DG LAB

Mempool Analysis & Simulation

Karl-Johan Alm
@kallewoof

C42A FF7C 61B3 E44A 1454 CD35 57AF 762D B335 3322

Agenda

- Why?
- What?
- How?
- So!

Why?

Background

"*Optimizing fee estimation via the mempool state*", Scaling Stanford 2017[1]

No tools to do fee rate analysis.

Unable to make comparisons of different strategies.

Even with ZMQ logs data is lost. Orphaned blocks & txs. Why care? Because they are missing pieces of a complete re-enactment of some point in time.

Want a way to record, and playback, the mempool.

[1]https://scalingbitcoin.org/stanford2017/Day2/Scaling-2017-Optimizing-fee-estimation-via-the-mempool-state.pdf

Why record/playback the mempool?

- Loss of information: timestamps, blocks, transactions.
- No good answer to "what happened at *t*=X..Y"
- No good way to simulate fee estimators
- No public information on what harvesters gather from mempool analysis.
- No good way to gauge "spam" vs "organic use".
- What prt of txs are likely miners' (i.e. not broadcasted but mined directly)
- MFF addresses this & as a bonus also addresses assumption that Bitcoin is somehow anonymous. (It isn't.)

We have no recording of the *mempool*, only of the resulting *chain*.

What?

A new tool for mempool analysis

MFF (**M**empool **F**ile **F**ormat)

- logs time of (re-)entry/exit/confirmation/invalidation
- logs entire raw data for transactions that were replaced (RBF, 2x-spend, ..)
- logs chain tip changes (block mined/orphaned, & which txs were in it)
- can seek on a per-block basis, but "find tx X" requires O(*n*), *n*=entire db

Library implementation is called *libbcq*, and is built on top of a database format called *CQDB*.

A new tool for mempool analysis

Client Type	Downloads	Keeps
Light Clients	Interesting blocks	Nothing
Pruned Full Nodes	All blocks & recent txs	Recent confirmed blocks & unconfirmed txs
Full Nodes	All blocks & recent txs	All confirmed blocks & unconfirmed txs
↑ MFF enabled	All blocks & recent txs	All blocks, unconfirmed + invalidated txs retaining order

A new tool for mempool analysis

Client Type	Downloads	Keeps
Light Clients	Interesting blocks	Nothing
Pruned Full Nodes	All blocks & recent txs	Recent **confirmed** blocks & unconfirmed txs
Full Nodes	All blocks & recent txs	All **confirmed** blocks & unconfirmed txs
↑ MFF enabled	All blocks & recent txs	All blocks, unconfirmed + **invalidated** txs retaining **order**

MFF so far (tiny mempool ZMQ dump)

Source	ZMQ dumps w/o block hex (only block hash); tiny mempool setting (10k tx cap)
Period	June 18 2018 ~ May 27 2019 (313 days, block #532421 ~ #578042, 45622 blocks)
Size on disk	6.8 GB (between 200-400 MB/cluster, avg 287 MB) ~> 22 MB/day
Entries	274822087 (274.8 million), with 16073 tx invalidations
Count dist	tx in=52.6% (23.3% ref), tx out=47.4%, tx invdt=0.01%, block mined=0.02%
Byte dist	tx in=84.8% (3.6% ref), tx out=7.6%, tx invdt=0.09%, block mined=7.5%
Top ref tx	db9539c40343c5c47bdaaa53e11e735dce3526daca8824476f5c10128e686ce4 (1901 refs)

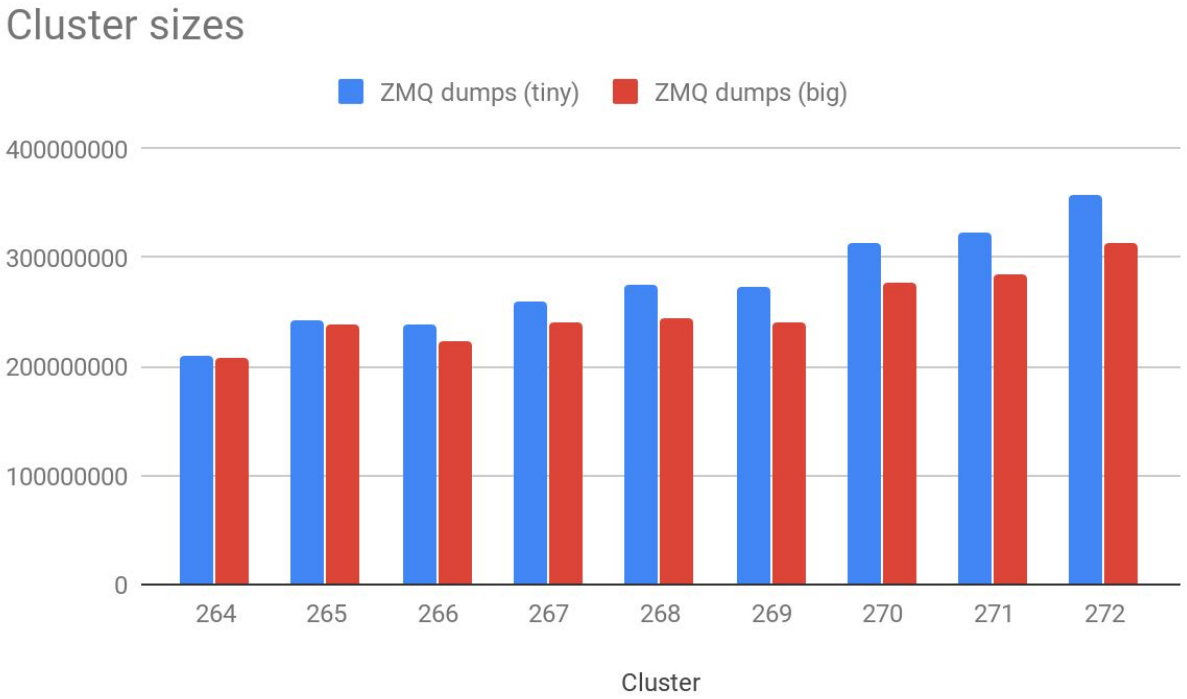
MFF so far (bigger mempool ZMQ dump)

Source	ZMQ dumps w/o block hex (only block hash); bigger mempool setting (200k tx cap)
Period	June 18 2018 ~ Nov 28 2018 (133 days, block #532421 ~ #551861, 19441 blocks)
Size on disk	6.0 GB (between 200-230 MB/cluster, avg 220 MB) ~> 15 MB/day
Entries	31758780 (31.8 million), with 55101 tx invalidations
Count dist	tx in=99.23% (1.34% ref), tx out=0.36%, tx invdt=0.16%, block mined=0.06%
Byte dist	tx in=94.49% (0.07% ref), tx out=0.03%, tx invdt=0.79%, block mined=3.78%
Top ref tx	c529e5b79ec7216c97b03c71cd5d0c60c6e087a7b5d7a428167baa6d3b011f35 (1434 refs)

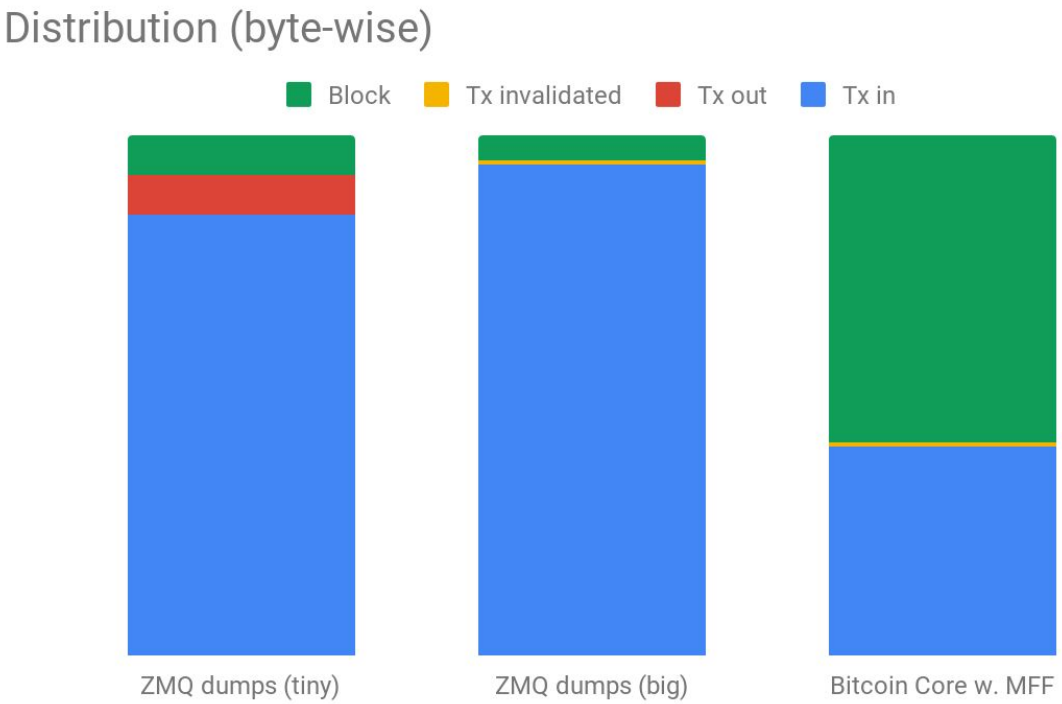
MFF so far (Bitcoin Core with MFF)

Source	Bitcoin network via patched Bitcoin Core (default settings)
Period	June 2 2019 ~ June 7 2019 (5 days, block #578885 ~ #579642, 758 blocks)
Size on disk	77 MB ~> 15 MB/day (~220 MB/cluster)
Entries	353487 (353k), with 1054 tx invalidations
Count dist	tx in=99.49% (0% ref), tx out=0%, tx invdt=0.30%, block mined=0.21%
Byte dist	tx in=40.43% (0% ref), tx out=0%, tx invdt=0.59%, block mined=58.98%
Top ref tx	da8bbd861efb37ccbae748b9eba7081caf9aad920658f0c480fa2733e1a8db74 (353 refs)

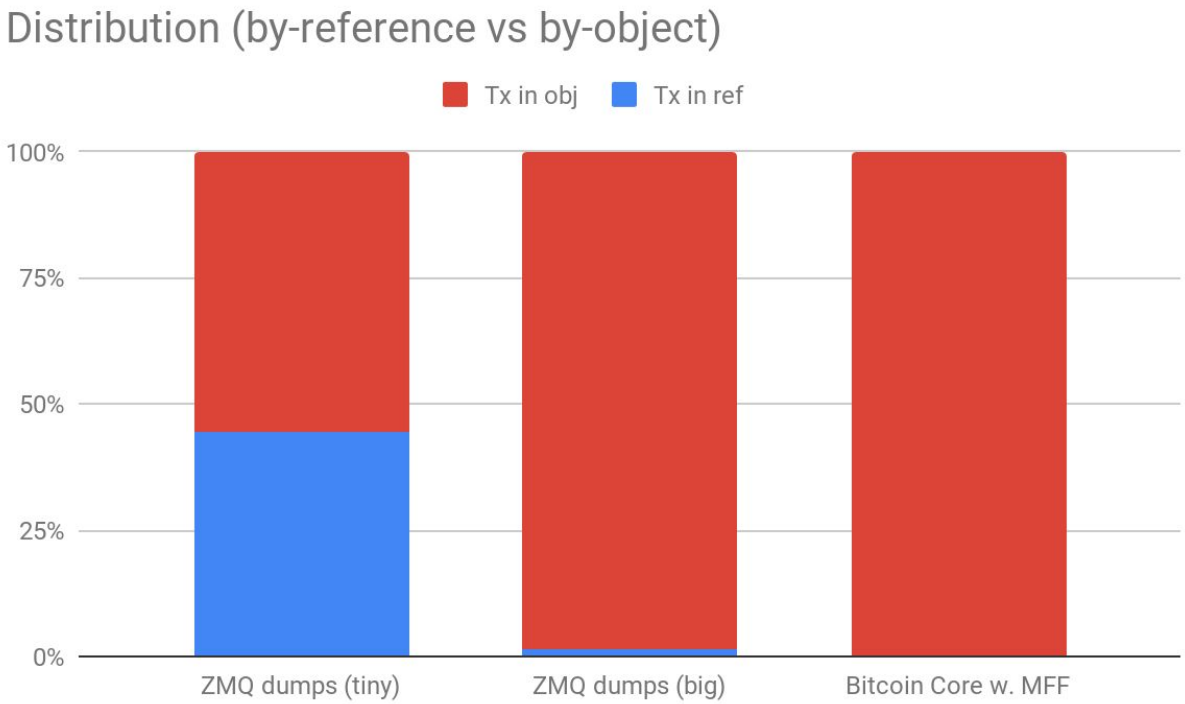
MFF so far


Cluster sizes
ZMQ dumps (tiny)
ZMQ dumps (big)
400000000
300000000
200000000
100000000
0
264
265
266
267
268
269
270
271
272
Cluster

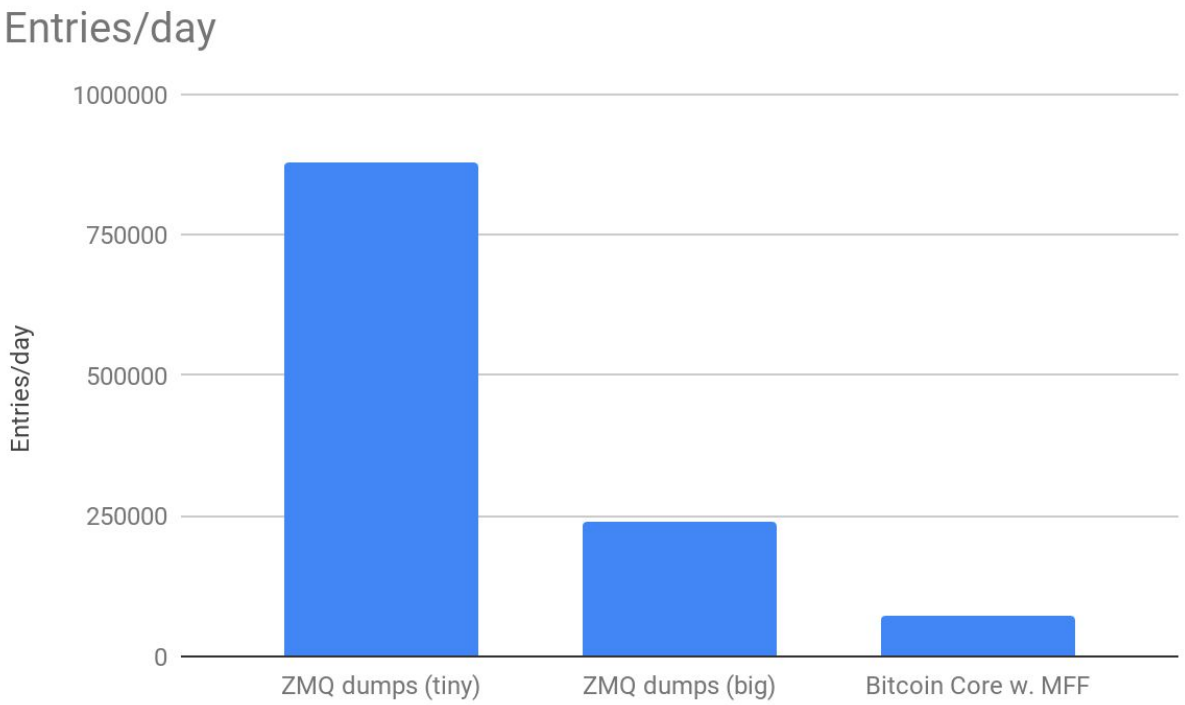
MFF so far


Distribution (byte-wise)
Block
Tx invalidated
Tx out
Tx in
ZMQ dumps (tiny)
ZMQ dumps (big)
Bitcoin Core w. MFF

MFF so far


Distribution (by-reference vs by-object)
Tx in obj
Tx in ref
100%
75%
50%
25%
0%
ZMQ dumps (tiny)
ZMQ dumps (big)
Bitcoin Core w. MFF

MFF so far


Entries/day
1000000
750000
500000
250000
0
Entries/day
ZMQ dumps (tiny)
ZMQ dumps (big)
Bitcoin Core w. MFF

db9539c40343c5c47bdaaa53e11e735dce3526daca8824476f5c10128e686ce4

0.54127445 BTC

Aug 22, 2016 5:27:46 AM

152790 Confirmations

Input Amount	0.57567445 BTC	**Inputs**	1
Output Amount	0.54127445 BTC	**Outputs**	2,501
Included in Block	426246	**Block Index**	1103
Fee	0.0344 BTC	**Size**	78.6 kB
Version	1	**Fee Size**	42.73 sat/B

3NrDMb926RAtn5tePWVS4xt3X4o6DP3EX5	0.57567445 BTC	1PzZY7BmiRCxUDyvsLnGCmPqrwsDhwL5QU	0.00010244 BTC
		1Ff4Lqkdu2PYM6vEqyeWiCYutJyzT2jNHQ	0.00010247 BTC
		1PV8vCySAt8Y2tugKLaGWcDqbR7pzbxKj5	0.00010275 BTC
		1771bDgVt5dV9WQstqJ2bYGzL2kY6eUe9Q	0.00010261 BTC
		1Pi6c5BKmQkp7wnybS2GZ5piUm46YZrYUk	0.00010115 BTC
		1P5KznQ347F9FVofr6pUJtBrwVquYqGPUo	0.00010025 BTC
		1EUuPg7XnHRTTgcb3DeBcCjse8oV36zk7h	0.00010201 BTC
		1NyaaBMfBenKE64sgKcaY8zjCTH3RCStdK	0.00010146 BTC
		14iPjkMWAafU3QL64vbjdjoXw8F4ZywA2k	0.00010165 BTC
		3LPkj9tN3A8FoLp4jbNjQPvBZvhaC8PYgU	0.00010051 BTC
		3CmLkG9rSoceVPbjHAjJhW1BK5DTsPwn7f	0.00010066 BTC

c529e5b79ec7216c97b03c71cd5d0c60c6e087a7b5d7a428167baa6d3b011f35

173.58652544 BTC

Jul 28, 2018 4:38:51 AM

45280 Confirmations

Input Amount	173.58652544 BTC	**Inputs**	1
Output Amount	173.58652544 BTC	**Outputs**	4,766
Included in Block	534030	**Block Index**	2
Fee	0.00 BTC	**Size**	156.5 kB
Version	2	**Fee Size**	0.00 sat/B

1GX28yLjVWux7ws4UQ9FB4MnLH4UKTPK2z	173.58652544 BTC	12Gv5KBYFeCWM1KK6bLgE7EMrvCqcfA3bj	6.2469019 BTC
		18Q1iaCnFHvWyUtyEGQ7kz4Ffu3o6LYbzj	4.68057245 BTC
		1PX6TvLnivTJ6EMSNyhSPQvrPkpMpMAYZC	3.20220787 BTC
		162Ds8SY7drJSY5wQZ2ZV8ke5d6DBB5gHE	2.96772044 BTC
		32nMmncHbnoh4X9EmbBNFpkiEnAe9cRpbz	2.64923419 BTC
		1JrUa7rP7gEpCQxkqddnkTF6G2AThESLUX	2.54528511 BTC
		1JdzjkxN9pAhmfRT6148UHsAPLM4QPYPqu	2.41818249 BTC
		14pF6frsf1E7Eh2gmqc7PDoDVLKU8uFC9n	2.33733515 BTC
		36EDpcgtkwwQEBooE2hqssb76d7vSze31k	2.29534299 BTC
		3Qyc2GqLczysrXKCMPeqUXDciyCZ42836q	2.19661271 BTC
		1373KHZNtbUwu2cEZFU7sTDky17uN1YY3b	2.19129749 BTC

da8bbd861efb37ccbae748b9eba7081caf9aad920658f0c480fa2733e1a8db74

209.29517266 BTC

Jun 2, 2019 2:54:51 AM

439 Confirmations

Input Amount	209.29517266 BTC	**Inputs**	1
Output Amount	209.29517266 BTC	**Outputs**	4,500
Included in Block	578872	**Block Index**	61
Fee	0.00 BTC	**Size**	147.1 kB
Version	2	**Fee Size**	0.00 sat/B

1GX28yLjVWux7ws4UQ9FB4MnLH4UKTPK2z	209.29517266 BTC	3DQGSZMJHJ4gxR9yW2WsVcFL9LZMrKPhco	0.07162107 BTC
		1JDAQG74AxNxtriRy2P5kJHXsuCgt8sd3R	0.00923398 BTC
		1LBtrCwKCkxXCpJWFgo37CptgR4Lf8nEzJ	0.06172097 BTC
		381fJVzBTcFCCQ5tB1dFFCZJv15LVeNe32	0.02458812 BTC
		1HUP763TixY7YLoqVHsFhMte892DUJZuU2	0.00614025 BTC
		1GTxqjtoHxHVEEryiAQdboezUNqBBG6Du8	0.01368035 BTC
		32PRzKARxD8dngcpL5PKyYaNFfMetjzBxH	0.02024015 BTC
		19CrwBmmEjJMjgVQh3goV2TDWfdcb2uRk8	0.00624168 BTC
		3EoRexR6228jaXHqrStRRN8vg9u6JTXHj8	0.00524562 BTC
		3JgLNo1SonBkGLc9SQaAbWPjEA7W4zgpBz	0.00726441 BTC
		1FUT4DmgtHUneDLdCHpE9v8m5eLc9BgFpD	0.00752153 BTC

How?

Brief overview

3 components, on top of each other:

Component	Description
CQDB	Seekable Sequential (**C**-kable Se**q**uential) DB (lib & spec)
BCQ	Bitcoin CQ (specialization of CQ for Bitcoin)
Implementations	libbcq branch (Bitcoin Core), MFF toolset (mff-findtx, …), etc.

CQDB

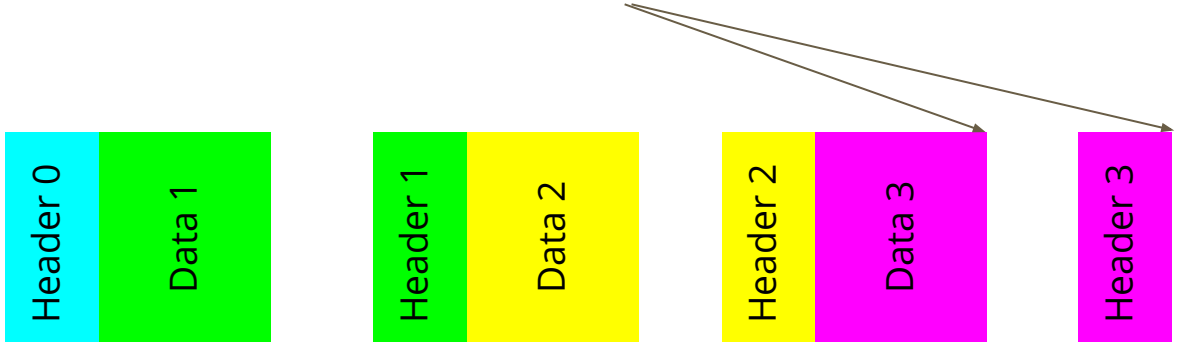
- Light-weight, space and memory efficient sequential database
- Data stored in independent clusters, each with a range of segments.
- Append-only. Chronological time restriction.
- Objects are stored on first reference, and referenced subsequently.

CQDB

Clusters stored as blocks of header+data pairs. Because of append-only nature, the header for the current cluster is actually stored as the header for (cluster + 1).

CQDB

Append-only, chronological → write index and data simultaneously, once.



CQDB

Serialize objects once, then use references to point back at their byte position 2nd+ time. Reader chooses what to remember. Seek back and re-deserialize on demand.

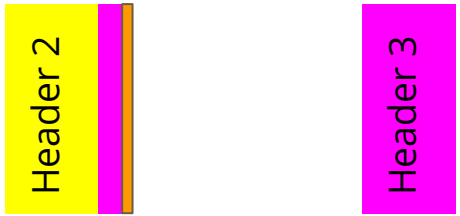
BCQ

BCQ is a CQDB where

- each segment corresponds to a block in the blockchain
- each cluster is 2016 blocks (i.e. one retargeting period)
- objects are transactions or references to such (e.g. outpoints)

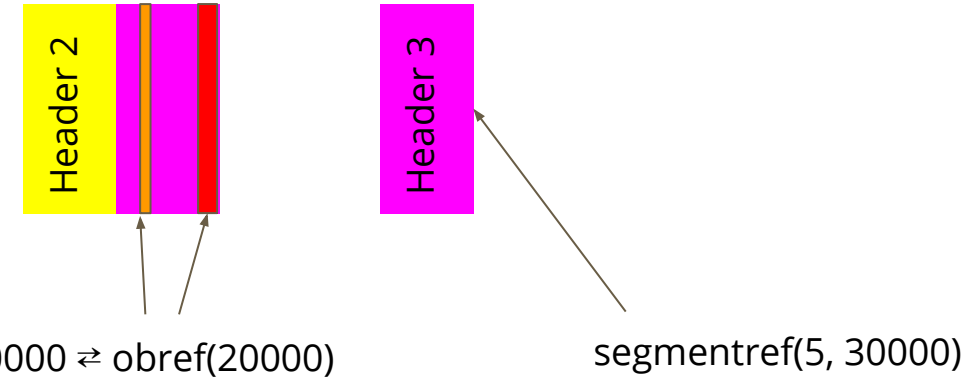
BCQ

Write txid 36e2f[...]384b into cluster 3, starting at byte position 10000.



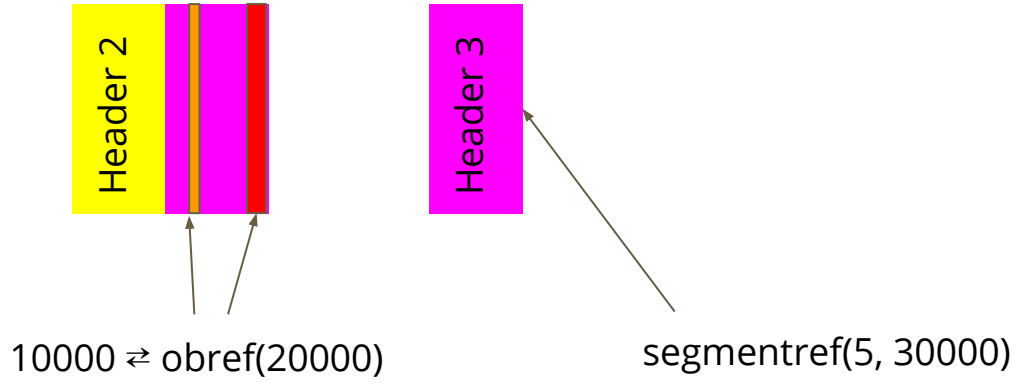
BCQ

Write txid 36e2f[...]384b into cluster 3, starting at byte position 10000.
Reference txid 36e2f[...]384b for block #5 inclusion at byte position 30000.
Reference is written as 20000 as a varint (`0x809b20`), the offset.
Also writes segment 5
ref to end of header 3.



BCQ

When I read block #5, I get "this tx is at <block start>-20000". So tx 36e2f... is aka "tx 10000". If I remember "tx at 10000", I am fine. If not, and I want/need it, I can seek back and read it.



BCQ

BCQ available as a patch for Bitcoin Core at:

https://github.com/kallewoof/bitcoin/tree/libcq

CQDB (libcqdb) is at: https://github.com/kallewoof/cqdb

MFF (libbcq) is at: https://github.com/kallewoof/mff

So!

52a72025d071a53b6fb59ffb040d7eb134dccb2cdef7214f1c6965a00ba270e0

1,546.67342722 BTC

Jun 2, 2019 12:20:55 PM

139 Confirmations

Hash	07538e610bc14d40a0f1cbe1a5da55b0a20af2f0a5656986bb751dc232112b54	**Inputs**	38
Input Amount	1,546.67929241 BTC	**Outputs**	15
Output Amount	1,546.67342722 BTC	**Block Index**	1357
Included in Block	578883	**Size**	6 kB
Fee	0.00586519 BTC	**VSize**	3 kB
Version	2	**Fee Size**	191.11 sat/B

bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	14.10362556 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	1.8 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	13.37758094 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	2.5 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	14.4719292 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	2.8 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	14.98231079 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	9.5 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	13.23596081 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	15.6 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	14.58976187 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	19.4 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	14.21744259 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	20.4 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	12.96116887 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	26.3 BTC
bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh54v3dn	14.25072295 BTC	3FxUA8godrRmxgUaPv71b3XCUxcoCLtUx2	28.3 BTC

Kalle Alm @kallewoof · 20h

.@btccom_official what on earth are you doing?
smartbit.com.au/tx/52a72025d07...

I can't think of a single reason why you would wanna send multiple times to the same address within the same transaction.

Bitcoin Transaction 52a72025d071a53b6fb59ffb040...
Smartbit Bitcoin Transaction. A bitcoin transaction is a value transfer between two addresses. Bitcoin ...
smartbit.com.au

2 3

BTC.com
@btccom_official

Replying to @kallewoof

split utxo for faster payouts. 1. max tx/tx chain size: 100KB 2. max fee: 0.1 btc

6:23 PM · Jun 2, 2019 · Twitter Web Client

What's it good for?

- Educational for people learning how Bitcoin works (e.g. seeing the flow of a transaction being RBF-bumped or double spent)
- Useful in general for scientific purposes, such as writing better algorithms for fee rate estimation, or analyzing spam vs not spam.
- Improved transparency (we know more precisely what *they* know)

A "double spend" (not really)

```
$ ./mff-findtx ~/.bitcoin-mff/mff 13c724dd61092d742c94295bd8e1ba9afce46815d559ae9ecd306ca42a992cbd
Sun Jun  2 03:51:54 2019: ----log begins----
Sun Jun  2 03:53:08 2019: mempool_in (first seen 13c724dd61092d742c94295bd8e1ba9afce46815d559ae9ecd306ca42a992cbd - 486 vbytes, 200000 fee, 411.523 fee rate (sat/vb), block #578887)
Sun Jun  2 04:24:50 2019: mempool_invalidated 13c724dd61092d742c94295bd8e1ba9afce46815d559ae9ecd306ca42a992cbd (replaced)
tx(hash=13c724dd61, ver=2, vin.size=3, vout.size=1, locktime=578887)
    txin(outpoint(86e715cfd4c4e82c9cb0418adccad738eb86336612a6b4a705cbc61a51c26567, 1), scriptSig=473044022052624926e2afe4, sequence=4294967293)
    txin(outpoint(d9dd89ae898cee34a6f7b4f0ee774ea27ab732e9ddf5ce824eea55be56e2278e, 0), scriptSig=47304402200675cc881a821c, sequence=4294967293)
    txin(outpoint(f93f39388f43df1c91dd990ba8871af66b330cdb284b3c89d96920f546997017, 0), scriptSig=483045022100af690b703bf1, sequence=4294967293)
    scriptWit()
    scriptWit()
    scriptWit()
    txout(value=3.07139667, scriptPubKey=76a91492e733d0f034cbb4a537deaa)

Sun Jun  2 04:30:53 2019: mempool_invalidated 4f6cbe80d787dba5e0461235c0d1e0e994620d584bdec4aa0a1d55da1d481550 (conflict)
tx(hash=4f6cbe80d7, ver=2, vin.size=3, vout.size=1, locktime=578887)
    txin(outpoint(86e715cfd4c4e82c9cb0418adccad738eb86336612a6b4a705cbc61a51c26567, 1), scriptSig=473044022060539516fdcfda, sequence=4294967293)
    txin(outpoint(d9dd89ae898cee34a6f7b4f0ee774ea27ab732e9ddf5ce824eea55be56e2278e, 0), scriptSig=473044022064797405b7a7f, sequence=4294967293)
    txin(outpoint(f93f39388f43df1c91dd990ba8871af66b330cdb284b3c89d96920f546997017, 0), scriptSig=483045022100ab0995e86f88, sequence=4294967293)
    scriptWit()
    scriptWit()
    scriptWit()
    txout(value=3.07039667, scriptPubKey=76a91492e733d0f034cbb4a537deaa)
 -> 13c724dd61092d742c94295bd8e1ba9afce46815d559ae9ecd306ca42a992cbd
Sun Jun  2 04:30:53 2019: block_mined (13c724dd61092d742c94295bd8e1ba9afce46815d559ae9ecd306ca42a992cbd in #578888=00000000000000000002235fedbc4aae14e714a8de6974a0de6696b440872909c)
Sun Jun  2 04:32:34 2019: mempool_in (first seen 99f263fa8f28655d71d173a485f56582b5305d67fbe3186ffaea770e4be72b67 - 223 vbytes, 448 fee, 2.009 fee rate (sat/vb), block #578888)
Sun Jun  2 05:17:22 2019: mempool_in (first seen d79eb11d5e4e5e95a989f9ec997c454395a054ddbee6383653826c3785b5cf70 - 222 vbytes, 200000 fee, 900.901 fee rate (sat/vb), block #578890)
```

Thank you for your time

Questions?

Github links etc:

CQDB: https://github.com/kallewoof/cqdb
BCQ/MFF: https://github.com/kallewoof/mff (with tools)
Patched Bitcoin Core:
https://github.com/kallewoof/bitcoin/tree/libcq
Mempool dumps available upon request.

Karl-Johan Alm

@kallewoof

C42A FF7C 61B3 E44A 1454 CD35 57AF 762D B335 3322