling record,
UU switching record,
UU-community measurement,
UU-UE measurement,
attach signaling,
detach signaling,
PDP Activation,
PDP Deactivation,
PDP Modification,
RAU,BSSGP
RANAP,
Relocation,
Service Request,
2G Paging

Each XDR is a UDP packet streamed to the collector

User Service Flow Record Fields

Time of Data Transmission
User IP Address
State Code
Network Code
Cell ID
Tracking Code
Location Code
Routing Area ID
2G/3G Network ID
User Location Information
IMSI
Subscriber number
IMEI
APN
Charging ID
SGSN User Plane Transmission IP
GGSN User Plane Transmission IP
Destination IP
SGSN User Plane TEID
GGSN User Plane TEID

User Port
Destination Port
TCP FIN times
Uplink Dropped Packets Number
Downlink Dropped Packets Number
Total Number Of Uplink Data Packets
Total Number Of Downlink Data Packets
Uplink Traffic
Downlink Traffic
Window Size
MSS Size
RST Direction
Bearing Layer Protocol
Fragments flag
SYN Number In TCP Linking
Successful Identification Of Three Shake Hands
SYN ACK Number In TCP Linking

ACK Number InTCP Linking
Uplink IP Fragment Number
Downlink IP Fragment Number
Disordered packet number of Uplink TCP
Disordered packet number of Downlink TCP
Retransmission packet number of Uplink TCP
Retransmission packet number of Downlink TCP
TCP RESET Number
Direction
Protocol Type
Response delay Of TCP Linking
Confirmation delay Of TCP Linking
Delay Between TCP Linking And The First Request
Delay Between he First Request And The First ACK

	Field Name	Name	Type	Length	Default Value	Meaning/Notes
2	Field Name	Name	Type	Length	Default Value	Meaning/Notes
3	CDR_HEADER	CDR Header Structure		52		All CDR Header Structure
4	trans_time	Time of Data Transmission	uint64_delay	8	0	Unit: cycle
5	user_ip	User IP Address	uint8	16		User IP Address（ipv4 or ipv6）
6	mcc	State Code	uint16	2	0xFFFF	State code
7	mnc	Network Code	uint16	2	0xFFFF	Network code
8	eci/cell_id	Cell ID	uint32	4	0xFFFFFFFF	Cell ID / (fill in eci when LTE， fill in ci when 2G， fill in sac when 3G)
9	tac	Tracking Code	uint16	2	0xFFFF	Tracking code / using in LTE
10	lac	Location Code	uint16	2	0xFFFF	Location code
11	rac	Routing Area ID	uint8	1	0xFF	Routing area ID
12	rat	2G/3G Network ID	uint8	1	0xFF	2G/3G network ID
13	uli	User Location Information	uint8	1	0xFF	0：CGI，1：SAI，2：RAI
14	reserved	Reserved	uint8	1		
15	imsi	IMSI	string	18	NULL	Subscriber IMSI
16	msisdn	Subscriber number	string	24	NULL	Subscriber number
17	imei	IMEI	string	18	NULL	IMEI
18	apn	APN	string	32	NULL	APN
19	charge_id	Charging ID	uint32	4	0	Charging information
20	sgsn_u_ip	SGSN User Plane Transmission IP	uint8	16	0	SGSN user plane transmission IP
21	ggsn_u_ip	GGSN User Plane Transmission IP	uint8	16	0	GGSN user plane transmission IP
22	dest_ip	Destination IP	uint8	16	0	Destination IP of application layer
23	sgsn_u_teid	SGSN User Plane TEID	uint32	4	0	SGSN User Plane TEID
24	ggsn_u_teid	GGSN User Plane TEID	uint32	4	0	GGSN User Plane TEID
25	user_port	User Port	uint16	2	0	User port number
26	dest_port	Destination Port	uint16	2	0	Destination port number
27	reserved1	Reserved1	uint8	1		
28	tcp_fin_pkts	TCP FIN times	uint8	1	0	
29	up_tcp_dropped_pkts	Uplink Dropped Packets Number	uint8	1	0	
30	down_tcp_dropped_pkts	Downlink Dropped Packets Number	uint8	1	0	
31	up_packets	Total Number Of Uplink Data Packets	uint32	4	0	Total number of uplink data packets
32	down_packets	Total Number Of Downlink Data Packets	uint32	4	0	Total number of downlink data packets
33	up_bytes	Uplink Traffic	uint32	4	0	Total uplink traffic, unit: Byte
34	down_bytes	Downlink Traffic	uint32	4	0	Total downlink traffic, unit: Byte
35	tcp_window_size	Window Size	uint16	2	0	Window size, the window after TCP linking negotiation
36	tcp_mss_size	MSS Size	uint16	2	0	MSS size
37	tcp_reset_direction	RST Direction	uint32	4	0	Launch direction of TCP connection resetting
38	proto_type	Bearing Layer Protocol	uint8	1	0xFF	Two types of bearing layer protocol 6 : TCP 17 : UDP
39	segment_flag	Fragments flag	uint8	1	4	1. Pure start fragment 2. Pure intermediate fragment (period) 3. Pure end fragment(RST/FIN) 4. Pure end fragment (Timeout) 5. Start&end fragment(RST/FIN) 6. Start&end fragment(Timeout)

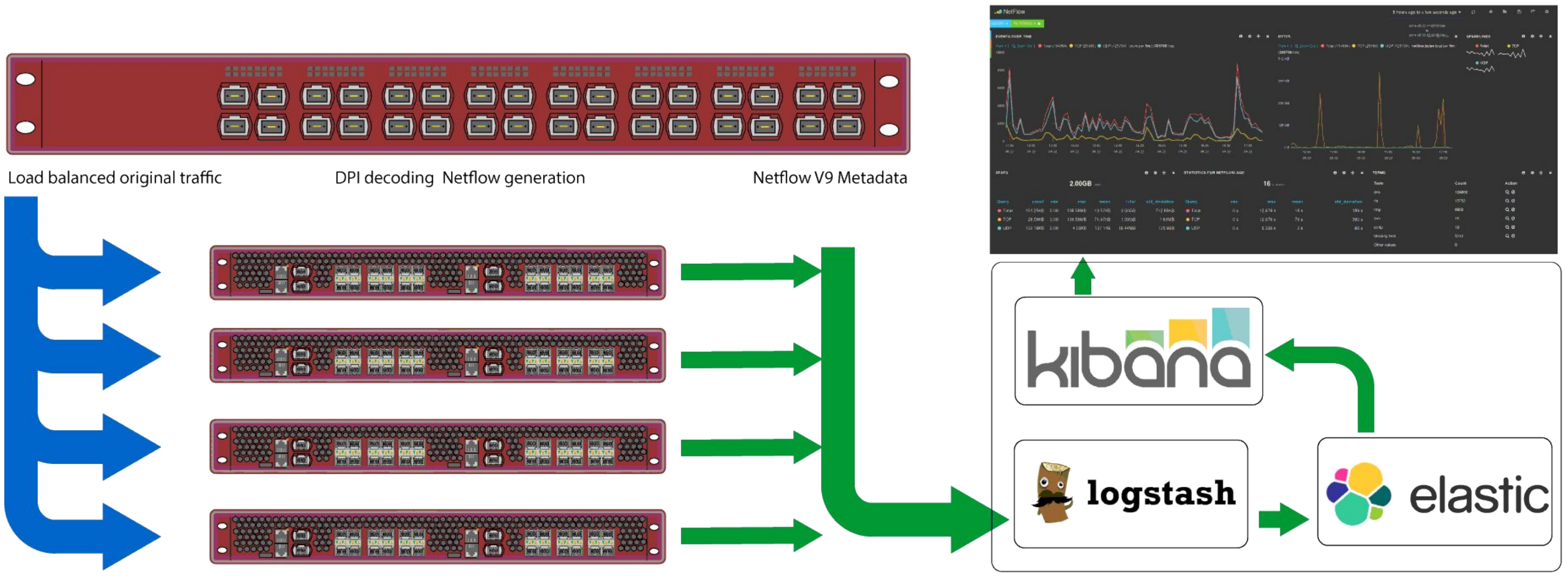
These are the fields which are available in the user service flow XDR

Big Data Integration Flow Monitoring for Data Centers

Example for a multiple 100 Gbit flow monitoring solution

Each Probe can support 60 Gbit traffic with DPI analysis enabled

and 120 Gbit traffic without DPI analysis

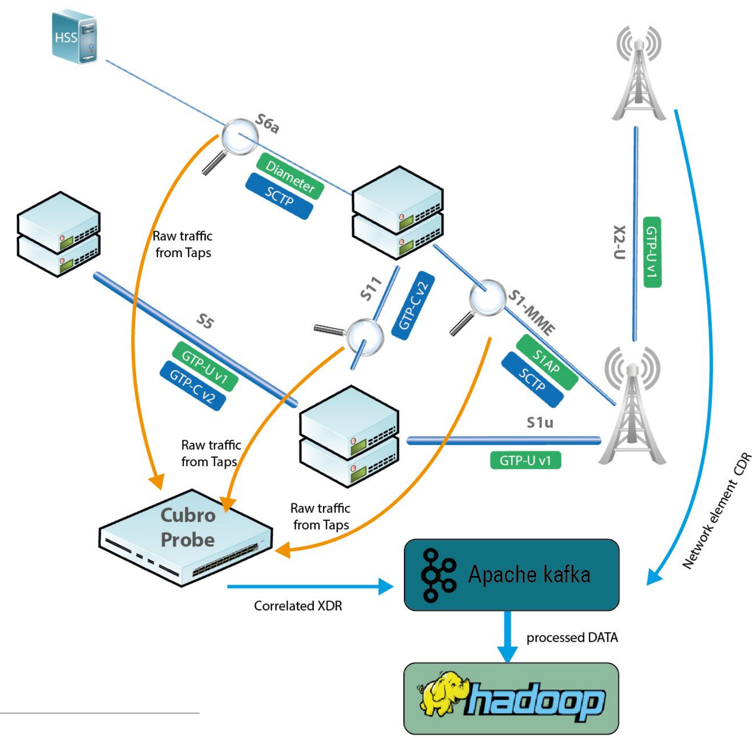


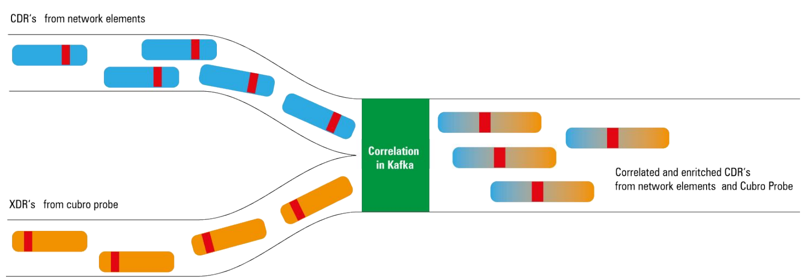
Typical Application Data Enrichment

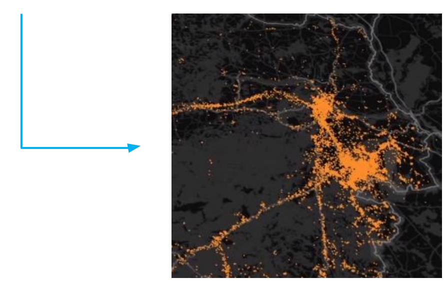
for Geolocation application in Mobile Networks

Enriching existing CDRs coming from network elements

To get a full data set for this Geolocation application.







Geolocation Application in Mobile Networks

The request was to generate subscriber based, high quality real time geolocation data to generate revenue out of this data for the MNO.

Challenge : Real time, all connected subscribers, better than cell id

- real time means a lot processing power is needed
- all subscribers means a lot information must be handled per second
- better location cell id means to get data from UTRAN interfaces or CDRs from eNodeB
  - real time correlation of all core interfaces with a IMSI refill rate better than 95 % (Cubro reached 97 - 98 %)
  - near real time correlation of core and UTRAN CDRs to combine the subscriber
  - information: terminal model, IMSI with the cell id, the antenna vector
  - receiving HF power from eNodeB to calculate an exact location.

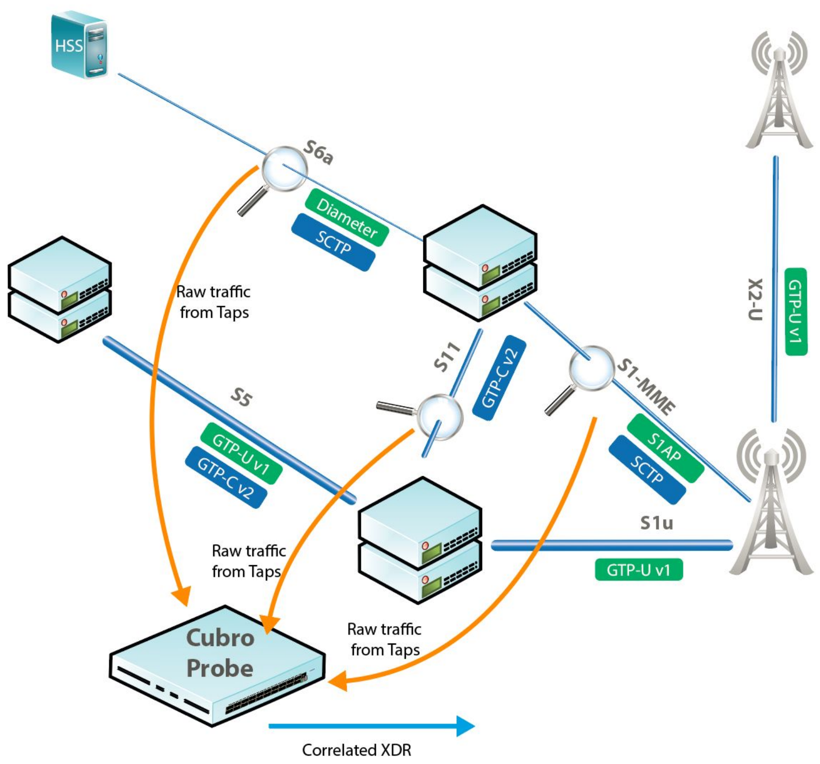
Geolocation Application in Mobile Networks

Challenge : Real time means a lot processing power is needed

Even in a smaller network there are multi-hundred thousand events per second.

All these events from all core interfaces (S1MME, S11, S6A, S3 , S10) must be captured, deciphered,

analyzed in real time to reach a high IMSI refill rate for each XDR



The probe produces several XDRs, typically one per interface, this is the first correlation stage on XDR level.

The second stage of correlation is to build one combined XDR.

This XDR holds several correlated information from multiple interfaces, like

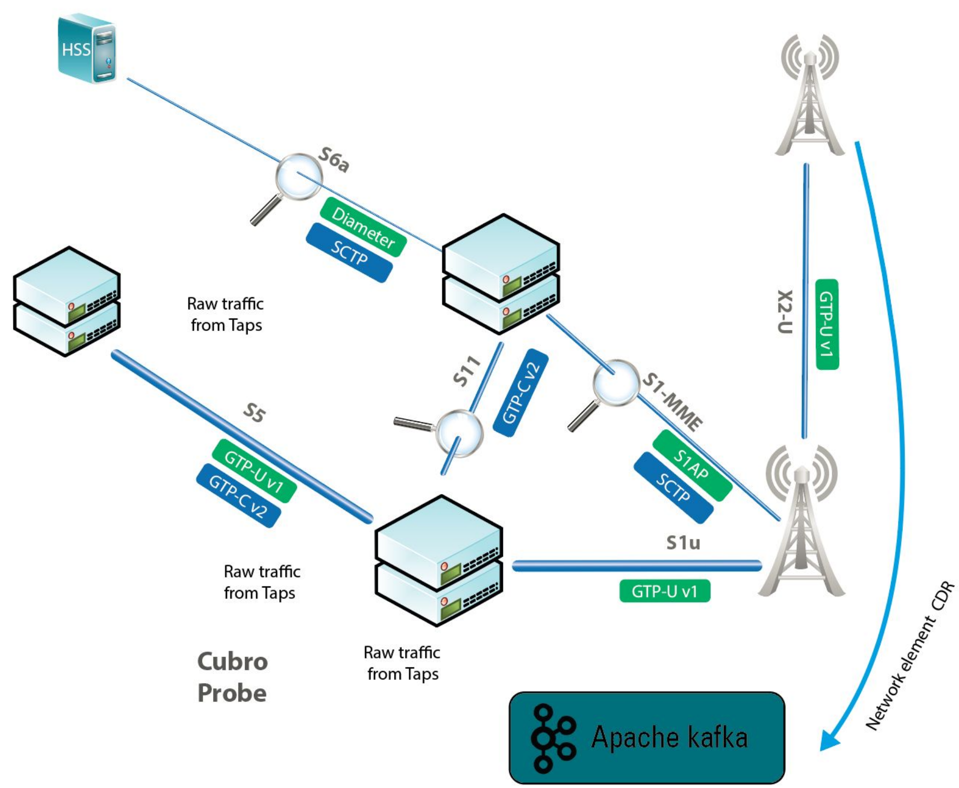
IMSI, IMEI, CELL ID, APN , Date, TIME and some more.

Geolocation Application in Mobile Networks

Challenge : Better location cell id means to get data from UTRAN interfaces or CDRs from eNodeB

To get this information there are two options - tap X2 interface or get the information from the eNodeB.

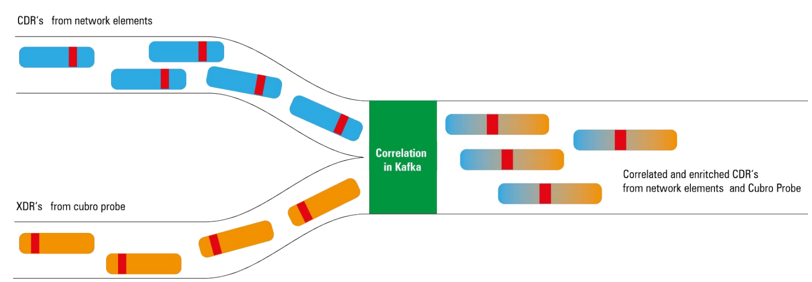
The difficulty is that this CDR from the eNodeB does not have all the information what is needed to fulfil the task.



This CDR delivers the cell id, the cell vector and the power but not much more.

That is the reason a third correlation stage is needed.

To correlate the core XDRs and the eNodeB CDRs to one resulting XDR.

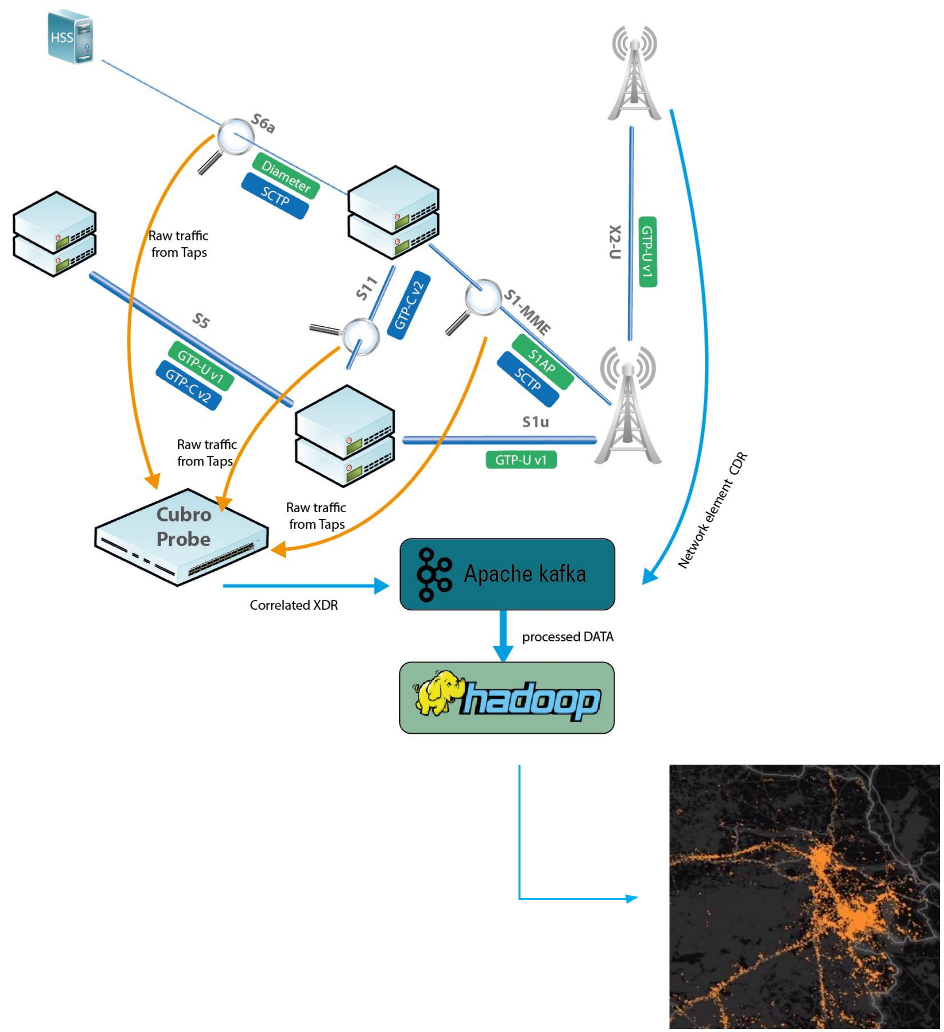


Geolocation Application in Mobile Networks

This third correlation stage is done in Apache Kafka cluster.

The resulting data is then stored in a hadoop cluster for further processing.

The real challenge is the quality of the XDRs, only with real time decoding of all interfaces - an IMSI refill rate of > 95 % is mandatory, to produce useful results in such an application.



The ROI on this project was 1 year.
Big Data could be treasure for MNO.
Cubro can help to harvest this !

Thank you

EMEA

Cubro Network Visibility
Ghegastraße 1030 Vienna,
Austria

Tel.: +43 1 29826660
Fax: +43 1 2982666399

Email: support@cubro.com

North America

Cubro US
337 West Chocolate Ave
Hershey, PA 17033

Tel.:717-576-9050
Fax.: 866-735-9232

Email: support@cubro.com

APAC

Cubro Asia Pacific

8, Ubi Road 2 #04-12 Zervex
Singapore 408538

Tel.: +65-97255386

Email: jl@cubro.com

Japan

Cubro Japan

8-11-10-3F, Nishi-Shinjuku,
Shinjuku,
Tokyo, 160-0023 Japan

Email: japan@cubro.com