Tvheadend - Feature #4263

MediaHighWay EPG Grabber

2017-03-04 11:44 - Adam W

Status:	New		**Start date:**	2017-03-04
Priority:	Normal		**Due date:**	
Assignee:	Adam Sutton		**% Done:**	0%
Category:	EPG - Grabbers		**Estimated time:**	0.00 hour
Target version:	999			

Description

This was first requested a good few years ago now -

http://tvheadend.org/issues/156

MediaHighWay EPG (MHW). This is transmitted over the air by some satellite providers in Europe using two PIDs on a designated transponder (210 and 211 for MHW1 and 561-566 for MHW2) and carries 5-7 days of EPG data. These providers only carry now and next EIT information otherwise.

The system was created by the Canal group, so it's historically been available through satellite providers they originally owned - Canal Sat France, Movistar+ Spain (Canal Satélite Digital), Canal NL, Sky Italia (Telepiu), Cyfra+ Poland.

I have only looked at the two providers on 19.2E - Movistar+ and Canalsat France. Since going HD only a couple of years ago, Canal France does not seem to carry this data any more on their 'home' transponder (12363V), they must use the internet or something else. However Movistar+ is still carrying MHW2 data on PIDs 561-566 on 10847V. I think MythTV can use this data, and also Enigma2. I put an OpenSPA image on my Wetek Play which has MHW2 Movistar+ capability built in and successfully downloads the 7 day EPG from 10847V. I don't know about the other providers, but they were all MHW1 (PIDs 210 and 211), except for Movistar.

Looking at a couple of the PIDs for Movistar on 10847V in hex view in Transedit, you can clearly see the EPG data - programme names on 563 and descriptions on 566. The other PIDs have similar on them (I think series link, channel names/SIDs etc).

In order to download the full EPG, you must be tuned to the transponder (parsing the PID data) for 6 minutes in order to receive all the programmes. The data is transmitted as tables over the PIDs, one table corresponds to channel names, one table corresponds to programme genres, one table corresponds to channel names, etc.

Just thought I would open this ticket with all the information I know so far, if anyone knows more and can take a look at it. There's lots of stuff out there as MediaHighway has been widely supported elsewhere since the early 2000s.

If I can get to grips with C and processing MPEG TS tables then I may be able to take a look at it myself!

History

#1 - 2017-03-06 21:27 - Mono Polimorph

Hi,

Use project EPG Collector:
http://sourceforge.net/projects/epgcollector/

It does that you search for. It's actively maintained, supports SAT>IP and other tuners, and can export to WMC and XMLTV.
The only drawback is that's Windows only... However, you can run it in a Virtual Machine and use the SAT>IP server capabilities of the TVHE.

#2 - 2017-04-22 12:08 - Jaroslav Kysela

- Target version set to 999

#3 - 2020-10-31 16:52 - Adam W

I finally got round to looking into this and made a little progress in investigating the format. We should be able to use the MHW2 data for creating bouquets as well as getting the EPG.

For Movistar+ on 19.2°E, 10847V carries MHW2 data as described above.

Channel details for the bouquet, including NID, TSID and SID for each service are carried on PID 561 in the table with TID 200 (0xC8) -

```
TID:          0xc8
```

```
Section length: 2263
Table Ext: 0
Number of Channels: 107
NID:      1 | TSID:  1038 | SID: 30402 | LA 1
NID:      1 | TSID:  1008 | SID: 29812 | LA 2
NID:      1 | TSID:  1046 | SID: 30508 | ANTENA 3
NID:      1 | TSID:  1060 | SID: 30610 | CUATRO
NID:      1 | TSID:  1060 | SID: 30604 | TELECINCO
NID:      1 | TSID:  1052 | SID: 29862 | LA SEXTA
NID:      1 | TSID:  1008 | SID: 29816 | #0
NID:      1 | TSID:  1054 | SID: 30358 | M. SERIES
NID:      1 | TSID:  1038 | SID: 30410 | M.SERIESMANÍ
NID:      1 | TSID:  1008 | SID: 29807 | FOX
NID:      1 | TSID:  1008 | SID: 29815 | AXN
NID:      1 | TSID:  1038 | SID: 30411 | TNT
NID:      1 | TSID:  1038 | SID: 30408 | COMEDYCENTR.
NID:      1 | TSID:  1060 | SID: 30608 | CALLE 13
NID:      1 | TSID:  1046 | SID: 30512 | COSMO
NID:      1 | TSID:  1038 | SID: 30415 | AMC
NID:      1 | TSID:  1008 | SID: 29800 | FOX LIFE
NID:      1 | TSID:  1008 | SID: 29809 | AXN WHITE
NID:      1 | TSID:  1060 | SID: 30614 | SYFY
NID:      1 | TSID:  1052 | SID: 29858 | MTV
NID:      1 | TSID:  1034 | SID: 30652 | FDF
NID:      1 | TSID:  1032 | SID: 30206 | NEOX
NID:      1 | TSID:  1032 | SID: 30215 | DISNEY+
NID:      1 | TSID:  1038 | SID: 30400 | M.ESTRENOS
NID:      1 | TSID:  1034 | SID: 30662 | M.CINEDOC&RO
NID:      1 | TSID:  1008 | SID: 29804 | M. ACCIÓN
NID:      1 | TSID:  1008 | SID: 29805 | M. COMEDIA
NID:      1 | TSID:  1008 | SID: 29806 | M. DRAMA
NID:      1 | TSID:  1046 | SID: 30518 | M.CINEESP
NID:      1 | TSID:  1038 | SID: 30407 | TCM
NID:      1 | TSID:  1034 | SID: 30657 | HOLLYWOOD
NID:      1 | TSID:  1054 | SID: 30373 | SUNDANCE
NID:      1 | TSID:  1032 | SID: 30201 | PARAMOUNT
NID:      1 | TSID:  1060 | SID: 30615 | M.LALIGA
NID:      1 | TSID:  1042 | SID: 30059 | M. LALIGA 1
NID:      1 | TSID:  1042 | SID: 30080 | M. LALIGA 2
NID:      1 | TSID:  1054 | SID: 30364 | M. LALIGA 3
NID:      1 | TSID:  1042 | SID: 30075 | M.LCAMPEONES
NID:      1 | TSID:  1046 | SID: 30516 | #VAMOS
NID:      1 | TSID:  1060 | SID: 30607 | M. DEPORTES
NID:      1 | TSID:  1052 | SID: 29860 | MOVISTAR F1
NID:      1 | TSID:  1060 | SID: 30601 | M. GOLF
NID:      1 | TSID:  1060 | SID: 30606 | #VAMOS BAR
NID:      1 | TSID:  1052 | SID: 29857 | EUROSPORT 1
NID:      1 | TSID:  1052 | SID: 29863 | EUROSPORT 2
NID:      1 | TSID:  1052 | SID: 29859 | GOL
NID:      1 | TSID:  1038 | SID: 30412 | TELEDEPORTE
NID:      1 | TSID:  1046 | SID: 30507 | CAZA Y PESCA
NID:      1 | TSID:  1008 | SID: 29802 | IBERALIA TV
NID:      1 | TSID:  1042 | SID: 30064 | TOROS
NID:      1 | TSID:  1060 | SID: 30616 | R. MADRID TV
NID:      1 | TSID:  1052 | SID: 29856 | BARÇA TV
NID:      1 | TSID:  1060 | SID: 30605 | NAT GEOGRAPH
NID:      1 | TSID:  1008 | SID: 29810 | NAT GEO WILD
NID:      1 | TSID:  1046 | SID: 30513 | HISTORIA
NID:      1 | TSID:  1046 | SID: 30511 | DISCOVERY
NID:      1 | TSID:  1060 | SID: 30602 | CANAL ODISEA
NID:      1 | TSID:  1046 | SID: 30509 | BLAZE
NID:      1 | TSID:  1038 | SID: 30409 | VIAJAR
NID:      1 | TSID:  1052 | SID: 29855 | ENERGY
NID:      1 | TSID:  1046 | SID: 30514 | CRIMEN
NID:      1 | TSID:  1038 | SID: 30414 | CANAL COCINA
NID:      1 | TSID:  1060 | SID: 30613 | CANAL DECASA
NID:      1 | TSID:  1052 | SID: 29850 | DIVINITY
NID:      1 | TSID:  1008 | SID: 29801 | NOVA
NID:      1 | TSID:  1052 | SID: 29864 | BE MAD
NID:      1 | TSID:  1052 | SID: 29852 | BABY TV
NID:      1 | TSID:  1008 | SID: 29803 | DISNEY JR
NID:      1 | TSID:  1038 | SID: 30404 | CANAL PANDA
NID:      1 | TSID:  1054 | SID: 30359 | NICK JR
NID:      1 | TSID:  1046 | SID: 30510 | NICKELODEON
NID:      1 | TSID:  1038 | SID: 30403 | DISNEY CH.
```

```
NID:     1 | TSID:  1052 | SID: 29854 | BOING
NID:     1 | TSID:  1038 | SID: 30401 | CLAN TVE
NID:     1 | TSID:  1078 | SID: 28656 | VH1
NID:     1 | TSID:  1054 | SID: 30361 | MEZZO
NID:     1 | TSID:  1064 | SID: 30755 | MEZZO HD
NID:     1 | TSID:  1032 | SID: 30210 | CLASSICA
NID:     1 | TSID:  1046 | SID: 30520 | 24 HORAS
NID:     1 | TSID:  1002 | SID:  5001 | BBC WORLD
NID:     1 | TSID:  1028 | SID:  4422 | CNN INT.
NID:     1 | TSID:  1052 | SID: 29851 | FOX NEWS
NID:     1 | TSID:  1091 | SID: 31220 | EURONEWS
NID:     1 | TSID:  1028 | SID:  4440 | AL JAZEERA
NID:     1 | TSID:  1022 | SID:  6906 | FRANCE 24
NID:     1 | TSID:  1012 | SID:  6382 | RTESPAÑOL HD
NID:     1 | TSID:     9 | SID:   125 | CNBC HD
NID:     1 | TSID:  1022 | SID:  6915 | TV5MONDE
NID:     1 | TSID:  1026 | SID: 10067 | BLOOMBERG
NID:     1 | TSID:  1111 | SID:  7290 | SKY NEWS
NID:     1 | TSID:  1012 | SID:  6383 | RT FRANCE HD
NID:     1 | TSID:  1020 | SID:  7008 | CUBAVISIÓN
NID:     1 | TSID:  1040 | SID: 31304 | TELESUR
NID:     1 | TSID:  1002 | SID:  5021 | NHK WORLD
NID:     1 | TSID:  1020 | SID:  7011 | ARIRANG TV
NID:     1 | TSID:  1060 | SID: 30603 | PLAYBOY TV
NID:     1 | TSID:  1046 | SID: 30505 | CANAL SUR A.
NID:     1 | TSID:  1012 | SID:  6385 | TV GALICIA
NID:     1 | TSID:  1008 | SID: 29811 | ARAGÓNTV INT
NID:     1 | TSID:  1054 | SID: 30350 | ALQUILER  1
NID:     1 | TSID:  1054 | SID: 30351 | ALQUILER  2
NID:     1 | TSID:  1042 | SID: 30055 | ALQUILER  3
NID:     1 | TSID:  1042 | SID: 30063 | ALQUILER  4
NID:     1 | TSID:  1042 | SID: 30053 | ALQUILER  5
NID:     1 | TSID:  1042 | SID: 30054 | ALQUILER  6
NID:     1 | TSID:  1050 | SID: 30804 | ALQUILER HD
NID:     1 | TSID:  1050 | SID: 30814 | ALQUILER2 HD
```

The table ext is the first byte of table data after the 3 bytes of TID and length etc. Ext 0 carries what looks like the SD bouquet, then Ext 2 is with HD channels swapped in (e.g. LA 1 HD instead of LA 1). Ext 3 seems to carry details of the Spanish digital terrestrial DTT channels (NID 8916)!

To start to figure out the decoding of the data, I've been using the Enigma2 code here which decodes each section -
https://github.com/openatv/enigma2/blob/0627006482416a94693d232c36f04a4d5debb7d7/lib/dvb/epgcache.cpp

Will need to figure out how to make use of this data in TVHeadend and write a grabber to get the programme names/summaries once I figure out how they work too.

#4 - 2020-10-31 17:03 - Adam W

To get the data above, you find tables with TID 200 (0xC8) on PID 561, then you look for byte 117 of the table data (after the header) which gives the number of channels (N) in the bouquet. After this byte, there are 8 bytes for each channel - 2 bytes network ID (e.g. 0x00 0x01 for Astra 1 with NID 1), 2 bytes TSID, 2 bytes SID, and 2 bytes unknown. At the end of this data, so byte 117+(8*N) there are the channel names, with the 1 byte before each name containing the length of the name in the last four bits (byte & 0xF).

Next to investigate is the programme names table (PID 644 TID 220 for 7 day EPG) and programme descriptions table (PID 642 TID 150).

#5 - 2020-11-02 18:18 - Adam W

Update - the channel names table is also present on PID 644, as are some of the programme descriptions (short ones). So you can get partial data from just that single PID. I've been investigating so far with a Python script and got information like this just from parsing the tables (150, 200 and 220) on PID 644 -

```
"81767282": {
    "channel": {
      "nid": 1,
      "tsid": 1050,
      "sid": 30819,
      "name": "LA 1 HD"
    },
    "start": "24/10/2020 20:00:00",
    "end": "24/10/2020 20:25:00",
    "title": "Telediario 2 Fin de Semana",
    "summary":
"Pres: Lluís Guilera. Informativo de Televisión Española de los fines de semana. Lluís Guilera y Lara Siscar s
e encargan de presentar las últimas noticias de ámbito nacional e internacional. Marcos López, mientras, desgr
ana la actualidad deportiva. "
  },
```

642 contains the description data where the programme has a longer description.

#6 - 2020-11-03 14:44 - Adam W

- File mhw2.py added

- File data_03-11.json added

Now getting all of the data, with a Python script based on the Enigma2 code here:
https://github.com/openatv/enigma2/blob/0627006482416a94693d232c36f04a4d5debb7d7/lib/dvb/epgcache.cpp

Using the TSDuck tstables tool (https://tsduck.io) to dump the tables from PIDs 642 and 644 to a binary file, from 360 seconds (six minutes) of dumped transport stream from TVHeadend. Then running the binary dump through the Python tool (and jq to pretty print the JSON) to get the EPG data.

```
curl -m 360 http://tvh_ip:9981/stream/mux/astra1_10847v_mux_link?pids=642,644 > 10847_mhw2.ts
tstables -p 642 -p 644 -t 150 -t 200 -t 220 -b mhw2_dump.bin 10847_mhw2.ts
cat mhw2_dump.bin | python3 mhw2.py | jq > data_03-11.json
```

I get ~18000 programmes, most of which have a subtitle/description that matches. Around 5000 don't, but most of these have a 'generic' description ID which I guess means they don't actually have a description. There are a handful of others where there might be a description that hasn't been parsed, or maybe 6 minutes isn't quite long enough to grab all the data.

This script could be reconfigured to output XMLTV I guess! But also should be possible to eventually make an OTA grabber directly in TVHeadend - I will get there!

#7 - 2020-11-04 17:27 - saen acro

- File EIT.json added

Current JSON result is similar to this one with come out from Mumudvb

Files

PID566.png	53.3 KB	2017-03-04	Adam W
PID563.png	49.3 KB	2017-03-04	Adam W
mhw2.py	6.05 KB	2020-11-03	Adam W
data_03-11.json	8.15 MB	2020-11-03	Adam W
EIT.json	350 KB	2020-11-04	saen acro