An analysis of dust in UTXO based cryptocurrencies

Cristina Pérez-Solà, Sergi Delgado-Segura, Guillermo Navarro-Arribas, Jordi Herrera-Joancomartí

Departament d’Enginyeria de la Informació i les Comunicacions
Universitat Autònoma de Barcelona

October 6th, 2018
1 Introduction

2 Definitions

3 Results

4 Conclusions
A **UTXO** is a transaction output that has not been spent yet.

When we talk about bitcoins we are actually referring to UTXOs.

The **UTXO set** is where all UTXOs are stored. We can see it as a wallet that includes all unspent bitcoins. No matter their type, *owner* nor value.
Properties of the UTXO set

- It is part of every full node.
- The Bitcoin value of a UTXO does not affect its size (bigger value ≠ bigger size).
- In general, the larger the output script of a UTXO, the more space it occupies in the set.
Goals

- How many unspent outputs are actually worth spending?
- How much space is every full not devoting to store not-worth-spending outputs?
Introduction

Definitions
- Dust
- Unprofitable

Results

Conclusions
How do we know if an output is worth spending?
Outputs worth spending

It depends on two factors:

- How much data such output contributes to a new transaction
- What is the fee rate we need (or want) to pay
Bitcoin Core defines **dust** as an output that costs more in fees to spend than the value of the output.

To compute the cost of spending an output, both its size and the size of the input are considered.

\[
\text{is\_dust(out)} = \begin{cases} 
1, & \text{out}_v < f \times (41 + 107/\alpha + \text{outs}) \\
0, & \text{otherwise}
\end{cases}
\]

where \(\alpha\) is 1 for non-segwit outputs and 4 otherwise.
Our definition: unprofitable outputs

**Unprofitable**

We define an **unprofitable** output as the output of a transaction that holds less value than the fee necessary to be spent, taking into account **only the size of the input** that will be needed to spend it.

\[
\text{is\_unprofitable}(\text{out}) = \begin{cases} 
1, & \text{out}_v < f \times \text{pred\_in}_s / \alpha \\
0, & \text{otherwise}
\end{cases}
\]

where:

\(\text{pred\_in}_s\) is the predicted size of the input that will spend output \(\text{out}\).

\(\alpha\) is 1 for non-segwit outputs and 4 otherwise.
Our definition: unprofitable outputs

...but how do we know the size of an input before we see it?

![Generic transaction structure diagram](image-url)
Our definition: computing the minimum size

\[ \text{min\_size} = \text{fixed\_size} + \text{variable\_size} \]

\[ \text{fixed\_size} = \text{outpoint} + n\text{Sequence} = 40\text{bytes} \]

\[ \text{variable\_size} = \text{scriptSig\_len} + \text{scriptSig} \]

depends on the UTXO type
Our definition: unprofitable outputs

Two different metrics for unprofitability:

- A **lower bound** on unprofitability, that will take into account the minimum size of the input;
- An **estimation** of unprofitability, that tries to estimate the real unprofitable rates taking into account data available in the blockchain.
Variable size: non-SegWit

Pay-to-PubKey (P2PK) outputs:

PUSH sig (1 byte) + sig (71 bytes)

Pay-to-PubkeyHash (P2PKH) outputs:

PUSH sig (1) + sig (71) + PUSH pk (1) + pk (33-65)

Pay-to-multisig (P2MS) outputs:

OP_0 (1) + (PUSH sig (1) + sig (71)) * req_sigs (1-20)

Pay-to-ScriptHash (P2SH) outputs:

∅
Variable size: SegWit

Pay-to-Witness-Public-Key-Hash (P2WPKH) outputs:
PUSH sig (1) + sig (72) + PUSH pk (1) + pk (33)

Pay-to-Witness-Script-Hash (P2WSH) outputs:
∅

Witness scripts discounted $\alpha = 1/4$
Public key sizes in the Bitcoin blockchain

Average PK size per block (P2PKH outputs)
## P2SH redeem scripts in the Bitcoin blockchain

<table>
<thead>
<tr>
<th>Redeem script</th>
<th>Number of inputs</th>
<th>Average input size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multisig</td>
<td>80,839,329</td>
<td>241.6</td>
</tr>
<tr>
<td>P2WPKH</td>
<td>7,961,073</td>
<td>23</td>
</tr>
<tr>
<td>P2WSH</td>
<td>5,544,793</td>
<td>35</td>
</tr>
<tr>
<td>Nonstd</td>
<td>112,354</td>
<td>169.98</td>
</tr>
<tr>
<td>P2PK</td>
<td>23,557</td>
<td>108.01</td>
</tr>
<tr>
<td>P2PKH</td>
<td>448</td>
<td>132</td>
</tr>
<tr>
<td>P2SH (Hash puzzle)</td>
<td>82</td>
<td>28.73</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>94,481,636</strong></td>
<td><strong>210.93</strong></td>
</tr>
</tbody>
</table>
P2SH sizes in the Bitcoin blockchain

![Graph showing average P2SH input script size over input height]
Dust UTXOs

![Graph showing the portion of UTXOs vs. fee rate (sat/byte).]

- Dust
- Non-profitable min.
- Non-profitable est.
Dust value

The graph shows the percentage of coins in circulation against the fee rate (sat/byte) for different scenarios:

- **Dust**
- **Non-profitable min.**
- **Non-profitable est.**

The x-axis represents the fee rate, while the y-axis shows the percentage of coins in circulation.
Unprofitability evolution (Bitcoin)

The graph illustrates the evolution of unprofitability for different fee rates (sat/byte) against the portion of UTXOs. The x-axis represents the fee rate in sat/byte, while the y-axis shows the portion of UTXOs. Different lines represent varying thresholds: 100K, 150K, 200K, 250K, 300K, 350K, 400K, 450K, and 500K. As the fee rate increases, the portion of UTXOs that are unprofitable also increases, indicating a higher risk of transactions not being included in blocks due to insufficient fees.
Is this really that bad?
Unprofitability evolution (Litecoin)

Fee rate (lit/byte)
Portion of UTXOs
Conclusions

- There is a fairly big percentage of dust in the UTXO set.
- The current implementation of the UTXO set can grow unbounded.
- The bigger the set gets, the less suitable it is to run a full node in low resource devices.
- Dust attacks can be performed to make the set grow.
Solutions?

- There has been proposals to mitigate this (TXO commitments by Peter Todd)
- Output consolidation when fees are low
- A good coin selection algorithm is important, specially for exchanges
An analysis of dust in UTXO based cryptocurrencies

Cristina Pérez-Solà, Sergi Delgado-Segura, Guillermo Navarro-Arribas, Jordi Herrera-Joancomartí

Departament d’Enginyeria de la Informació i les Comunicacions
Universitat Autònoma de Barcelona

October 6th, 2018