Introduzione alle Blockchain

P. Rullo

Corso di Basi di Dati

Corso di Laurea in Informatica

Unical

Introduzione: Blockchain e criptovalute

- La Blockchain (BC) nasce come la tecnologia sottostante il Bitcoin
- Come tale, gli usi primari della BC comprendono i pagamenti e altre transazioni finanziarie
- Il Bitcoin è una criptovaluta: valuta ‘nascosta', nel senso che è visibile/utilizzabile solo conoscendo le ‘chiavi di accesso' pubblica e privata
- La criptovaluta non esiste in forma fisica (anche per questo viene definita ‘virtuale'), ma si genera e si scambia esclusivamente per via telematica. Non è pertanto possibile trovare in circolazione dei bitcoin in formato cartaceo o metallico

Introduzione: Blockcahin e criptovalute

- La Blockchain supporta transazioni peer-to-peer, eliminando la necessità di un intermediario di fiducia che verifica le operazioni - un ruolo che è necessario quando i partecipanti non si conoscono o non si fidano
- Quindi non c'è nessuna banca o nessun soggetto terzo che verifica le transazioni
- È invece la rete, nel suo complesso, che verifica le transazioni attraverso un "meccanismo di consenso" decentralizzato

Definizioni preliminari - Funzione Hash

- Una funzione hash associa ad una stringa di lunghezza m arbitraria, detta *messaggio*, una stringa di lunghezza n fissa, detta *digest*
- **Esempio**. Usando il calcolatore hash (SHA-256)

https://hash.online-convert.com/it/generatore-sha256

il messaggio “Una funzione hash associa ad una stringa di lunghezza m arbitraria, il messaggio, una stringa di lunghezza n fissa, il digest.” viene convertito nel seguente digest di 32 caratteri (256 bit):

39de4f0a0344364aee639973ab032824318da561025a0f1070b8445c74ff5e32

Definizioni preliminari - Funzione Hash

- La funzione non è iniettiva – più messaggi associati allo stesso digest
- Messaggi di m bit e digest di n bit
  - $p=2^m$ possibili messaggi
  - $q=2^n$ possibili digest
- Ad esempio, con m=128 e n=32, vi sono $r=p/q=2^{96}$ messaggi per digest (assumendo una distribuzione uniforme). La probabilità che due messaggi collidano è circa pari a $1/2^{32}$
- Quello delle collisioni non è un fenomeno desiderato, in quanto il digest si usa come una impronta digitale, quindi univoca, di ogni dato documento

Definizioni preliminari - Funzione Hash

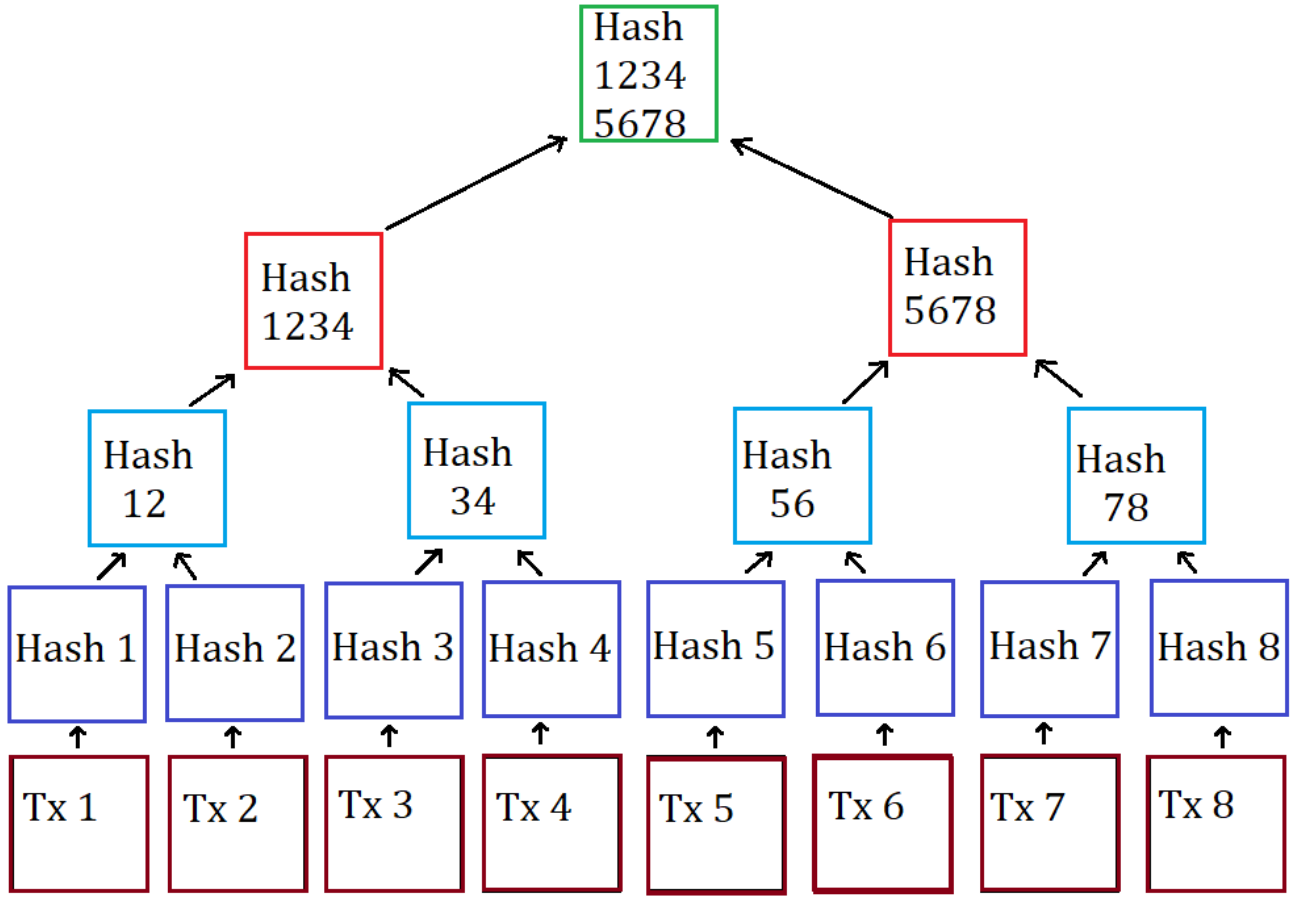
Proprietà

- Deve essere facile calcolare il digest D di un messaggio M
- Dato il digest D, deve essere difficile trovare un messaggio M che ha generato D, cioè, tale che D=hash(M) -- unidirezionalità o non invertibilità delle funzioni di hash
- Deve essere difficile modificare un messaggio senza modificare il relativo digest -- resistenza debole alle collisioni
- Deve essere difficile trovare due messaggi M e M’ che abbiano lo stesso digest D, cioè, D = hash(M) = hash(M’) -- resistenza forte alle collisioni

Definizioni preliminari - Funzione Hash

- Effetto valanga: Si consideri la seguente stringa
  - S1: “Una funzione hash associa ad una stringa di lunghezza m arbitraria, il messaggio, una stringa di lunghezza n fissa, il digest.”
- digest(S1) = 39de4f0a0344364aee639973ab032824318da561025a0f1070b8445c74ff5e32
- Modifichiamo S1 eliminando solo il simbolo finale “.”
  - S2: “Una funzione hash associa ad una stringa di lunghezza m arbitraria, il messaggio, una stringa di lunghezza n fissa, il digest”
- digest(S2) = 2a1515d3032583a1aa12416ba4765e4469740adb48e98a0f2c9476f958be1372

Definizioni preliminari - Merkle tree



- Tx_1, …, Tx_n sono stringhe che rappresentano transazioni
- Si calcola $h_1 = hash(Tx_1)$, …, $h_n = hash(Tx_n)$, che rappresentano le foglie del MT
- si calcola quindi $h_{1,2} = hash(h_1,h_2)$, …, $h_{n-1,n} = hash(h_{n-1}, h_n)$, che rappresentano i nodi padri deli nodi foglia
- il processo viene reiterato fino a quando non si genera un unico nodo, il Merkle root.

La Blockchain – struttura dati

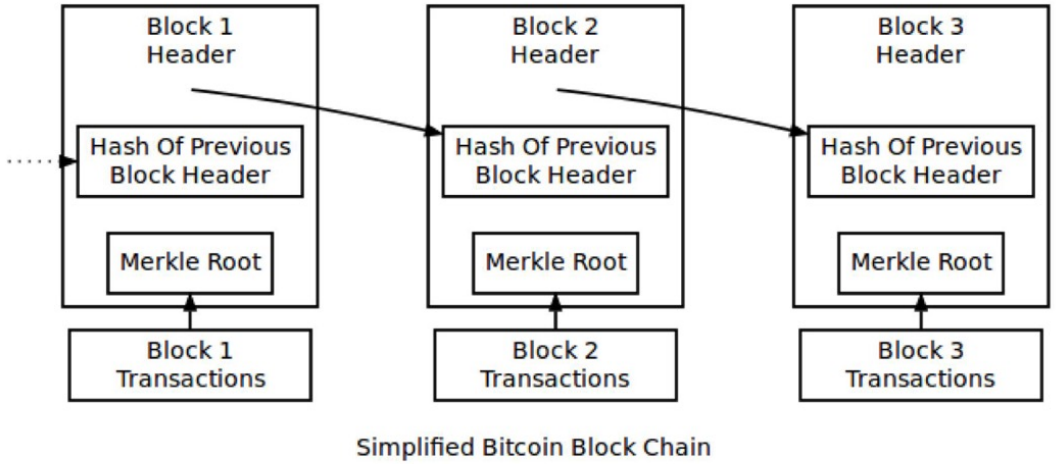
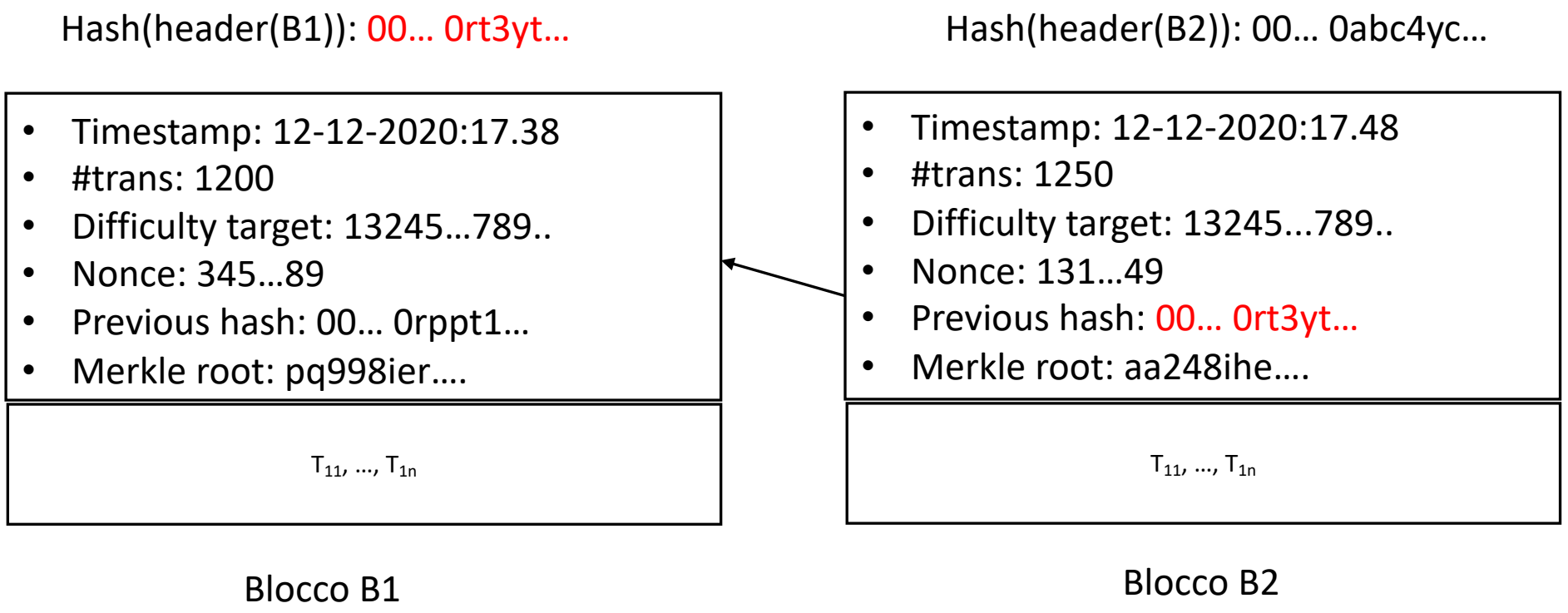


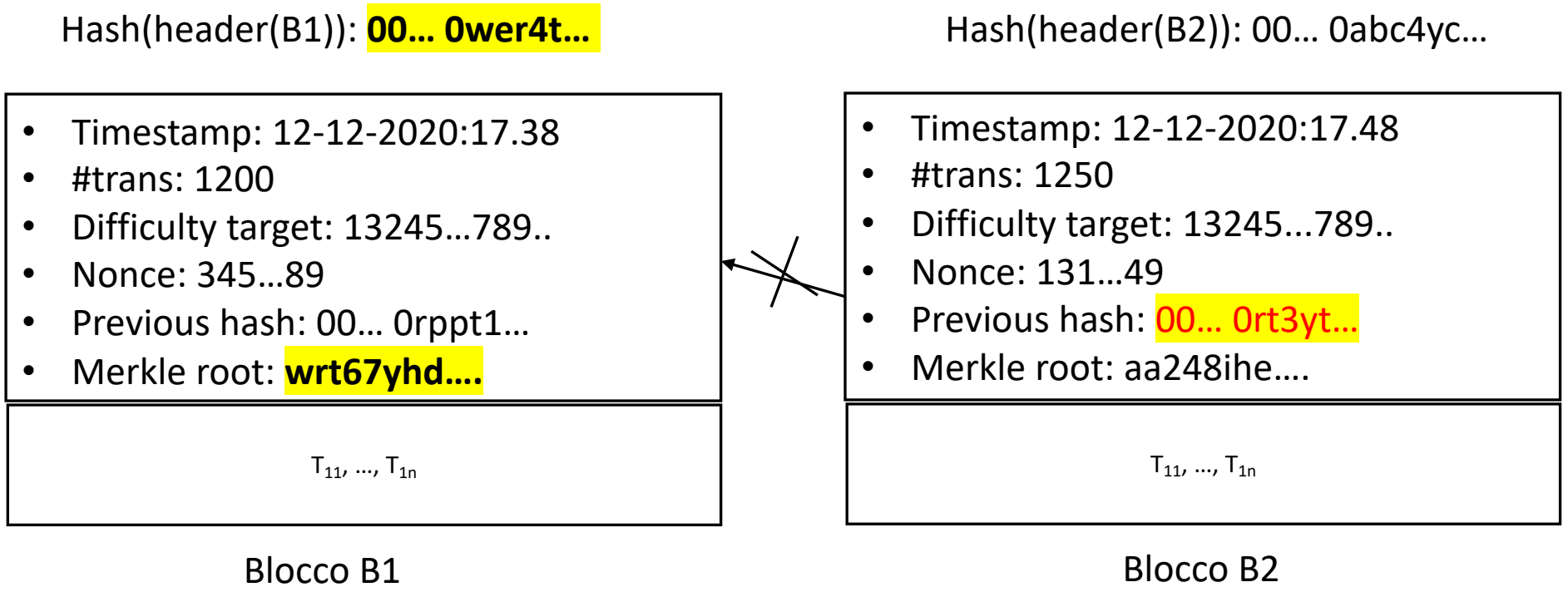
Fig. 6. Schematizzazione semplificata della Blockchain dei Bitcoin.

- E’ una catena di blocchi
- Ogni blocco è formato da
  - un header
  - un blocco di transazioni
- Header
  - Data e ora (timestamp)
  - Nr. di transazioni contenute nel blocco
  - Difficulty target
  - Nonce
  - Hash (digest) del block header precedente
  - Merkle root del blocco delle transazioni

L’integrità della catena



L’integrità della catena



- Viene modificata qualche transazione in B1
- Cambia il Markle root
- Cambia l’header
- Cambia hash(header)

- Il valore del Previous hash in B2 non è corretto
- Si rompe la catena!

L’integrità della catena

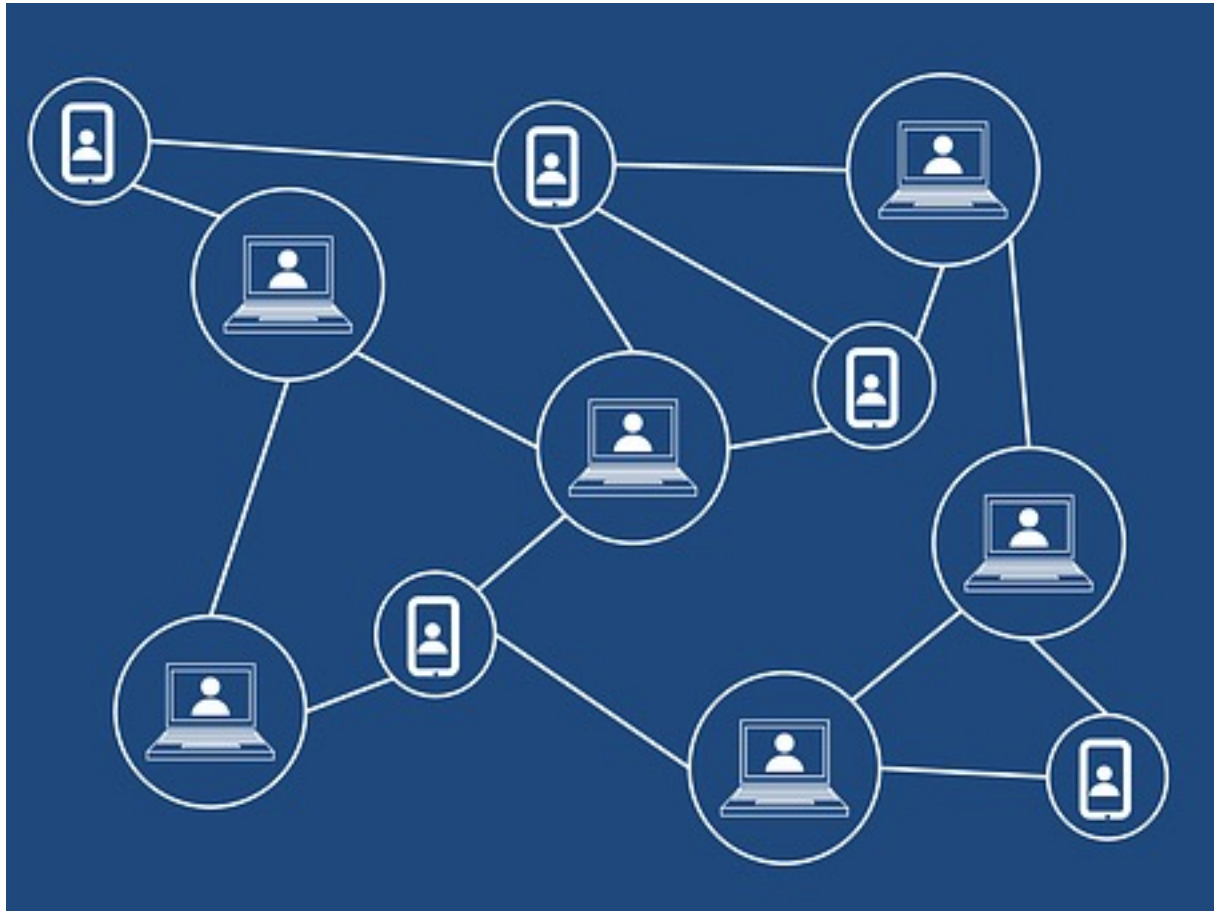
- Rimedio: al fine di ripristinare il collegamento tra B1 e B2
  - aggiorno il valore del *previous hash* di B2 al nuovo valore hash(header(B1))= 00… 0wer4t….
  - Modificando il *previous hash* di B2, cambia l’header di B2 e, quindi, il suo hash
  - si rompe il link tra B2 e il blocco B3
  - il problema si reitera per tutti i blocchi a seguire!!

L’integrità della catena

- Verificare se due BC sono uguali è facile: basta verificare se gli hash header dell’ultimo blocco sono uguali
- L’hashing rende ogni manomissione della BC evidente

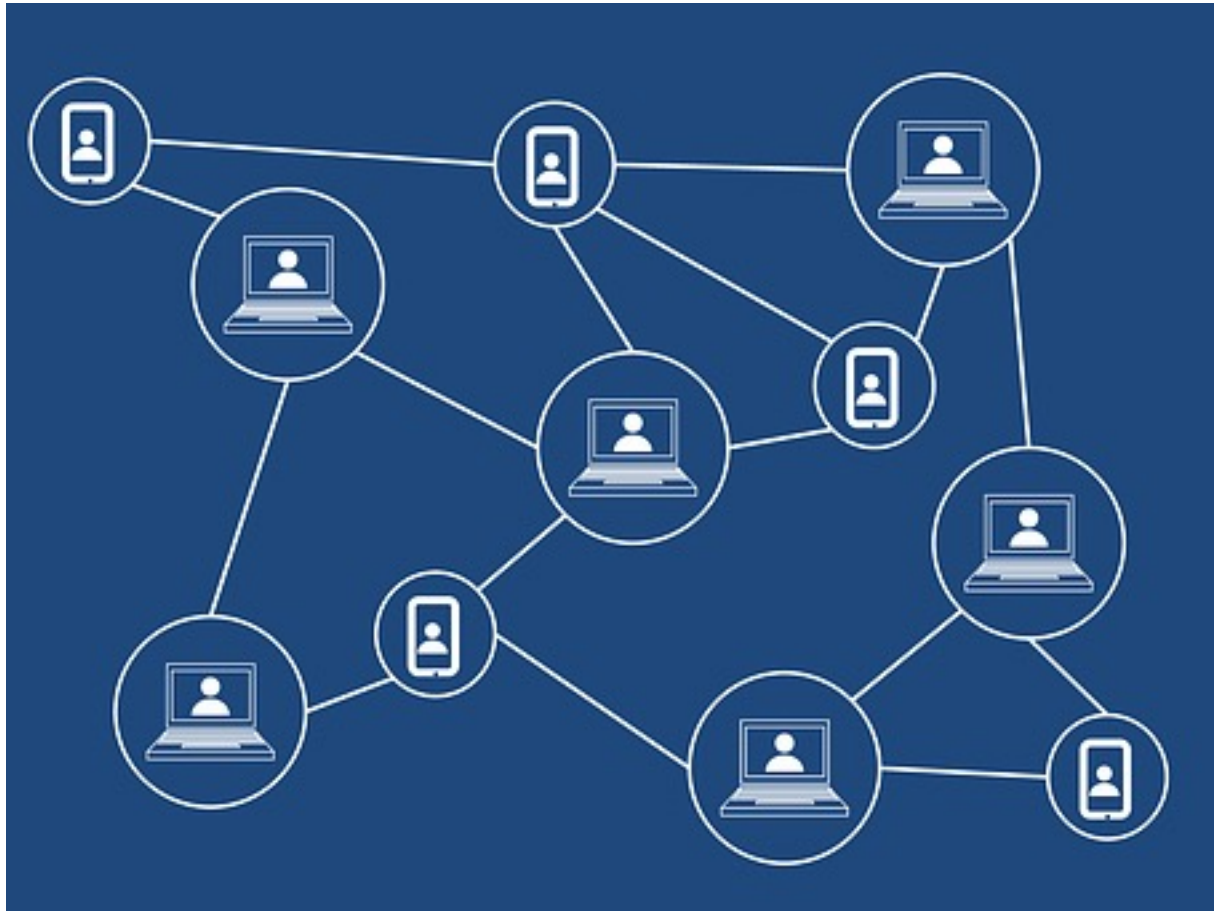
Il network della blockchain

- La blockchain è replicata su ogni nodo della rete - architettura *distribuita*
- La rete è *decentralizzata:* non esiste un server centrale e tutti i nodi sono considerati uguali – svolgono le stesse funzioni (*peer-to-peer*)
- La rete è caratterizzata dall'assenza di una autorità centrale



Il network della blockchain

- Due tipi di nodi: completi (*full node*) e leggeri (*light node*)
- Nodo completo ospita una istanza della BC
- Nodo leggero viene solo usato da utente finale per eseguire transazioni



Transazioni

- Una transazione è un record che registra il trasferimento di un asset digitale (ad esempio, Bitcoin) da un mittente ad un destinatario
- dati di una transazione
  - indirizzo del mittente (pagante)
  - indirizzo del destinatario (ricevente)
  - ammontare da trasferire
  - firma digitale del mittente
  - chiave pubblica del ricevente

Transazioni

Il pagante A vuole trasferire bitcoin al ricevente B:

- Il ricevente B comanda alla sua applicazione su PC o smartphone di creare un **indirizzo**
- Il ricevente B invia l'indirizzo al mittente A tramite qualunque mezzo: mail, messaggio, QR code, ecc.
- Il pagante A
  - inserisce nel suo software l'indirizzo di B e l’ammontare VAB da inviare
  - specifica l'ammontare della commissione da pagare al miner
  - conferma la transazione

Validazione delle transazioni

- Una volta confermata dal pagante, una transazione T viene inviata ai nodi (full node) della rete per la validazione
- Ogni nodo applica il protocollo della blockchain indipendentemente dagli altri

La filosofia di base della BC è che la validazione delle transazioni e dei blocchi debba essere fatta da tutti i partecipanti alla rete, senza una autorità centrale

Validazione delle transazioni

- Quando un nodo riceve una transazione, il protocollo prevede che vengano fatte alcune verifiche, come:
  - L’ammontare da trasferire deve essere nella disponibilità del pagante
  - Il pagante deve essere chi dice di essere
  - ….

Validazione delle transazioni

- Dopo che un nodo (full node) ha localmente verificato la transazione T, procede al suo invio agli altri nodi della rete (T viene propagata), che a loro volta validano T
- Se T non risulta essere valida, allora la sua propagazione viene bloccata
- Le transazioni che risultano valide vengono quindi propagate a tutti i nodi della rete

Consenso e Mining

- Tuttavia, prima di essere *confermata*, una transazione deve essere inserita in un nuovo blocco che sarà aggiunto ad ogni singola istanza della BC
- Ciò deve avvenire con il *consenso* di tutta la rete, in quanto la rete è un sistema peer-to-peer, democratico e anonimo
- **Problema di fiducia**: è possibile, in assenza di una autorità centrale, e in una rete in cui i nodi non danno informazioni sulla propria identità, fidarsi di tutti i partecipanti, cioè, che ogni singolo nodo applichi correttamente il protocollo di validazione?
- Risposta: NO

Consenso e Mining

- *Sybil Attack*: un miner disonesto (attaccante) potrebbe creare molti nodi anonimi (anche migliaia o milioni) apparentemente indipendenti, e prendere il controllo della rete per validare proprie transazioni non valide!
- Bisogna creare meccanismi robusti di *consenso* distribuito, che "forzano" la rete alla applicazione delle regole previste dal protocollo, al riparo da attacchi Sybil
- **Proof of Work** svolto dai **miner**

Proof of Work Mining

- I miner sono utenti della rete disposti a pagare un prezzo per partecipare al processo di validazione dei blocchi, in cambio di un ricompensa
- Quando un miner M riceve una transazione T, esegue i controlli standard e, se l'esito è positivo
  - inserisce T nella lista di transazioni *unconfirmed*, chiamata *mempool*
  - appena possibile, inserisce T in un blocco B da validare – M sceglie prima le transazioni che prevedono una ricompensa più alta
  - quando B è pronto, M inizia il processo di **mining basato sul PoW**
  - se questo ha successo, M trasmette il blocco B alla rete

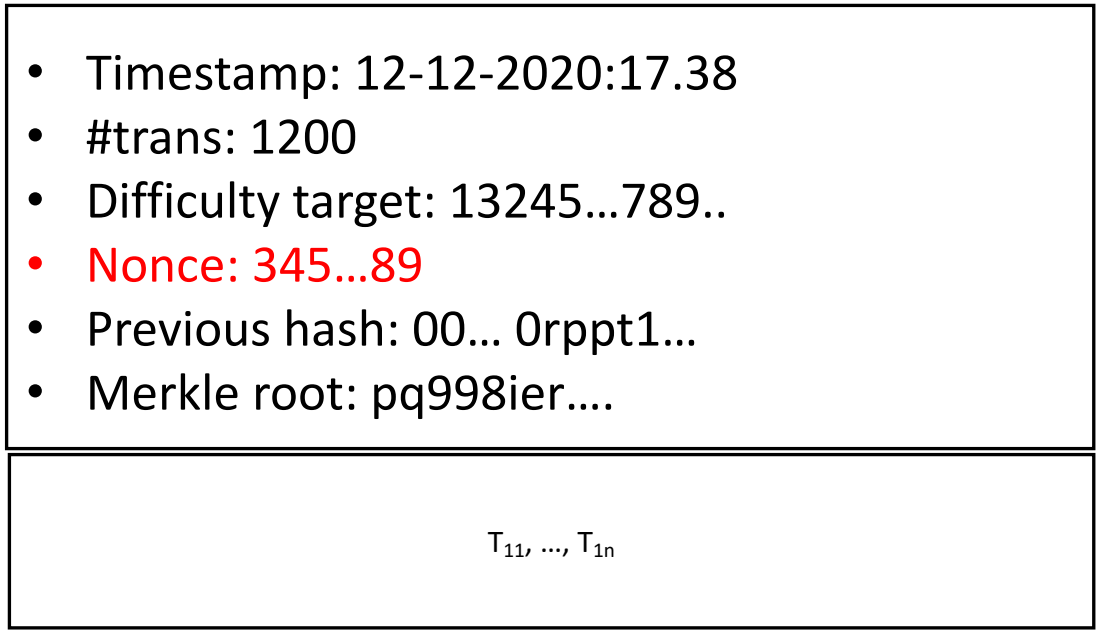
Proof of Work Mining

- Il PoW (Proof of Work) è una tecnica di mining che restringe il processo di validazione dei blocchi ai miner
- Questi devono disporre di un hardware molto veloce e costoso (in termini di consumi energetici) per la partecipazione ad una competizione che richiede la risoluzione di un problema matematico
- Il “vincitore” propone alla rete il *suo* blocco che, una volta verificato dai singoli nodi, viene accettato definitivamente

La Competizione Matematica

- **Problema matematico** del PoW: trovare un hash (SHA2-256) dell'header del blocco in fase di validazione che inizi con un numero *k* di zeri
- Problema computazionalmente difficile da risolvere per tentativi
- Ad ogni tentativo, l'argomento dell'hash (header) è variato incrementando il *nonce*

La Competizione Matematica



- Cambia il nonce per modificare l'header alla ricerca di un digest che soddisfa il difficulty target

La Competizione Matematica

Quanti tentativi sono in media necessari per trovare una soluzione?

- Supponiamo che il digest dell'header sia di n=256 bit, e che il difficulty target richieda hash che iniziano con k=32 zeri. Il numero di configurazioni binarie di n bit con k zeri iniziali è pari a 2^{n-k}. Pertanto, la probabilità che una configurazione soddisfi il target è

$$p = \frac{2^{n-k}}{2^n} = \frac{1}{2^k}$$

- Ne consegue che, per k=32, mediamente devono essere fatti 2^{32} tentativi prima di trovare un hash valido.
- Più alto è il valore di k, più basso è il valore del target da generare, maggiore è la difficoltà del problema.

La Competizione Matematica

Quanto vale k nella rete Bitcoin*?*

- Nella rete Bitcoin
  a. vengono validati 5 blocchi al minuto (throughput)
  b. l'hashrate totale è di circa 2^{70} hash/min
- Quindi, con k=70 viene validata in media una transazione/min
- Ne consegue il numero di zeri iniziali del target è k=68

La Competizione Matematica

Quanto vale k nella rete Bitcoin?

- Se aumenta l'hashrate totale, a parità di k
  - aumenta il throughput – blocchi/min
  - aumentano le ricompense ai miner
  - aumenta la generazione di bitcoin
  - diminuisce il valore del bitcoin
- Il target può essere variato per mantenere la rete in equilibrio

La Competizione Matematica

Qual è la probabilità p(M) che un miner M vinca la competizione PoW?

$$p(M) = \frac{hashrate(M)}{hashrate\ totale} = \frac{hashrate(M)}{\sum_{\forall\ miner\ X} hashrate(X)}$$

- L'hashrate totale nel 2021 è pari a circa 10^{20}hash/sec
- L'hashrate di un processore ASIC (progettato per il mining) è di circa 14 TH/s.
- Quindi se M possiede un unico ASIC

$$p(M) \approx 0{,}00000014$$

La conferma delle transazioni

- Il primo miner M che risolve il problema di ricerca dell'hash vince la competizione
- A questo punto, M invia alla rete il blocco B
- Ogni nodo che riceve B esegue le seguenti operazioni:
  - calcola l'hash dell'header di B (che contiene il nonce) e verifica che rispetti il difficulty target (numero di zeri iniziali) – questo è un calcolo veloce
  - se la verifica ha successo, il nodo valida le transazioni in B
  - se anche questa verifica ha successo, il nodo accetta il blocco e lo inserisce nella propria istanza della BC. Ogni transazione in B è a questo punto *confermata*

La conferma delle transazioni

- Se il miner M ha invece “imbrogliato” (cioè, ha inviato alla rete il blocco B come valido nonostante contenga qualche transazione non valida, oppure l’hash generato non rispetta il *difficulty target*), il blocco viene semplicemente *ignorato* dai nodi che lo hanno ricevuto
- Il miner perde la ricompensa

Perché il PoW protegge dal Sybil Attack

- Se un miner vuole avere una accettabile probabilità di successo deve dotarsi di una potenza di calcolo adeguata – processori specializzati nel calcolo del SHA256
- Mining farm
- Tuttavia, acquisire la maggioranza dei nodi costa troppo
- Il PoW, quindi, pur non escludendo la possibilità teorica di un Sybil attack, certamente ne abbassa la probabilità rendendolo praticamente inattuabile.

Perché il PoW protegge dal Sybil Attack

Quanto costa acquisire la maggioranza dei nodi*?*

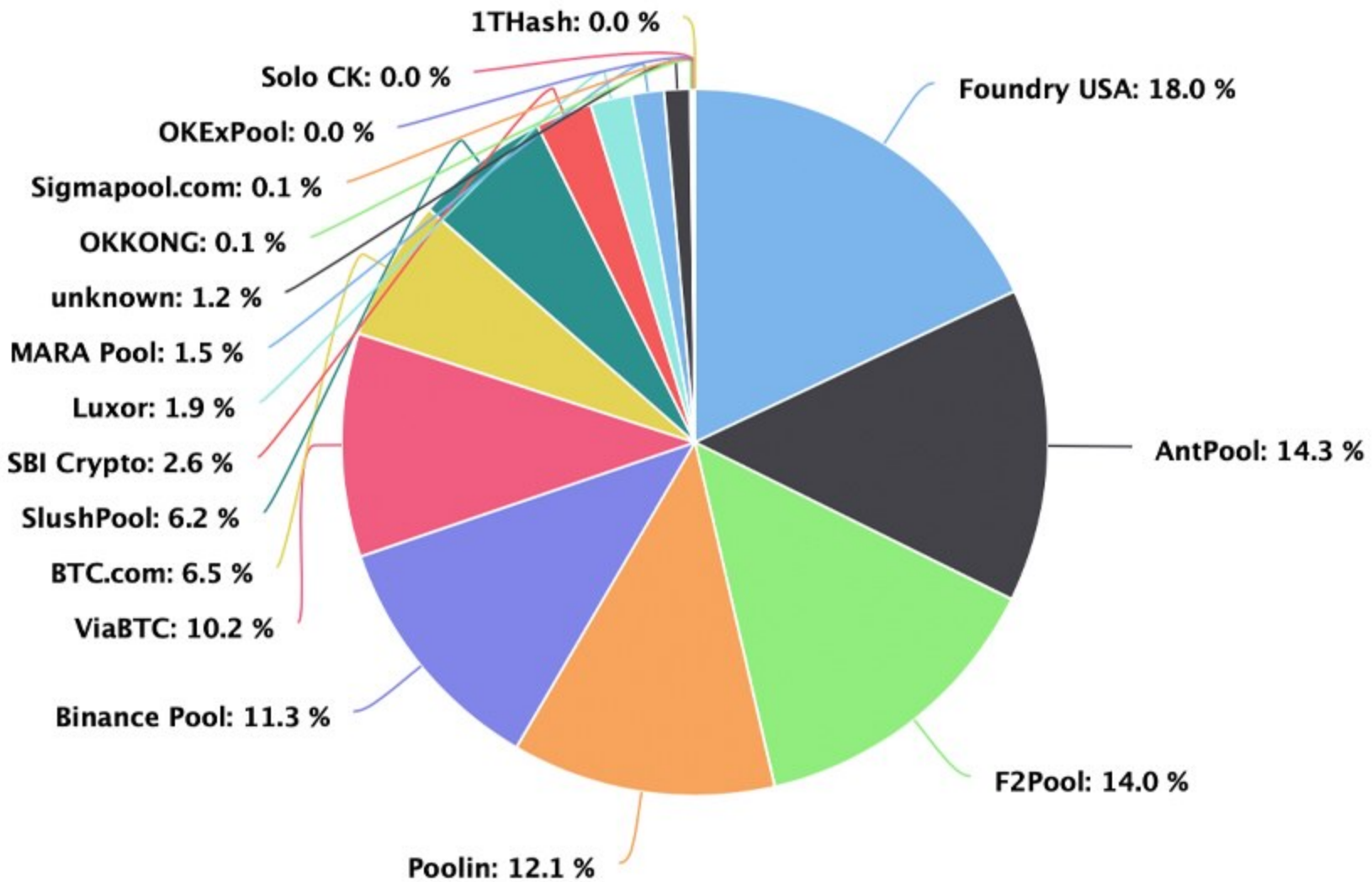
- L'hashrate totale rete Bitcoin (agosto 2021) = 100 milioni di tera-hash per secondo (100 EH/s)
- Affinché un miner (o un gruppo di miner) abbia una potenza di calcolo superiore al 50% dell'hashrate totale, deve quindi dotarsi di hardware per immettere in rete una potenza pari a 100 EH/s
- Il prezzo di un comune dispositivo per il mining è di circa 0,40€ per GH/s
- Il costo complessivo sarebbe quindi di circa $0{,}4*10^{11}$ € (40 miliardi di euro).
- Ai costi dell'investimento in dispositivi hw, andrebbero poi aggiunti gli enormi costi dell'energia necessaria per farli funzionare.

Ricompensa dei miner

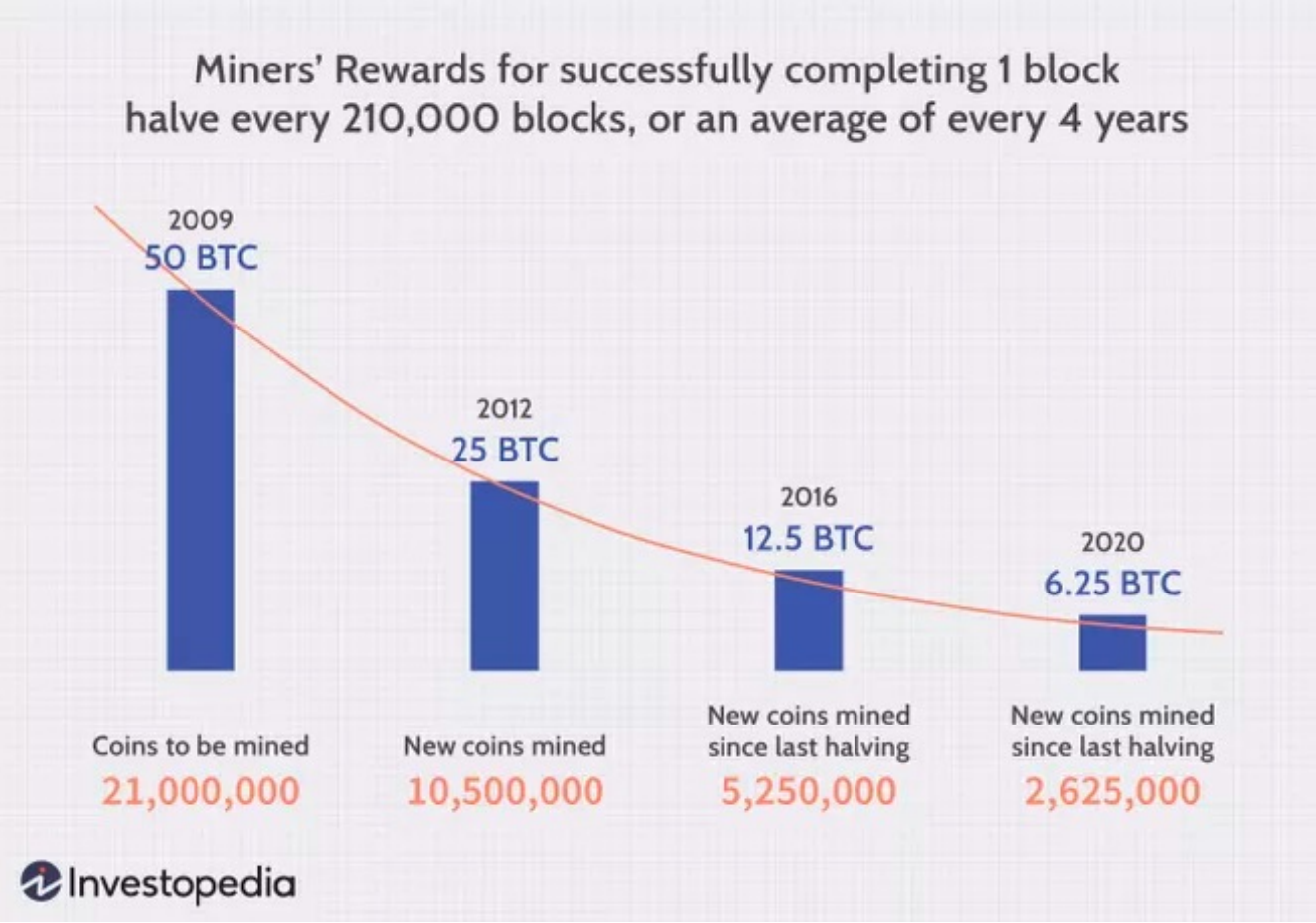
- I miner, attualmente, vengono remunerati con un premio di 6,25 bitcoin per ogni blocco confermato
- Il mining quindi produce una remunerazione di circa 37,5 bitcoin l'ora (6 blocchi l'ora), circa 330.000 bitcoin l'anno
- Con il bitcoin che oggi è quotato a circa 40.000 euro, ciò significa che la ricompensa complessiva è di cerca 13 miliardi di euro all'anno

Pool Distribution (calulate by blocks)

All 1 Y 3 M 1 M 1 W 3 D 24 H



Dimezzamento della ricompensa



- *halving: d*e ricompense per l'estrazione di Bitcoin si riducono di circa la metà ogni quattro anni
- Aumenta il valore del Bitcoin

La sicurezza della rete

- Tre fattori
  - tecniche crittografiche
  - architettura distribuita con ridondanza dei dati
  - protocollo del consenso - PoW

La sicurezza della rete - Crittografia

- Un nodo, non possedendo le chiavi private relative agli indirizzi di altri utenti, non può sottrarre fondi altrui e accreditarli nel proprio indirizzo.
- Ciò perché la crittografia a protezione del sistema è praticamente ineludibile, considerando che è impossibile ricavare la chiave privata a partire da quella pubblica o dall'indirizzo.

La sicurezza della rete - Ridondanza

- La BC è un database *resiliente* grazie alla sua struttura ridondante
- I dati memorizzati sulla Blockchain non possono infatti andare persi perché sono replicati su tutti i nodi della Blockchain
- Uno o più nodi possono guastarsi, essere sottoposti ad attacchi di haker, ecc., ma è praticamente impossibile che tutti i server della rete siano colpiti contemporaneamente da eventi negativi

La sicurezza della rete - Consenso

- L'*immutabilità* dei dati della BC deriva dal protocollo del consenso, dal quale consegue questo principio generale:

modificare in locale la propria BC privata può essere più o meno facile, ma fare accettare le modifiche al resto della rete, se non conformi al protocollo, è molto difficile

- Solo disponendo di almeno il 50% dell'hashrate totale è possibile un double-spending attack

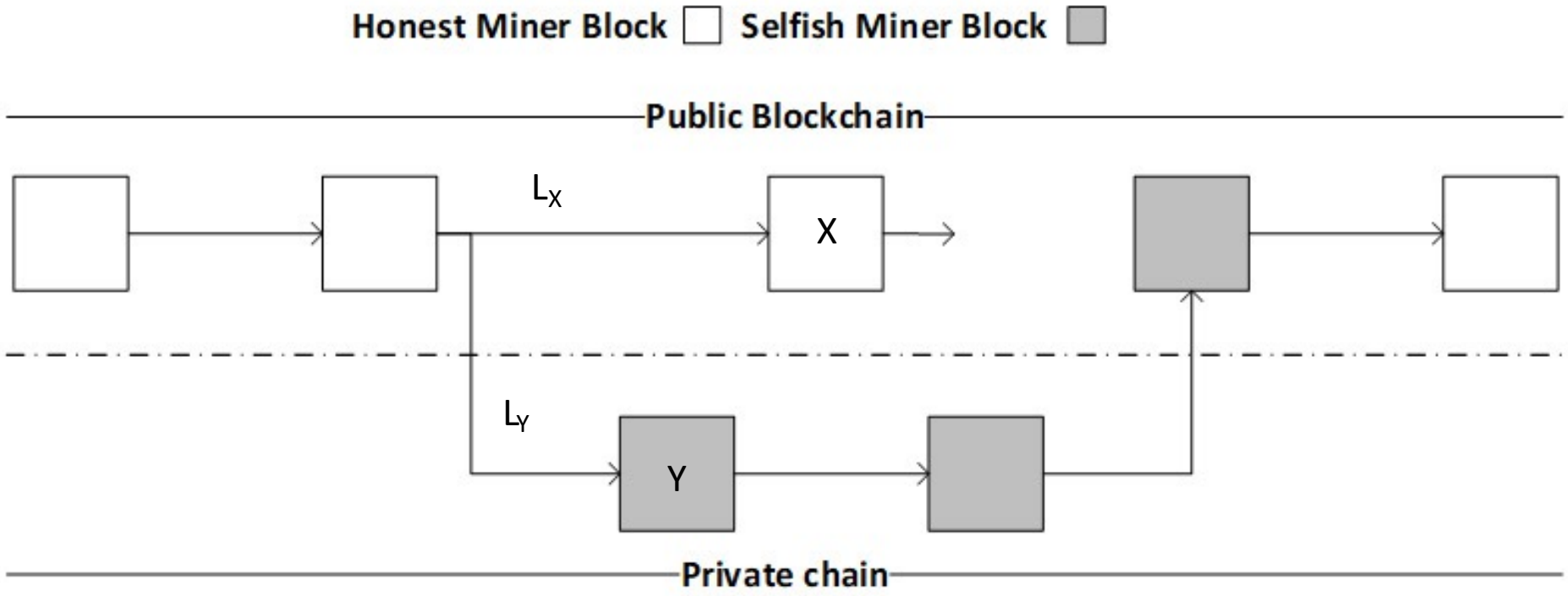
Double Spending

- Alice ha nel suo wallet 1 BTC, e tenta di spenderlo con due transazioni separate, A e B (per acquistare due beni ognuno del valore di 1 BTC)
- Se A viene inserita per prima in un blocco e confermata, la seconda transazione B verrà rigettata dalla rete
- Pertanto, un miner M disonesto che vuole utilizzare più volte lo stesso ammontare di denaro, dovrà ideare una strategia di attacco che consenta di eludere il controllo sulla compatibilità di A e B

Double Spending

1. M invia alla rete la transazione A con cui acquista un bene presso il commerciante C1
2. M crea segretamente un altro blocco Y (*selfish mining*) che contiene un'altra transazione B, in conflitto con A, con cui spende, a favore del commerciante C2, lo stesso denaro già speso con C1
3. Il blocco Y non viene comunicato alla rete, e viene agganciato alla BC locale di M allo stesso blocco cui è agganciato X, creando una biforcazione
4. La catena Lx, che contiene X, è quella nota alla rete e include le transazioni oneste, A compresa; la catena Ly non è invece visibile alla rete, ma è solo locale al nodo M

Double Spending



Double Spending

5. Quando il commerciante C1 vede confermata la transazione A, procede alla spedizione del bene
6. Mentre la catena Lx continua a crescere per il lavoro dei miner onesti, M continua a validare nuovi blocchi aggiungendoli alla catena privata Ly
7. se M ha sufficiente hashrate, prima o dopo Ly diventa più lunga di Lx. Quando ciò avviene, M rilascia Ly alla rete (un blocco alla volta, incominciando da Y)

Double Spending

8. A questo punto i nodi onesti passano, in accordo con il protocollo della BC, a lavorare sulla catena più lunga Ly in cui è presente B; M fa quindi proprie tutte le ricompense per la risoluzione dei nuovi blocchi di Ly
9. I blocchi della catena onesta Lx diventano orfani, cosicché le rispettive transazioni dovranno essere ri-confermate; ma, poiché A è inconsistente con B (in quanto ha speso il denaro che utilizza A), essa verrà rigettata
10. Il risultato netto è che solo il commerciante C2 riceverà il pagamento, mentre il miner M è entrato in possesso sia del bene acquistato presso C1 che di quello acquistato presso C2

Double Spending

- La possibilità di successo del *double spending* è chiaramente legata alla hashrate di M. In particolare, vale quanto segue:
  - se M controlla più del 50% dell'hashrate totale, *certamente* riuscirà nel tentativo fraudolento – in altri termini, se riesce a minare blocchi più rapidamente di tutti gli altri nodi della rete messi assieme, riuscirà nel tentativo di creare una catena più lunga
  - se invece l'hashrate di M è inferiore alla soglia del 50%, la probabilità di successo decresce esponenzialmente.

Controindicazioni del PoW

- Eccessivo consumo di energia: una elevata hashrate richiede elevati consumi di energia, e relativi danni ambientali. Il Bitcoin, che è basato sul PoW, consuma circa lo 0,3% della produzione mondiale di energia. Questo è il costo che la rete deve pagare per garantire sicurezza e immutabilità
- Concentrazione del mining: creazione di grandi *mining farm* e *mining pool* per mettere in campo un potere computazionale elevatissimo. Con la conseguenza che l'hashrate totale è aumentata, ed è diminuita la probabilità di vincere la gara da parte dei nodi poco attrezzati