



CREDO monitor: smartphone application data pre-processing

Oleksandr Sushchov

CREDO Week 2018, Cracow, 02 October 2018

CREDO: the quest for unexpected



CREDO Monitor: a citizen science tool for “fishing”

Requirements:

- Easy to get and to use
- Not boring
- Automated
- Functional

Expectations from the users:

- Training/tuning the algorithm(s)
- Helping the scientists
- Attracting new users of mobile/pocket detectors

Data processing

2 data types: **detections** and **pings**

1. **Detections'** acquisition (<https://github.com/credo-science/credo-webapp/tree/master/credoapiv2>) → **credo.txt**

```
{"detections": [{"user_id": 4125, "latitude": 41.6782314, "timestamp": 1535882719183, "altitude": 0.0, "visible": true, "longitude": 12.5179708, "height": 768, "frame_content": "i_IMAGE_ENCODING_ggg==", "width": 1024, "team_id": 829, "source": "api_v2", "provider": "network", "y": 615, "x": 178, "device_id": 3628, "id": 1299697, "time_received": 1535882720895, "accuracy": 72.9}, {"user_id": 4935, "latitude": 50.3516815, "timestamp": 1535882818143, "altitude": 0.0, "visible": false, "longitude": 18.8530087, "height": 720, "frame_content": "i_IMAGE_ENCODING_2_g==", "width": 1280, "team_id": 999, "source": "api_v2", "provider": "network", "y": 583, "x": 1, "device_id": 3751, "id": 1299698, "time_received": 1535882818739, "accuracy": 1125.0},
```

2. **Detections'** formatting → **general_info.txt**

Device_id	id	accuracy	alt	latitude	long	timestamp	provider	time_received	filter
3628	1299697	72.9	0.0	41.6782314	12.5179708	1535882719183	network	1535882720895	true
3751	1299698	1125.0	0.0	50.3516815	18.8530087	1535882818143	network	1535882818739	false
3751	1299699	1125.0	0.0	50.3516815	18.8530087	1535882865143	network	1535882866000	false
1312	1299700	33.0	0.0	49.4906744	19.0532604	1535882874059	network	1535882871241	true
1276	1299701	25.0	319.0	50.63121873	19.63160585	1535882735431	gps	1535882884382	true
3751	1299702	1125.0	0.0	50.3516815	18.8530087	1535882920334	network	1535882920950	false

Diagram illustrating data processing steps:

- active_list.txt**: Points to the **Device_id** column.
- timestamp.txt**: Points to the **timestamp** column.
- artefacts.txt**: Points to the **time_received** column.
- signals.txt**: Points to the **filter** column.

Data processing

3. **Pings'** acquisition (<https://github.com/credo-science/credo-webapp/tree/master/credoapiv2>) →
credo_ping.txt

```
{"pings": [{"user_id": 1153, "on_time": 600248, "timestamp": 1535882732878, "delta_time": 600103, "id": 599705, "time_received": 1535882734420, "device_id": 1600}, {"user_id": 4935, "on_time": 566687, "timestamp": 1535882746814, "delta_time": 600260, "id": 599706, "time_received": 1535882749006, "device_id": 3751}, {"user_id": 1163, "on_time": 600033, "timestamp": 1535882760187, "delta_time": 600033, "id": 599707, "time_received": 1535882765995, "device_id": 1570}, {"user_id": 4980, "on_time": 600166, "timestamp": 1535882801369, "delta_time": 600133, "id": 599708, "time_received": 1535882804451, "device_id": 3873}, {"user_id": 4125, "on_time": 600139, "timestamp": 1535882832574, "delta_time": 600155, "id": 599709, "time_received": 1535882839538, "device_id": 3201},
```

4. **Pings'** formatting → formatted.txt

User_id	ontime	timestamp	delta-time	id	time_received	device_id
4125	600139	1535882832574	600155	599709	1535882839538	3201
1849	600213	1535882854132	600416	599710	1535882857444	1779
4125	600098	1535882859799	600117	599711	1535882861385	3628
1510	565917	1535882861373	600206	599712	1535882863730	1397
3971	600267	1535882875502	600301	599713	1535882878347	3194
157	599940	1535882884533	600343	599714	1535882890222	3673

sum → ontime.txt

ping_list.txt

Data updating

1. Request edition

```
{  
  "since": 1535882713000,  
  "until": 1538474713000,  
  "limit": 500000,  
  "data_type": "detection"  
}
```

2. Time_received sorting

Timestamp	Time_received
1535882865143	1535882818739
1535882719183	1535882866000
1535882874059	1535882871241

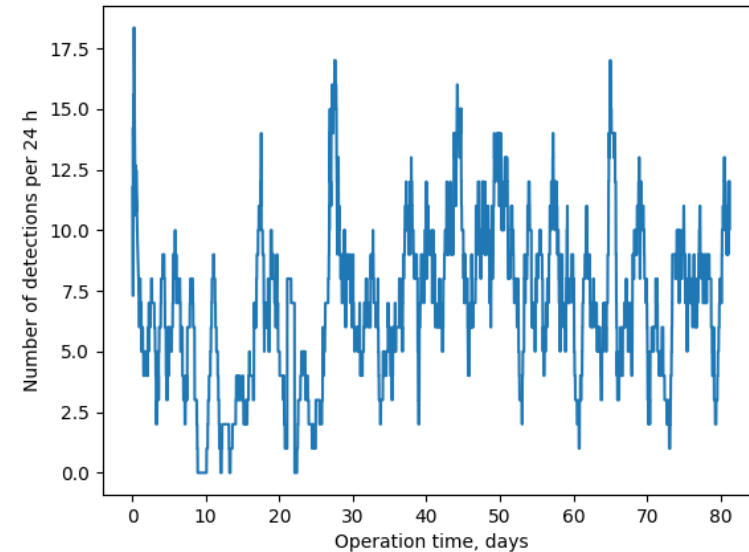
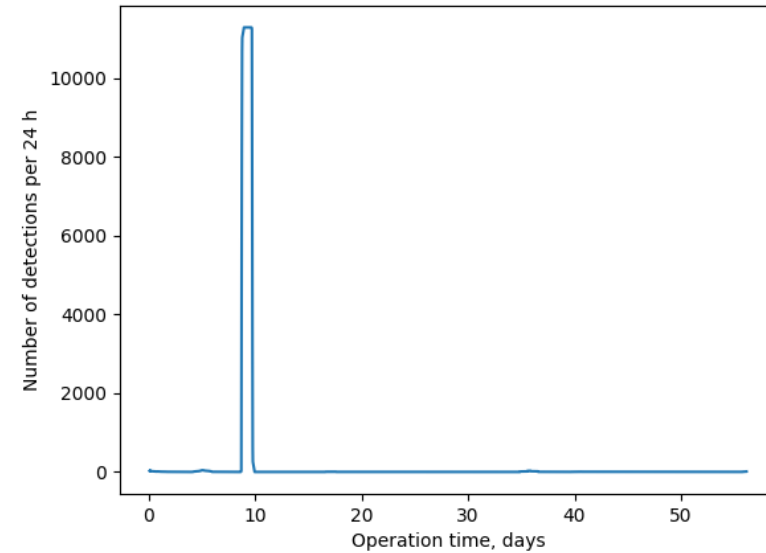
3. “Synchronization”

4. The previous steps with the recent data

Users selection

Cuts:

1. Detections' number > 400
2. Total ontime > 100 hours
3. Rate < 1000 events/24 h
4. Manual check



Accounting for ontime

1. **Timestamp sorting**
2. **Cumulative ontime calculation**
3. **Defining “edges” of 24h ontime parts**
4. **Splitting timestamp.txt → timestamp_i.txt**

Kevin's algorithms

**Plots, to be
presented on
Thursday**

Bottlenecks

1. Timestamps in ping processing

	Timestamp	Time_received	
~10 min	1534882832574	1535882839538	} ~0.5 sec
~10 min	1534883432578	1535882839660	
~10 min	1534884032580	1535882839775	
~10 min	1534884632573	1535882839890	
~10 min	1534885232572	1535882840007	
~10 min	1534885832569	1535882840122	

Bottlenecks

2. User_id vs. Device_id



Emanuele Maria Latorre

Team: [Divulgazione Libera](#)

Number of detections: 14154 (2nd place)

Time looking for particles: 5444h 59m (2nd place)

> 226 days;
Started operating
since 17 June



User detections

↕ date	↕ img
2018-10-02 15:00:26.611	
2018-10-02 14:57:33.510	
2018-10-02 14:38:22.910	

Outlook

1. Developing a user-friendly and easy-to-use interface
2. Reducing computational time (more efficient languages? DB?)
3. Improving algorithms (multiplets of higher order? Scrambled maps pregeneration?)
4. Improving the filtering algorithms (or adding new filters)

To get more information please visit:

<https://github.com/credo-science/credo-webapp>

<https://credo2.cyfronet.pl/redmine/projects/credo-monitor/wiki> - internal CREDO
space

Visit us on credo.science
and join CREDO science 😊

Cosmic-Ray Extremely Distributed Observatory (CREDO)

ables a strategy for a global analysis of cosmic-ray data to reach the sensitivity to extremely extended cosmic-ray phenomena, them super-preshowers, invisible for individual detectors or observatories. So far, the cosmic-ray research has been oriented detecting single air showers only, while the search for ensembles of cosmic-ray events induced by super-preshowers is a scientific terra incognita.

[Read More](#)