

Cardano on-the-wire specification

Nicholas Clarke

February 28, 2023

Contents

1	Introduction	1
1.1	Notation	1
1.2	Relation to the abstract specification and to the Cardano implementation	2
2	Core datatypes	2
2.1	Primitives	3
2.2	Basic Cardano Types	3
2.3	Addresses	4
2.4	Transactions	4
2.5	Shared Seed Computation	5
2.6	Delegation	6
2.7	Updates	7
2.8	Blocks	9
2.8.1	Epoch Boundary Blocks	10
3	Requirements on the binary format	10
3.1	Cryptographic properties	10
3.2	Dependencies on the binary format	10
4	Binary specification	10
A	Current binary format	10

1 Introduction

This document describes the binary serialisation formats used for the Cardano blockchain.

It proceeds as follows: section 2 describes the core types which must be serialised for the purposes of Cardano communication. We then detail the requirements on the binary format in 3. In section 4 we describe the explicit binary serialisation of the core types satisfying these requirements.

Appendix A details the current binary format.

1.1 Notation

Lists Lists of type A are written as $[A]$.

Non-empty lists Non-empty lists of type A are written as $[A]^+$.

Sets Sets of type A are written as $\{A\}$.

Maps A map from items of type A to items of type B is written as $A \rightarrow B$.¹

Records We denote records as:

$$\langle record \rangle = \{ \begin{array}{l} , \text{ fieldName} :: \text{type} \\ , \text{ fieldName2} :: \text{type2} \end{array} \}$$

Variants We denote variants as:

$$\langle variant \rangle = \begin{array}{l} \text{Option1} \\ | \text{Option2} \\ | \text{Option3.} \end{array}$$

Optional values We denote an optional value of type A as $()|A$

We will use capitalised words for variant constructors and lower-case terms for field names, as per standard Haskell syntax.

1.2 Relation to the abstract specification and to the Cardano implementation

This document straddles the middle-ground between the rule-based specification given in [1] and the concrete implementation at [2]. As such, it tries to compromise between the differing presentations of how the blockchain is constructed.

In general, we take the following approach:

- Where possible, we take our nomenclature from [1]. So, for example, we refer to the type of transactions as \mathcal{T} rather than \mathbf{Tx} . Where there is no name given in [1] we generally invent a new identifier in the similar style.
- Structure, on the other hand, typically derives from the code in [2], with some exceptions:
 - We drop `newtype` wrappers.
 - We flatten nested records where doing so would not lead to significant confusion (for example, by requiring convoluted or length names to contextualise fields).
 - In general, we tend to be fairly free with translating between isomorphic representations where it increases the readability of this spec.
 - We monomorphize any polymorphic variables to the type they are instantiated to when they are serialised.
 - We ignore `AsBinary` and its ilk. These provide ‘typed’ wrappers around `ByteString` in order to defer encoding/decoding. They only confuse an abstract definition.

2 Core datatypes

The types in this section are derived from a combination of the blockchain spec and Haskell datatypes in the codebase.

¹We use this terminology for consistency with [1], where these are partial functions.

2.1 Primitives

Primitive types:

\mathbb{B} Boolean value

Word8 8-bit word

Word32 32-bit word

Word64 64-bit word

String Arbitrary UTF8-encoded string type. We do not draw any distinctions as to whether this is implemented as a **String** or **Text**.

Bytes Arbitrary string of bytes.

\mathbb{I} arbitrary precision integers.

2.2 Basic Cardano Types

We start with sets of identifiers. These are represented in code as Blake2b-256 hashes. We also sometimes deal with these hashes directly.

- Transaction identifiers $\mathcal{I}_{\mathcal{T}}$
- Block identifiers $\mathcal{I}_{\mathcal{B}}$
- Update identifiers $\mathcal{I}_{\mathcal{UP}}$
- Concrete hashes **Hash**

We also have addresses, which are represented as Blake2b-224 hashes.

- Address (agent) identifiers $\mathcal{I}_{\mathcal{A}}$
- Stakeholder identifiers $\mathcal{I}_{\mathcal{K}}$

Epochs are identified by their index as a **Word64**. Slots are identified as a pair of epoch identifier and a **Word16** index within the slot.

- Epoch indices $\mathcal{I}_{\mathcal{E}}$
- Slot indices $\mathcal{I}_{\mathcal{S}}$

Cardano public keys are elliptic curve DH keys on curve 25519.

- Public keys χ_{pub}
- Signatures χ_{sig}

2.3 Addresses

We have a set of addresses \mathcal{A} . An address has an identifier, a type, and some additional data held in an `attributes` field. In particular, each address contains a stake distribution $\mathcal{A}_{\text{distr}}$.

$\langle \mathcal{A}_{\text{distr}} \rangle$	= Bootstrap SingleKey $\mathcal{I}_{\mathcal{A}}$ MultiKey $\mathcal{I}_{\mathcal{A}} \rightarrow \mathbf{Word64}$.
$\langle \mathcal{A}_{\text{type}} \rangle$	= PubKey Script Redeem Unknown Word8 .
$\langle \mathcal{A}_{\text{attr}} \rangle$	= { , pkDerivationPath :: () Bytes , stakeDistr :: $\mathcal{A}_{\text{distr}}$, unparsed :: Word8 \rightarrow Bytes }.
$\langle \mathcal{A} \rangle$	= { , root :: $\mathcal{I}_{\mathcal{A}}$, attributes :: $\mathcal{A}_{\text{attr}}$, type :: $\mathcal{A}_{\text{type}}$ }.

Figure 1: Address Types

2.4 Transactions

We have a set of transactions \mathcal{T} .

$\langle \mathcal{T}_{\text{in}} \rangle$	= Valid $\mathcal{I}_{\mathcal{T}} \times \mathbf{Word32}$. Invalid Word8 \times Bytes.
$\langle \mathcal{T} \rangle$	= { , inputs :: $[\mathcal{T}_{\text{in}}]$, outputs :: $[\mathcal{I}_{\mathcal{A}} \times \mathbf{Word64}]^+$ }
$\langle \mathcal{T}_{\text{proof}} \rangle$	= { , txCount :: Word32 , merkleRoot :: Hash , witnessHash :: Hash }

Figure 2: Transactions

We also have transaction witnesses, which provide proof that the transaction has authority to spend its inputs.

$\langle \mathcal{TW}_{pk} \rangle$	= { , key :: χ_{pub} , signature :: χ_{sig} }
$\langle \mathcal{TW}_{script} \rangle$	= { , validator :: Word16 × Bytes , redeemer :: Word16 × Bytes }
$\langle \mathcal{TW}_{redeem} \rangle$	= { , key :: χ_{pub} , signature :: χ_{sig} }
$\langle \mathcal{TW} \rangle$	= PublicKey \mathcal{TW}_{pk} Script \mathcal{TW}_{script} Redeem \mathcal{TW}_{redeem} Unknown Word8 Bytes

Figure 3: Transaction witnesses

2.5 Shared Seed Computation

Shared seed computation deals with commitments, openings, shares and certificates.

We start with some basic types used for VSS. VSS uses its own public key cryptography scheme.

- VSS public keys $\mathcal{VSS}_{\chi_{pub}}$
- VSS secret \mathcal{VSS}_{sec}
- VSS encrypted and decrypted shares \mathcal{VSS}_{enc} and \mathcal{VSS}_{dec}
- VSS secret proof \mathcal{VSS}_{proof}

$$\begin{aligned}
\langle SSC_{comm} \rangle &= \chi_{pub} \times ((VSS_{\chi_{pub}} \rightarrow VSS_{enc}) \times VSS_{proof}) \times \chi_{sig} \\
\langle SSC_{open} \rangle &= \mathcal{I}_{\mathcal{K}} \rightarrow VSS_{sec} \\
\langle SSC_{share} \rangle &= \mathcal{I}_{\mathcal{K}} \rightarrow (\mathcal{I}_{\mathcal{K}} \rightarrow [VSS_{dec}]^+) \\
\langle SSC_{cert} \rangle &= \{ \\
&\quad , \text{vsskey} :: VSS_{\chi_{pub}} \\
&\quad , \text{signingKey} :: \chi_{pub} \\
&\quad , \text{expiry} :: \mathcal{I}_{\mathcal{E}} \\
&\quad , \text{signature} :: \chi_{sig} \} \\
\langle SSC \rangle &= \text{Commitments } [SSC_{comm}] \times [SSC_{cert}] \\
&\quad | \text{Openings } SSC_{open} \times [SSC_{cert}] \\
&\quad | \text{Shares } SSC_{share} \times [SSC_{cert}] \\
&\quad | \text{Certificates } [SSC_{cert}] \\
\langle SSC_{proof} \rangle &= \text{CommitmentsProof } \mathbf{Hash} \times \mathbf{Hash} \\
&\quad | \text{OpeningsProof } \mathbf{Hash} \times \mathbf{Hash} \\
&\quad | \text{SharesProof } \mathbf{Hash} \times \mathbf{Hash} \\
&\quad | \text{CertificatesProof } \mathbf{Hash}
\end{aligned}$$

Figure 4: Shared seed computation

2.6 Delegation

The heavyweight delegation system is used to allow stakeholders to authorise other parties to issue blocks on their behalf.

$$\begin{aligned}
\langle \mathcal{DLG} \rangle &= \{ \\
&\quad , \text{epoch} :: \mathcal{I}_{\mathcal{E}} \\
&\quad , \text{issuer} :: \chi_{\mathbf{pub}} \\
&\quad , \text{delegate} :: \chi_{\mathbf{pub}} \\
&\quad , \text{certificate} :: \chi_{\mathbf{sig}} \} \\
\langle \mathcal{DLG}_{sig} \rangle &= \{ \\
&\quad , \text{psk} :: \mathcal{DLG} \\
&\quad , \text{sig} :: \chi_{\mathbf{sig}} \} \\
\langle \mathcal{DLG}^l \rangle &= \{ \\
&\quad , \text{epoch} :: \mathcal{I}_{\mathcal{E}} \times \mathcal{I}_{\mathcal{E}} \\
&\quad , \text{issuer} :: \chi_{\mathbf{pub}} \\
&\quad , \text{delegate} :: \chi_{\mathbf{pub}} \\
&\quad , \text{certificate} :: \chi_{\mathbf{sig}} \} \\
\langle \mathcal{DLG}_{sig}^l \rangle &= \{ \\
&\quad , \text{psk} :: \mathcal{DLG}^l \\
&\quad , \text{sig} :: \chi_{\mathbf{sig}} \}
\end{aligned}$$

Figure 5: Delegation

2.7 Updates

This section covers the types used to orchestrate updates of the system. We also introduce the block version and transaction fee policy in this section, since they are only serialised as part of the update system.

$\langle \mathcal{B}_{ver} \rangle$	= Word16 × Word16 × Word8
$\langle \mathcal{T}_{feepol} \rangle$	= Linear \mathbb{I} \mathbb{I} Unknown Word8 Bytes
$\langle \mathcal{B}_{ver\Delta} \rangle$	= { , scriptVersion :: () Word16 , slotDuration :: () \mathbb{I} , maxBlockSize :: () \mathbb{I} , maxHeaderSize :: () \mathbb{I} , maxTxSize :: () \mathbb{I} , maxProposalSize :: () \mathbb{I} , mpcThd :: () Word64 , heavyDelThd :: () Word64 , updateVoteThd :: () Word64 , updateProposalThd :: () Word64 , updateImplicit :: () Word64 , softforkRule :: () (Word64 × Word64 × Word64) , txFeePolicy :: () \mathcal{T}_{feepol} , unlockStakeEpoch :: () $\mathcal{I}_{\mathcal{E}}$ }
$\langle \mathcal{UP}_{data} \rangle$	= { , appDiffHash :: Hash , pkgHash :: Hash , updaterHash :: Hash , mdHash :: Hash }
$\langle \mathcal{UP}_{prop} \rangle$	= { , blockVersion :: \mathcal{B}_{ver} , blockVersionMod :: $\mathcal{B}_{ver\Delta}$, softwareVersion :: String × Word32 , data :: String → \mathcal{UP}_{data} , attributes :: Word8 → Bytes , from :: χ_{pub} , signature :: χ_{sig} }
$\langle \mathcal{UP}_{vote} \rangle$	= { , voter :: χ_{pub} , proposalId :: $\mathcal{I}_{\mathcal{UP}}$, vote :: \mathbb{B} , signature :: χ_{sig} }
$\langle \mathcal{UP} \rangle$	= { , proposal :: () \mathcal{UP}_{prop} , votes :: [\mathcal{UP}_{vote}] }

Figure 6: Updates

2.8 Blocks

A block consists of a block header and a block body. The block header consists of verification for the various components in the block body.

$\langle \mathcal{B}_{sig} \rangle$	= Signature χ_{sig} ProxySigLight \mathcal{DLG}_{sig}^1 ProxySigHeavy \mathcal{DLG}_{sig}
$\langle \mathcal{B}_{cons} \rangle$	= { , slotId :: \mathcal{I}_S , leaderKey :: χ_{pub} , difficulty :: Word64 , signature :: \mathcal{B}_{sig} }
$\langle \mathcal{B}_{headex} \rangle$	= { , blockVersion :: \mathcal{B}_{ver} , softwareVersion :: String \times Word32 , extraProof :: Hash , attributes :: String $\rightarrow \emptyset$ }
$\langle \mathcal{B}_{proof} \rangle$	= { , txProof :: \mathcal{T}_{proof} , sscProof :: \mathcal{SSC}_{proof} , dlGProof :: Hash , updProof :: Hash }
$\langle \mathcal{B}_{head} \rangle$	= { , prevBlock :: \mathcal{I}_B , bodyProof :: \mathcal{B}_{proof} , consensusData :: \mathcal{B}_{cons} , extraData :: \mathcal{B}_{headex} }
$\langle \mathcal{B}_{body} \rangle$	= { , txPayload :: $[\mathcal{T} \times [\mathcal{TW}]]$, sscPayload :: \mathcal{SSC} , dlGPayload :: $[\mathcal{DLG}]$, updPayload :: \mathcal{UP} }
$\langle \mathcal{B} \rangle$	= { , header :: \mathcal{B}_{head} , body :: \mathcal{B}_{body} , extra :: Word8 \rightarrow Bytes }

Figure 7: Blocks

2.8.1 Epoch Boundary Blocks

In addition to regular blocks, epoch boundary blocks contain a list of slot leaders for a given epoch. They are not conventionally distributed as part of the blockchain, but can be requested as part of catch-up and as such form part of the on-the-wire protocol.

$\langle \mathcal{B}'_{cons} \rangle$	= { , epoch :: $\mathcal{I}_{\mathcal{E}}$, chainDifficulty :: Word64 }
$\langle \mathcal{B}'_{head} \rangle$	= { , prevBlock :: $\mathcal{I}_{\mathcal{B}}$, bodyProof :: Hash , consensusData :: \mathcal{B}'_{cons} , extraData :: Word8 \rightarrow Bytes }
$\langle \mathcal{B}' \rangle$	= { , header :: \mathcal{B}'_{head} , body :: $[\mathcal{I}_{\mathcal{X}}]^+$, extra :: Word8 \rightarrow Bytes }

Figure 8: Epoch Boundary Blocks

3 Requirements on the binary format

3.1 Cryptographic properties

3.2 Dependencies on the binary format

4 Binary specification

References

- [1] *Rule-based specification of the blockchain logic*. Erik de Castro Lopo, Nicholas Clarke & Arnaud Spiwack.
- [2] *Cryptographic currency implementing Ouroboros PoS protocol*. <https://github.com/input-output-hk/cardano-sl/>
- [3] *RFC 7049 Concise Binary Object Representation*. <http://cbor.io>
- [4] *Concise data definition language (CDDL): a notational convention to express CBOR data structures* <https://tools.ietf.org/html/draft-ietf-cbor-cddl-00>

A Current binary format

This section documents the current binary serialisation format, as of 2018-06-03.

The current blockchain is serialized using CBOR[3]. Consequently we present the current description as a CDDL[4] document. Terms in use below should be interpreted in that context

and with reference to Appendix E of [4], which defines a standard prelude available to such things.

```
; Cardano Byron blockchain CBOR schema
```

```
block = [0, eblock]
        / [1, mainblock]
```

```
mainblock = [ "header" : blockhead
              , "body" : blockbody
              , "extra" : [attributes]
              ]
```

```
eblock = [ "header" : ebbhead
           , "body" : [+ stakeholderid]
           , extra : [attributes]
           ]
```

```
u8 = uint .lt 256
u16 = uint .lt 65536
u32 = uint
u64 = uint
```

```
; Basic Cardano Types
```

```
blake2b-256 = bytes .size 32
```

```
txid = blake2b-256
blockid = blake2b-256
updid = blake2b-256
hash = blake2b-256
```

```
blake2b-224 = bytes .size 28
```

```
addressid = blake2b-224
stakeholderid = blake2b-224
```

```
epochid = u64
slotid = [ epoch: epochid, slot : u64 ]
```

```
pubkey = bytes
signature = bytes
```

```
; Attributes - at the moment we do not bother deserialising these, since they
; don't contain anything
```

```
attributes = { * any => any }
```

```
; Addresses
```

```

addrdistr = [1] / [0, stakeholderid]

addrtype = &("PubKey" : 0, "Script" : 1, "Redeem" : 2) / (u64 .gt 2)
addrattr = { ? 0 : addrdistr
             , ? 1 : bytes}
address = [ #6.24(bytes .cbor ([addressid, addrattr, addrtype])), u64 ]

; Transactions

txin = [0, #6.24(bytes .cbor ([txid, u32]))] / [u8 .ne 0, encoded-cbor]
txout = [address, u64]

tx = [[+ txin], [+ txout], attributes]

txproof = [u32, hash, hash]

twit = [0, #6.24(bytes .cbor ([pubkey, signature]))]
      / [1, #6.24(bytes .cbor ([[u16, bytes], [u16, bytes]]))]
      / [2, #6.24(bytes .cbor ([pubkey, signature]))]
      / [u8 .gt 2, encoded-cbor]

; Shared Seed Computation

vsspubkey = bytes ; This is encoded using the 'Binary' instance
            ; for Scrape.PublicKey
vsssec = bytes ; This is encoded using the 'Binary' instance
        ; for Scrape.Secret.
vssenc = [bytes] ; This is encoded using the 'Binary' instance
        ; for Scrape.EncryptedSi.
        ; TODO work out why this seems to be in a length 1 array
vssdec = bytes ; This is encoded using the 'Binary' instance
        ; for Scrape.DecryptedShare
vssproof = [bytes, bytes, bytes, [* bytes]] ; This is encoded using the
        ; 'Binary' instance for Scrape.Proof

sscomm = [pubkey, [{vsspubkey => vssenc}, vssproof], signature]
sscomms = #6.258([* sscomm])

sscopens = {stakeholderid => vsssec}

sscshares = {addressid => [addressid, [* vssdec]]}

ssccert = [vsspubkey, pubkey, epochid, signature]
ssccerts = #6.258([* ssccert])

ssc = [0, sscomms, ssccerts]
      / [1, sscopens, ssccerts]
      / [2, sscshares, ssccerts]
      / [3, ssccerts]

```

```

sscproof = [0, hash, hash]
           / [1, hash, hash]
           / [2, hash, hash]
           / [3, hash]

; Delegation

dlg = [ epoch : epochid
        , issuer : pubkey
        , delegate : pubkey
        , certificate : signature
      ]

dlgsig = [dlg, signature]

lwdlg = [ epochRange : [epochid, epochid]
        , issuer : pubkey
        , delegate : pubkey
        , certificate : signature
      ]

lwdlgsig = [lwdlg, signature]

; Updates

bver = [u16, u16, u8]

txfeepol = [0, #6.24(bytes .cbor ([bigint, bigint]))]
           / [u8 .gt 0, encoded-cbor]

bvermod = [ scriptVersion : [? u16]
          , slotDuration : [? bigint]
          , maxBlockSize : [? bigint]
          , maxHeaderSize : [? bigint]
          , maxTxSize : [? bigint]
          , maxProposalSize : [? bigint]
          , mpcThd : [? u64]
          , heavyDelThd : [? u64]
          , updateVoteThd : [? u64]
          , updateProposalThd : [? u64]
          , updateImplicit : [? u64]
          , softForkRule : [? [u64, u64, u64]]
          , txFeePolicy : [? txfeepol]
          , unlockStakeEpoch : [? epochid]
        ]

updata = [ hash, hash, hash, hash ]

```

```

upprop = [ "blockVersion" : bver
           , "blockVersionMod" : bvermod
           , "softwareVersion" : [ text , u32 ]
           , "data" : #6.258([text , updata])
           , "attributes" : attributes
           , "from" : pubkey
           , "signature" : signature
           ]

upvote = [ "voter" : pubkey
           , "proposalId" : updid
           , "vote" : bool
           , "signature" : signature
           ]

up = [ "proposal" : [? upprop]
       , votes : [* upvote]
       ]

; Blocks

difficulty = [u64]

blocksig = [0, signature]
           / [1, lwdlgsig]
           / [2, dlgsig]

blockcons = [slotid , pubkey , difficulty , blocksig]

blockheadex = [ "blockVersion" : bver
               , "softwareVersion" : [ text , u32 ]
               , "attributes" : attributes
               , "extraProof" : hash
               ]

blockproof = [ "txProof" : txproof
              , "sscProof" : sscproof
              , "dlgProof" : hash
              , "updProof" : hash
              ]

blockhead = [ "protocolMagic" : u32
             , "prevBlock" : blockid
             , "bodyProof" : blockproof
             , "consensusData" : blockcons
             , "extraData" : blockheadex
             ]

blockbody = [ "txPayload" : [* [tx , [* twit ]]]

```

```
    , "sscPayload" : ssc
    , "dlgPayload" : [* dlg]
    , "updPayload" : up
  ]
```

; Epoch Boundary Blocks

```
ebbcons = [ epochid , difficulty ]
```

```
ebbhead = [ "protocolMagic" : u32
            , "prevBlock" : blockid
            , "bodyProof" : hash
            , "consensusData" : ebbcons
            , "extraData" : [attributes]
            ]
```