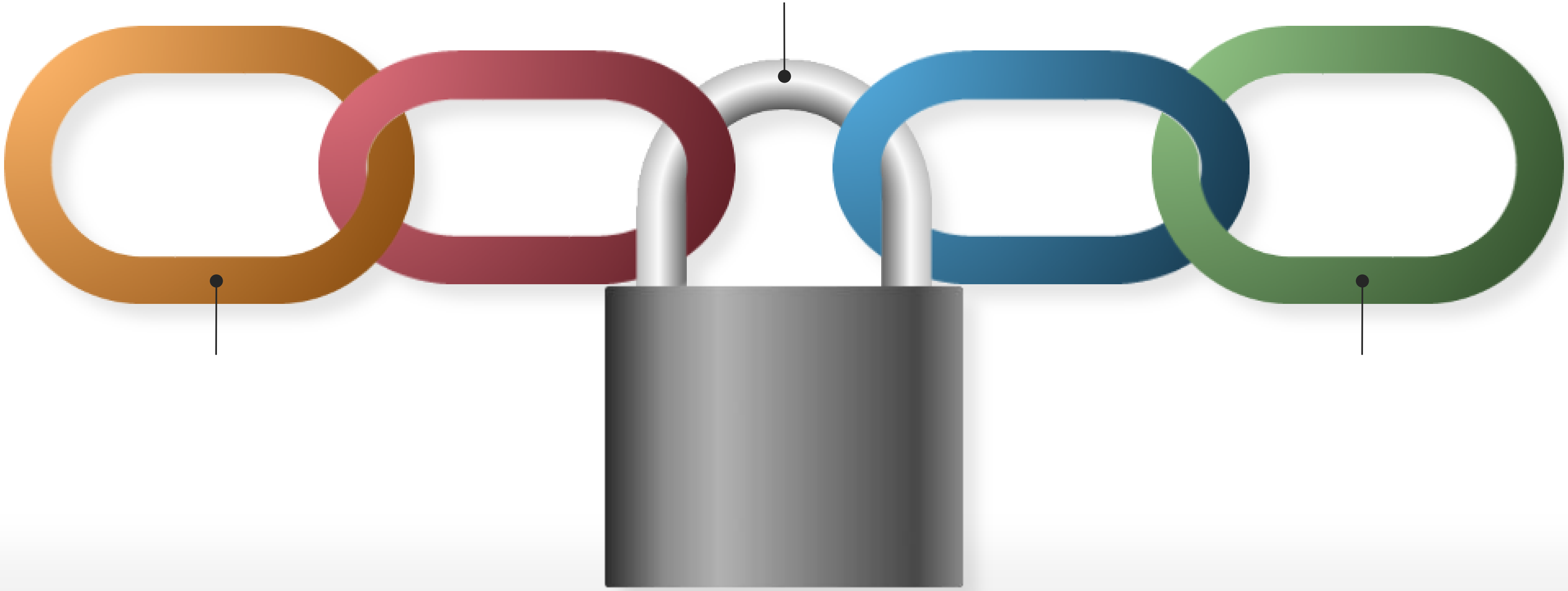
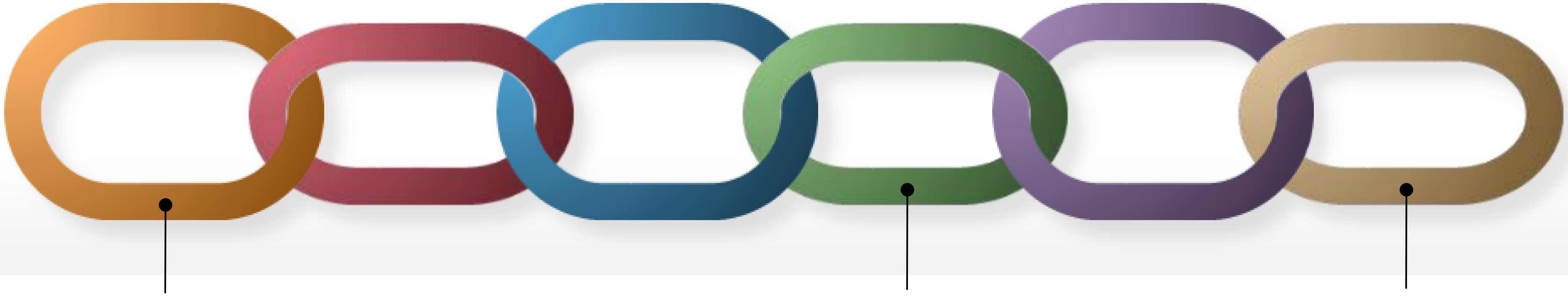


Data Privacy in Blockchains: Theory and Practice

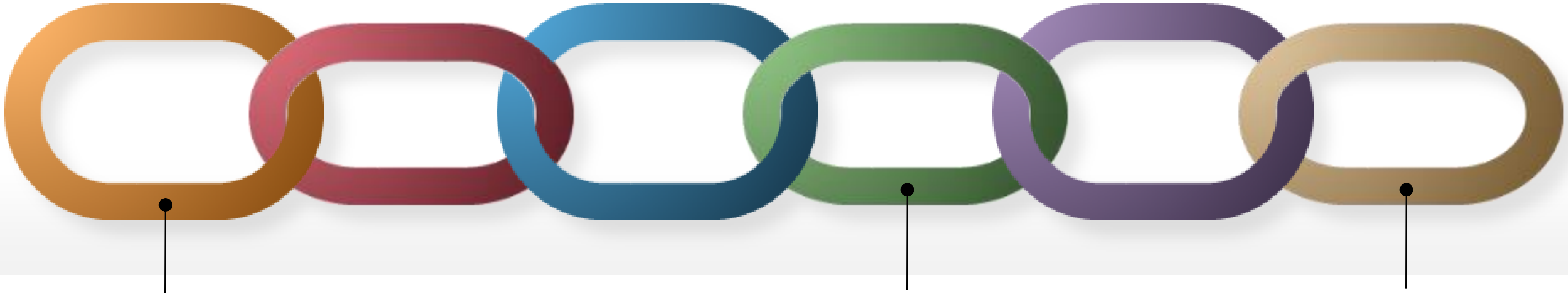
Ivan Visconti

Università di Salerno (DIEM)



Warm up

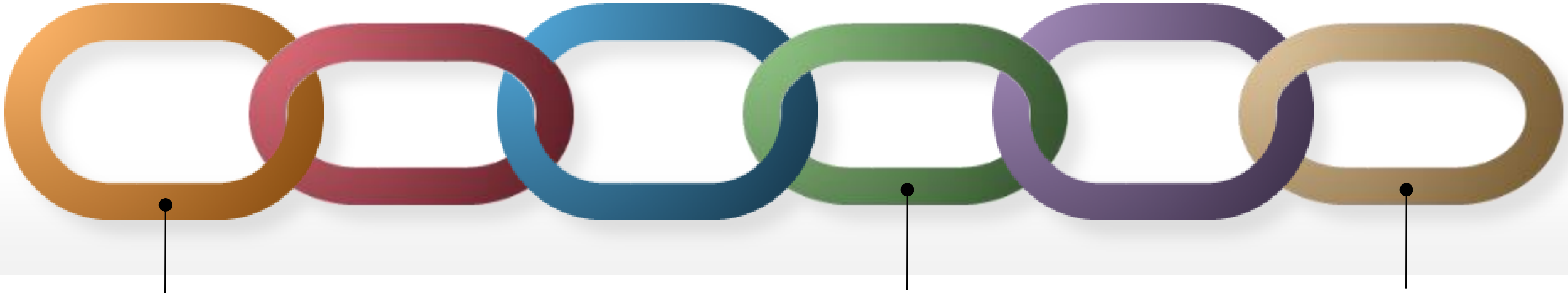
a blockchain must guarantee immutability of the past



Warm up

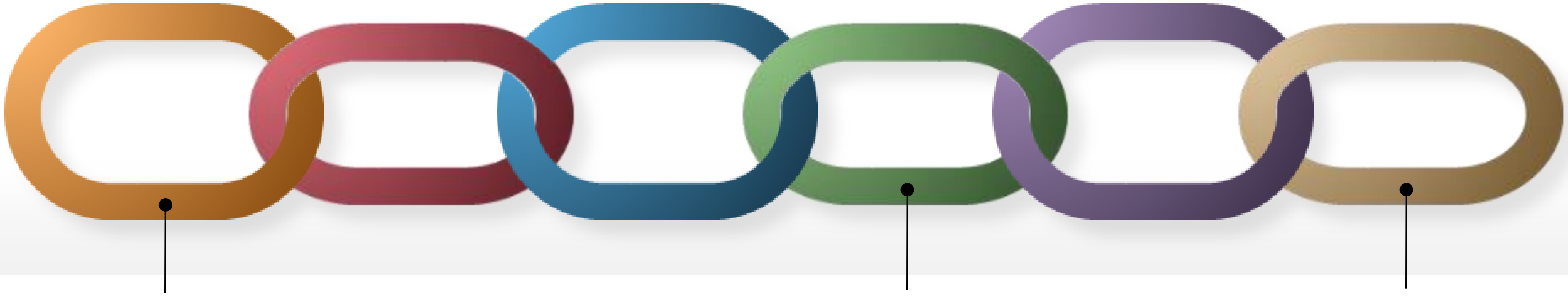
a blockchain must guarantee immutability of the past

WRONG!



Warm up

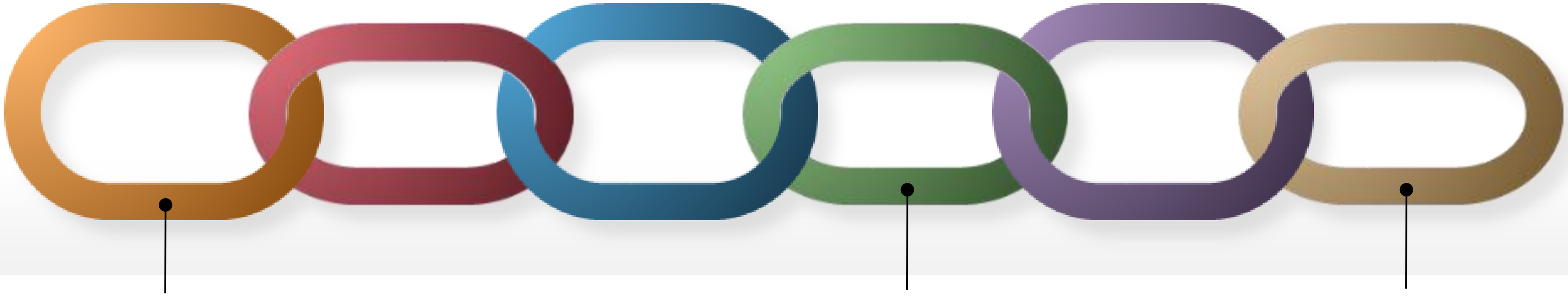
removing data from the Bitcoin blockchain, still allowing everyone to verify transactions, is almost impossible
(it requires a huge amount of hashing power)



Warm up

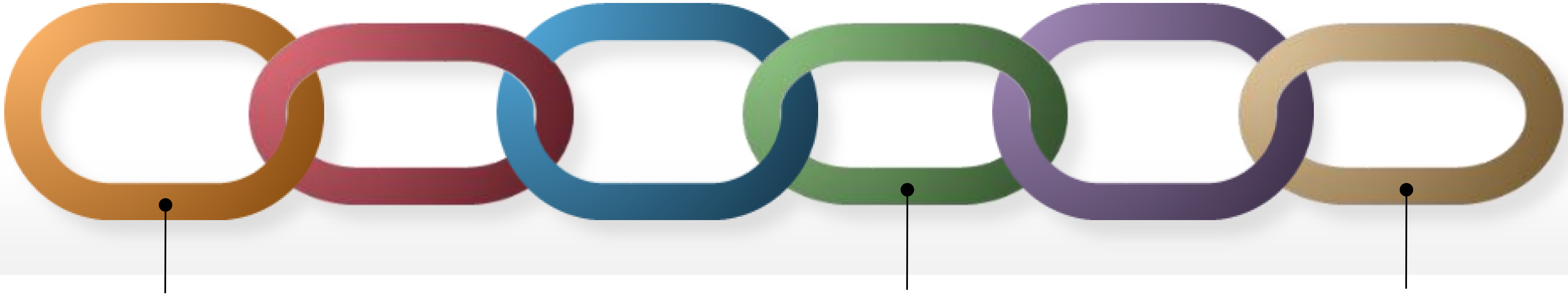
removing data from the Bitcoin blockchain, still allowing everyone to verify transactions, is almost impossible
(it requires a huge amount of hashing power)

WRONG!



Warm up

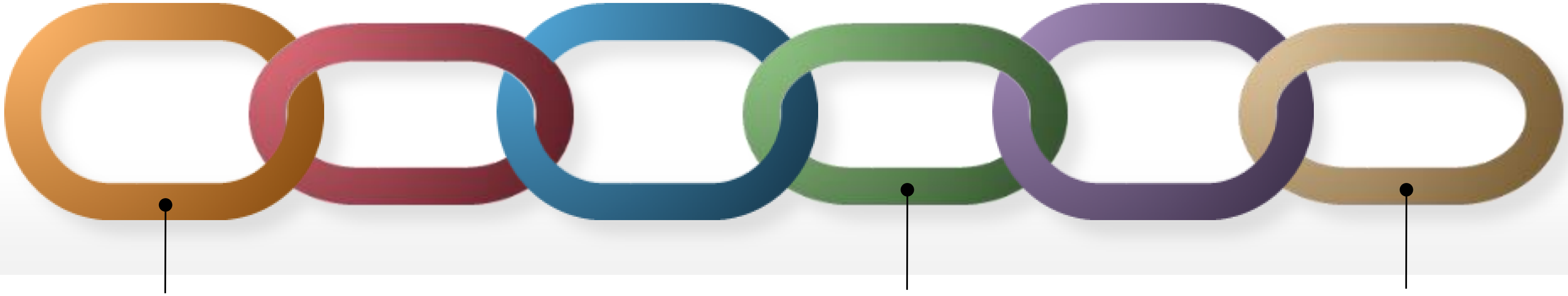
cryptographically hashing data on a blockchain allows to exploit the power of a blockchain keeping data **confidential**



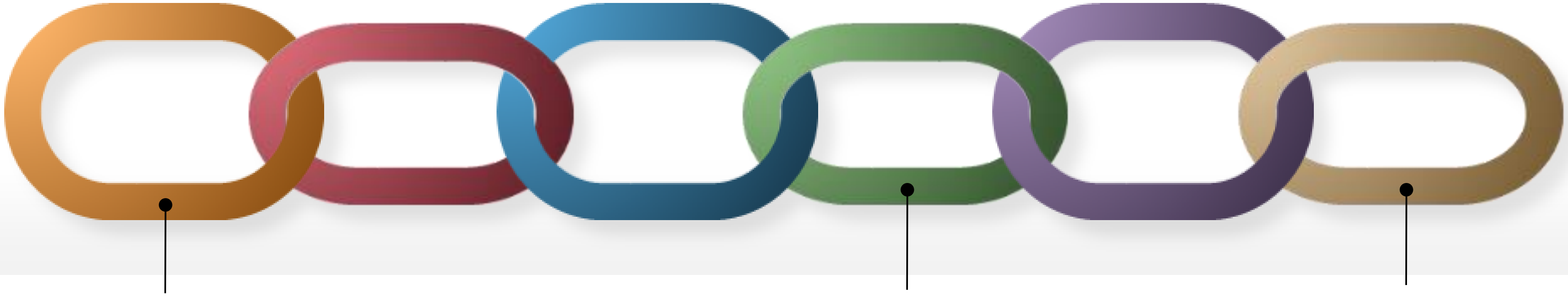
Warm up

cryptographically hashing data on a blockchain allows to exploit the power of a blockchain keeping data **confidential**

PROBLEMATIC!

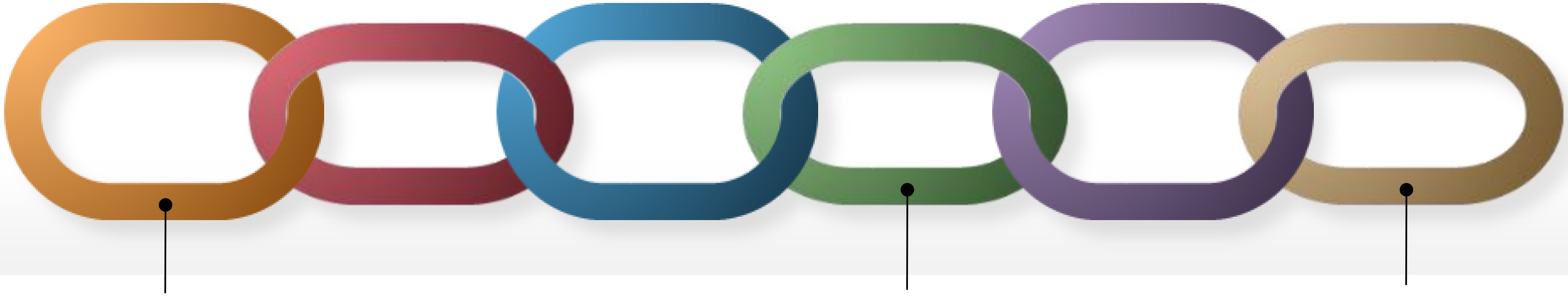


No worries, there will be lots of good news



Blockchain (informal) Definition

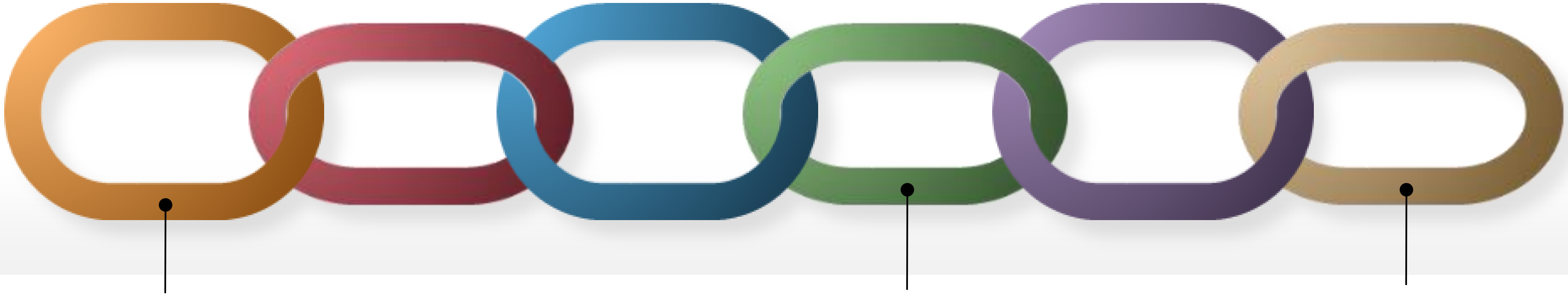
a blockchain is a *decentralized* computer that *publicly* runs programs (smart contracts); each program receives inputs (transactions)



Blockchain (informal) Definition

a blockchain is a *decentralized* computer that *publicly* runs programs (smart contracts); each program receives inputs (transactions)

public verifiability holds also without being permanently online

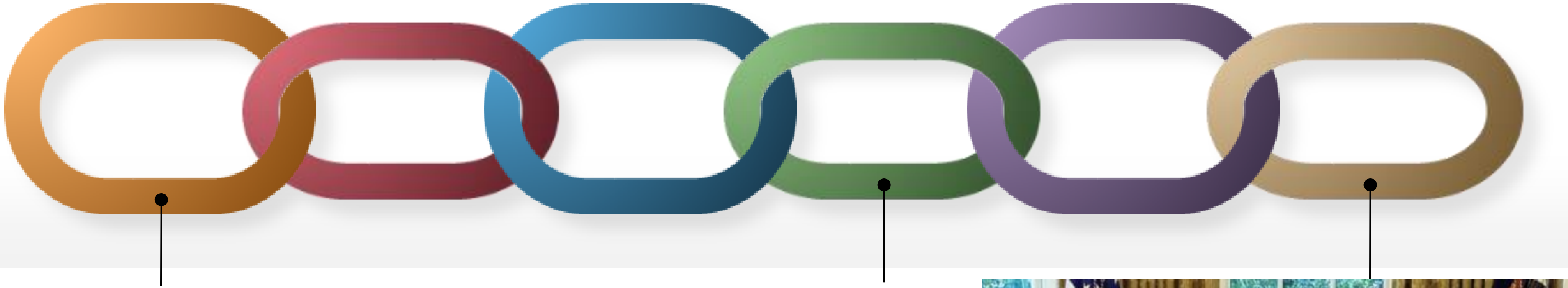


Blockchain (informal) Definition

a blockchain is a *decentralized* computer that *publicly* runs programs (smart contracts); each program receives inputs (transactions)

public verifiability holds also without being permanently online

if you would like to see a formal definition then you should look at rigorous property like *chain growth, quality, consistency see*
[Analysis of the Blockchain Protocol in Asynchronous Networks – Pass, Seeman, Shelat 2016]



Resilience and Transparency

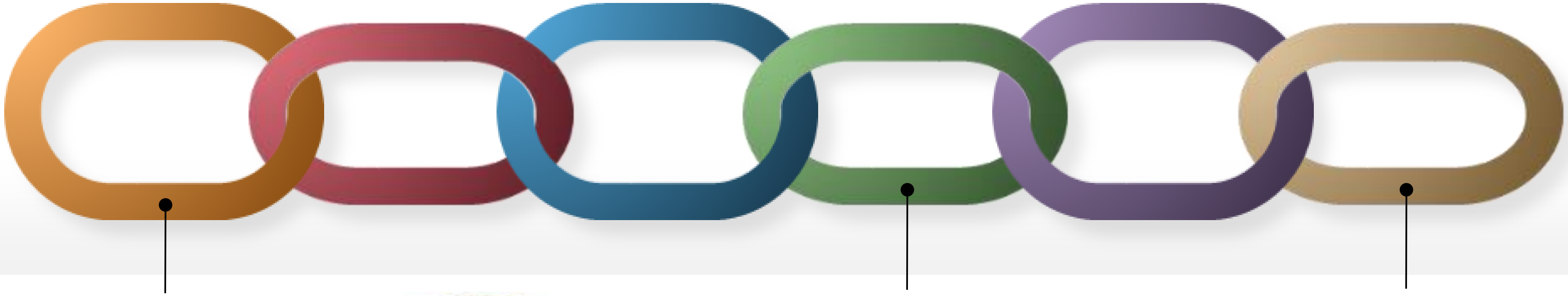
a decentralized computer can work perfectly even in case of a large scale attack

everyone can verify the correctness of the current state of the execution of any program

==> publicly verifiable integrity check of processes

In other words....

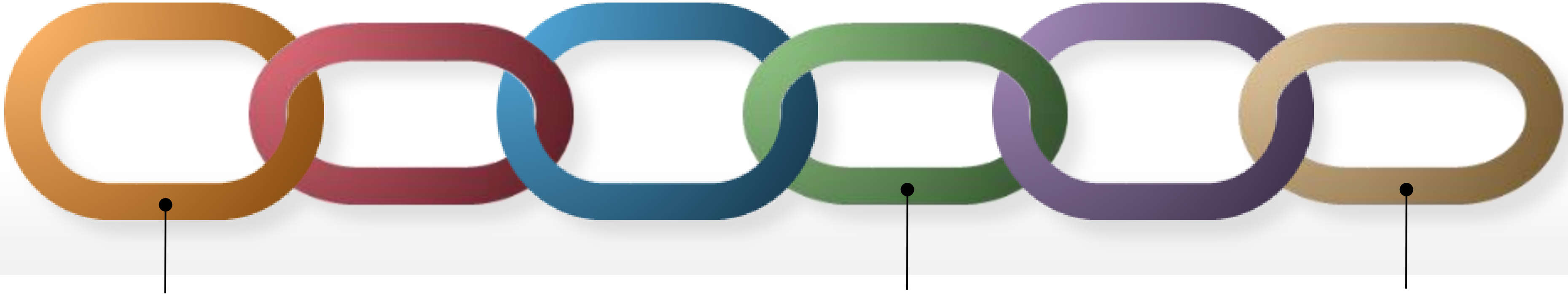
we can consider Blockchain technology as a disruptive tool against counterfeiting of processes



Bitcoin

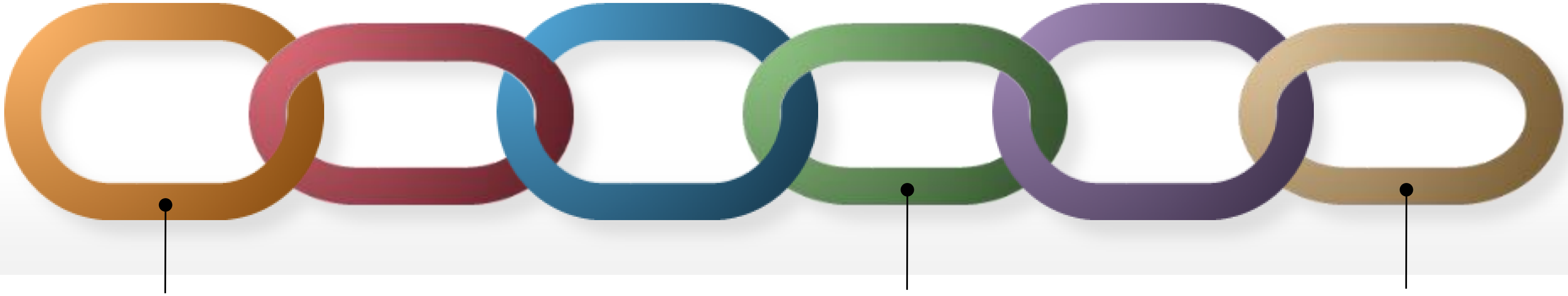
A blockchain running two programs:

1) one to mint coins and assign them through a lottery
2) one to transfer coins among wallets



Ethereum

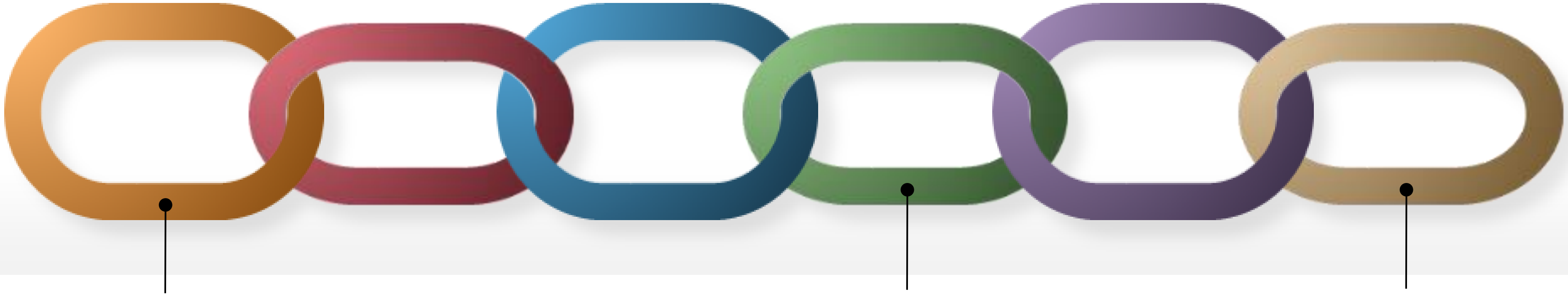
supports generic programs



Ethereum

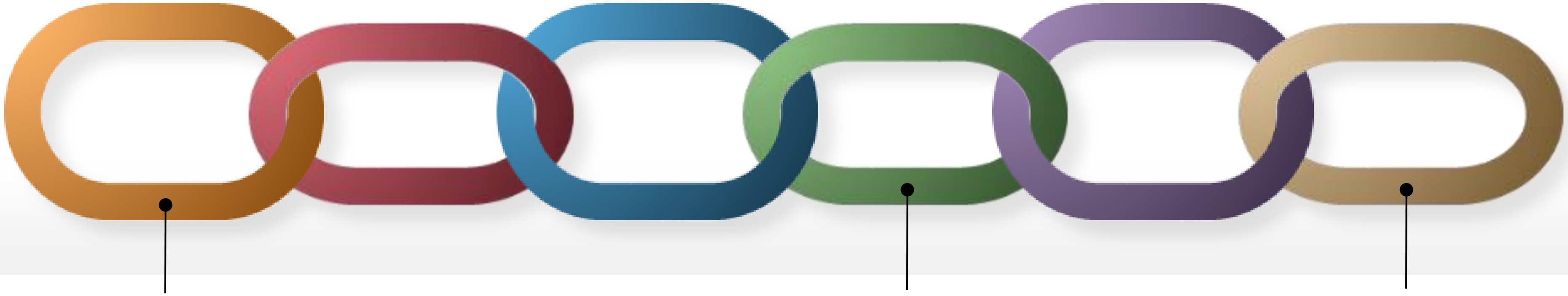
supports generic programs

Vitalik Buterin claimed that you can’t have *scalability*, *decentralization* and *security* at the same time…
This is the blockchain **Trilemma**… (more later about it)



You might want to use a Blockchain every time you are afraid of cheating

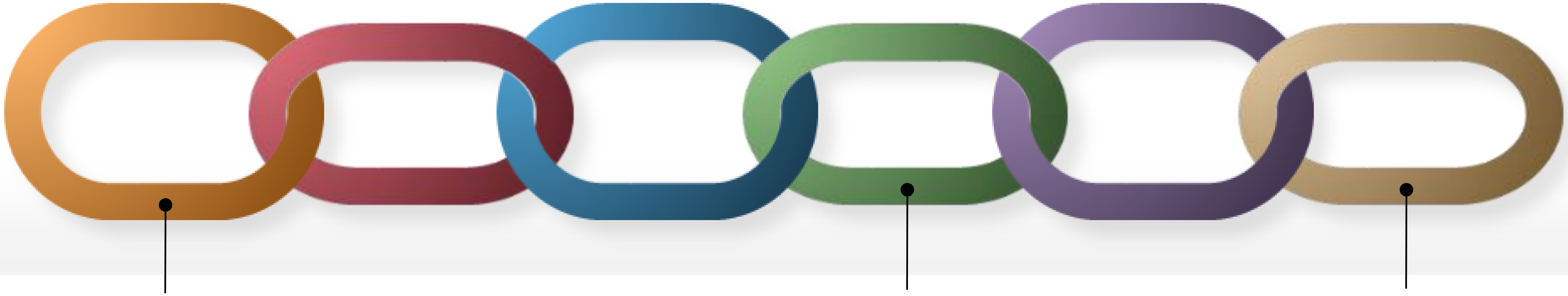
- e-voting, supply chain
- lotteries, games
- or more generically: any problem trivially resolved with the help of a trusted third party



Do you see any problem?
What about privacy?

it does not seem that you can have public verifiability/transparency along with privacy

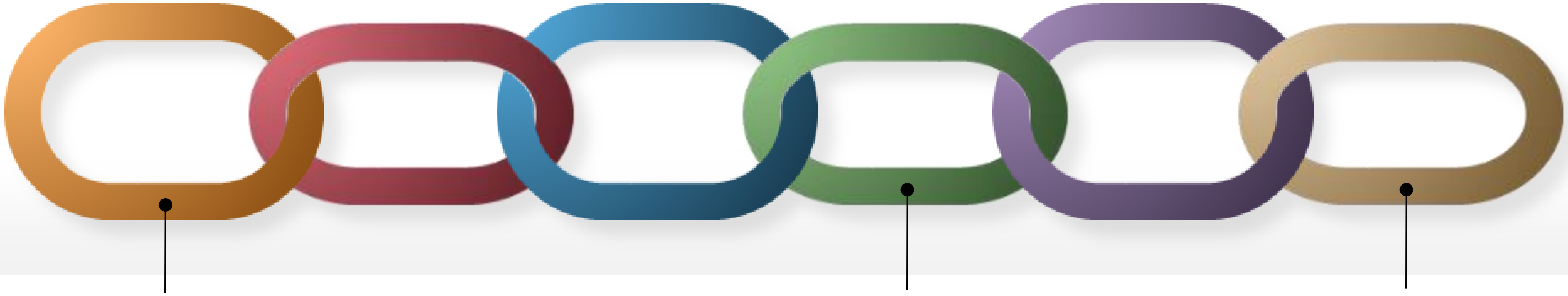
how can we use the integrity of a blockchain still preserving privacy?



Two classical goals in Cryptography

- Data Integrity (e.g., digital signatures)
- Data Confidentiality (e.g., encryption)

sometimes we need both simultaneously (e.g., e-commerce)



Two classical goals in Cryptography

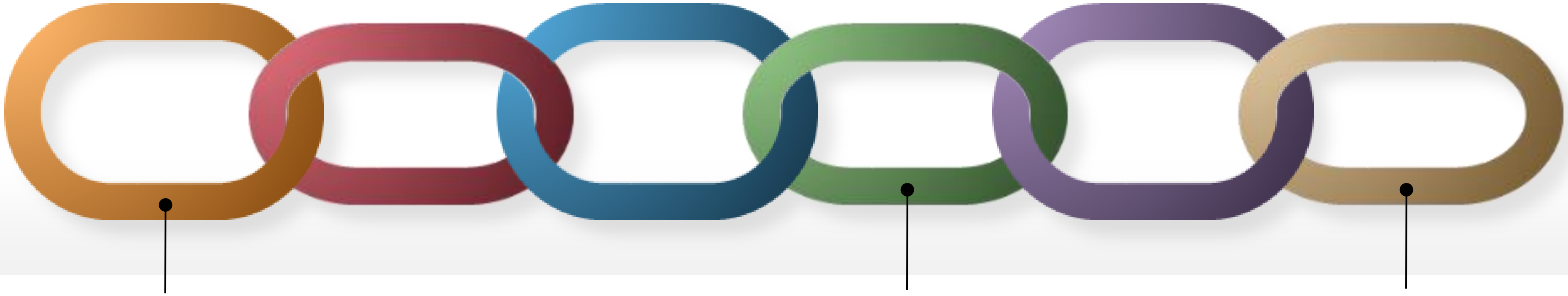
- Data Integrity (e.g., digital signatures)
- Data Confidentiality (e.g., encryption)

sometimes we need both simultaneously (e.g., e-commerce)

Blockchains are friendly with data integrity
Blockchains are hostile with data confidentiality

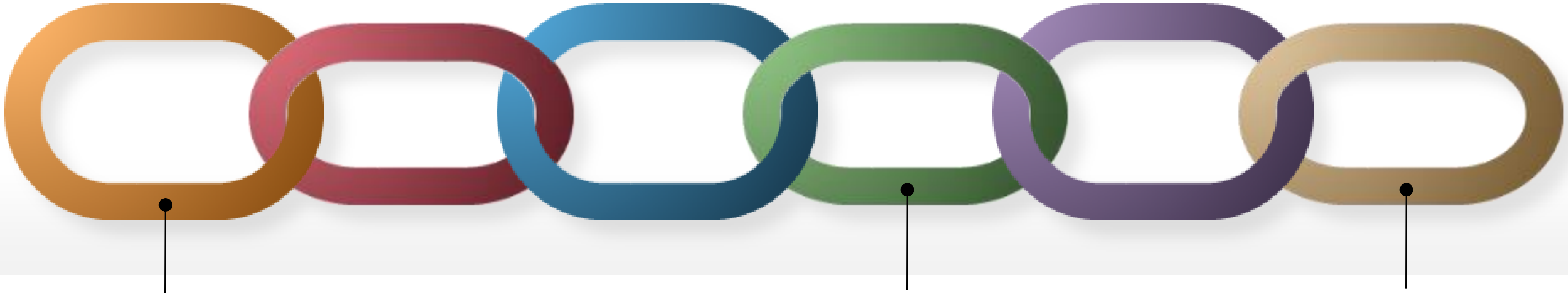
Blockchain: Privacy and GDPR

are immutability, public verifiability and compatible with current (and future) GDPR?



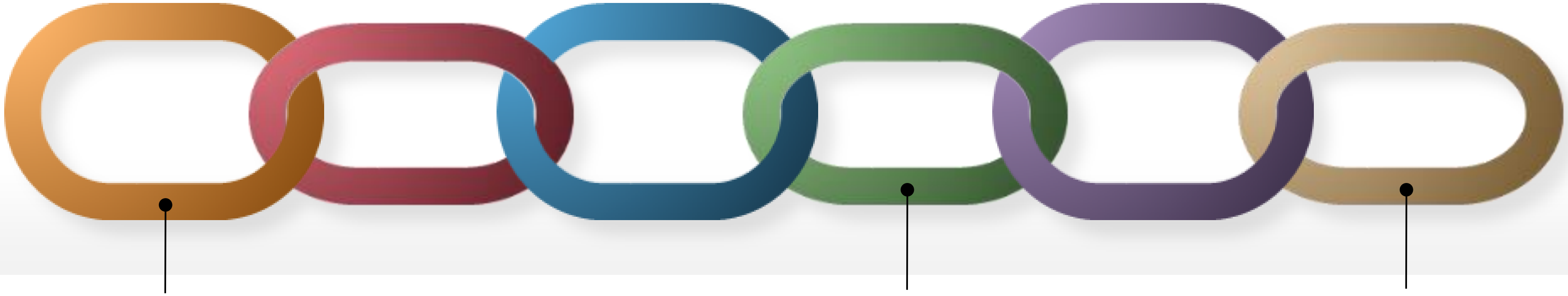
Blockchain Technology another revolution after the Internet?

Internet has been secured using standard cryptographic tools that are instead insufficient for privacy in blockchains (more details later)



Ethereum

Vitalik Buterin claimed that you can’t have *scalability*, *decentralization* and *security* at the same time…
This is the blockchain **Trilemma**…

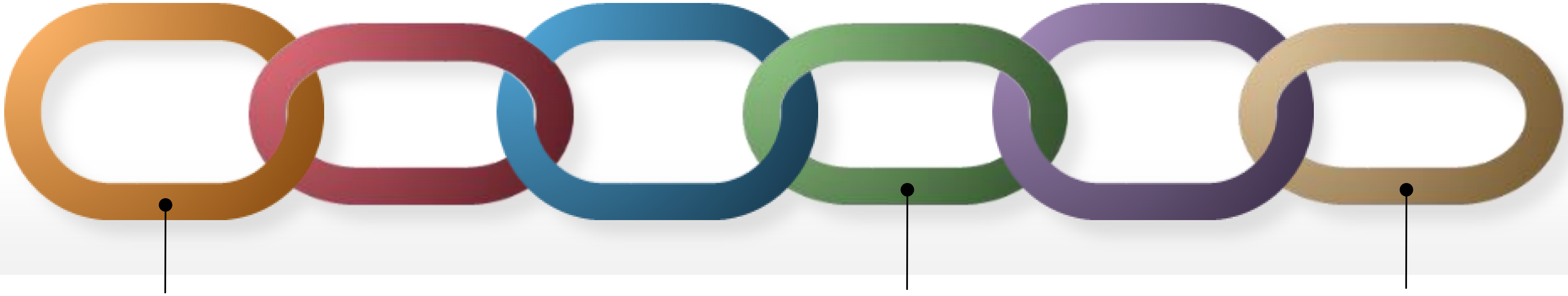


Ethereum

Vitalik Buterin claimed that you can’t have *scalability*, *decentralization* and *security* at the same time… This is the blockchain **Trilemma**…

It’s even worse:
we also need *compliance* with laws (e.g., GDPR...)
in other words, we have a **Quadrilemma** now…

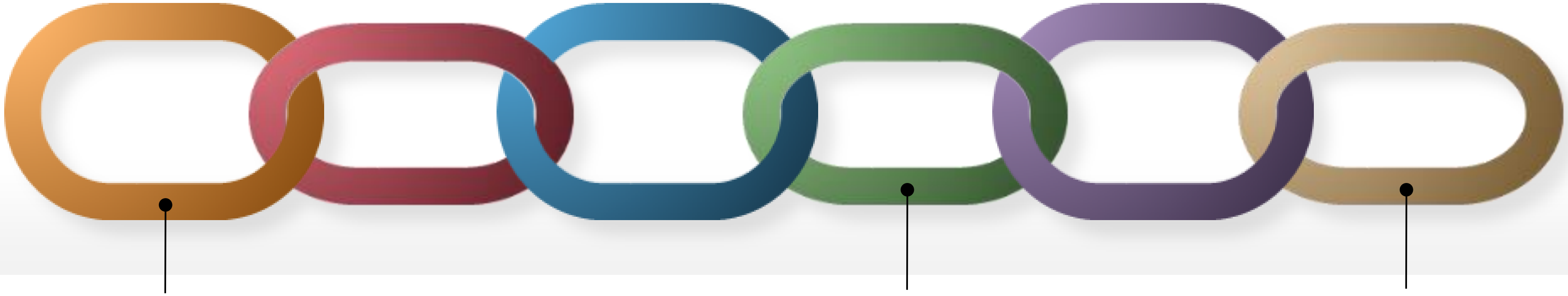
Concordium



Summing up

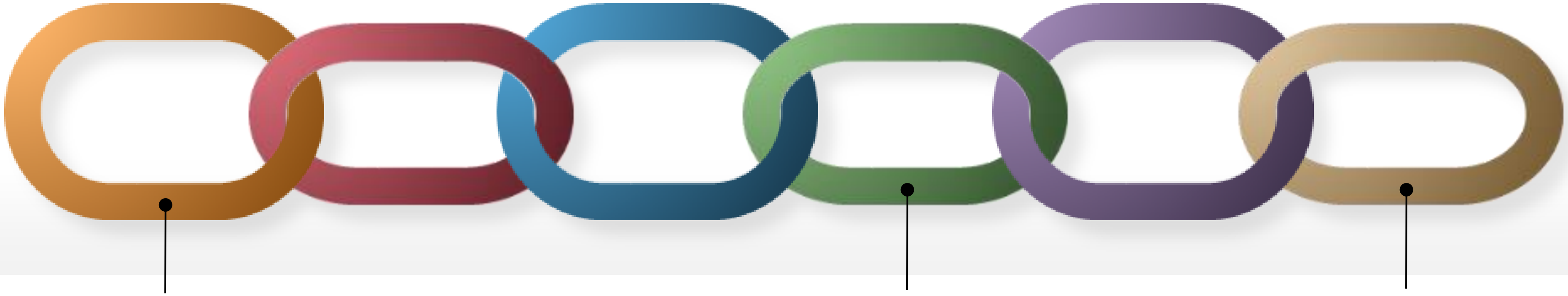
if you want to use a Blockchain for applications involving confidential data then:

- keep in mind that there is no much trust around
- keep in mind that it must be efficient
- keep in mind that it is must be secure
- keep in mind that you must obey to regulations



Challenge number 1: right to be forgotten…

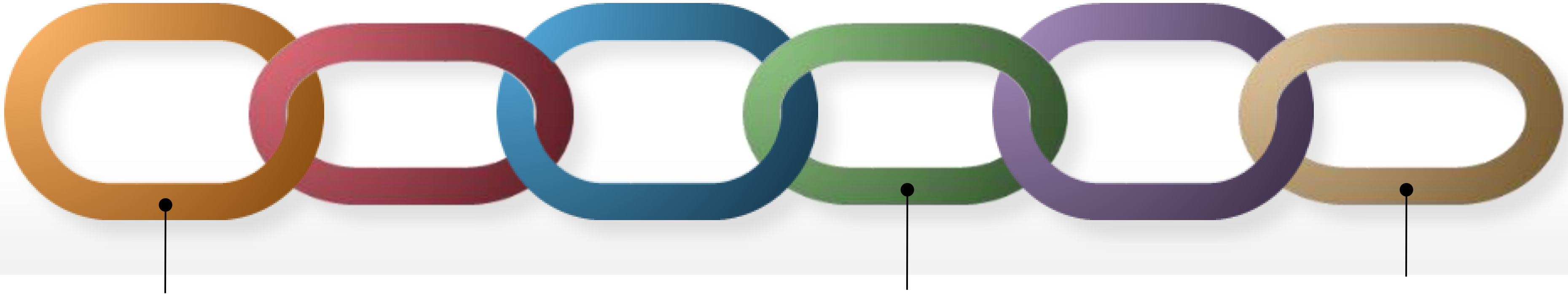
no way… blockchains are immutable no?



Challenge number 1: right to be forgotten…

no way… blockchains are immutable no?

that’s what many blockchain enthusiasts say… but…

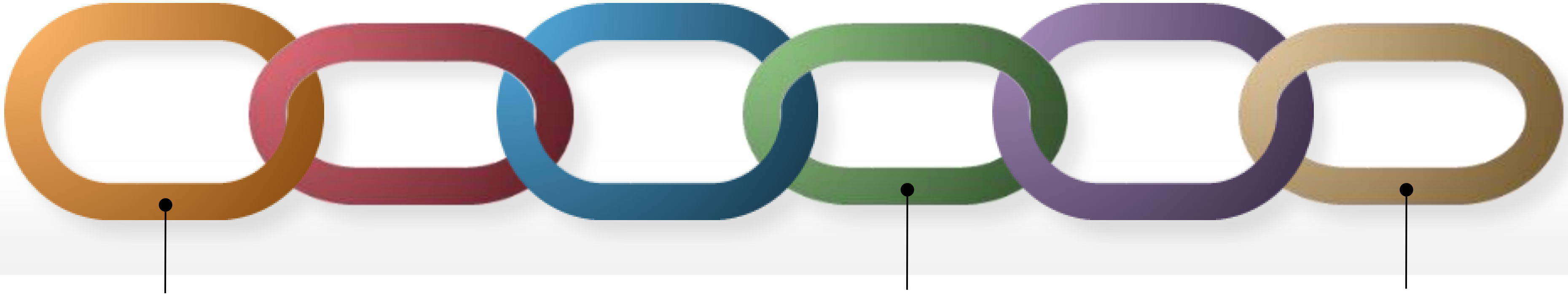


Blockchain Definition

a blockchain is a *decentralized* computer that *publicly* runs programs (smart contracts)

each program waits for some input (a transaction) to perform some *public* task

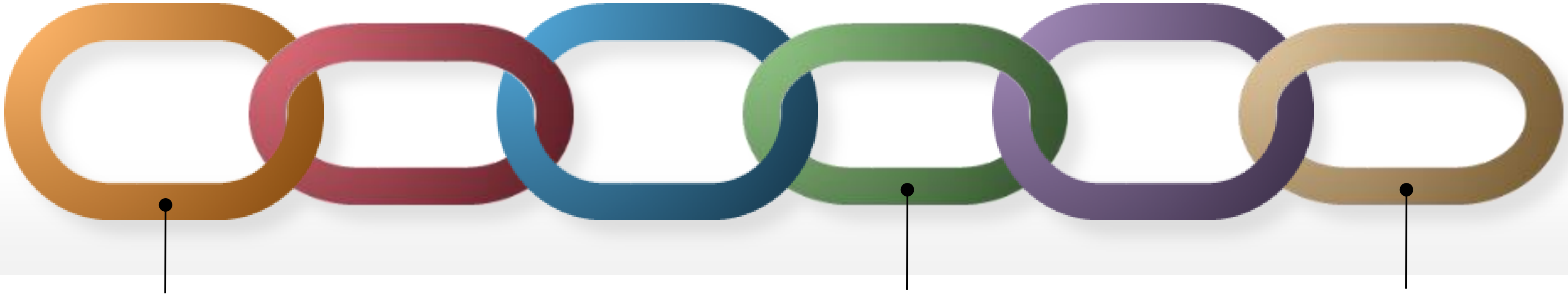
public verifiability holds also without being permanently online (… this does not necessarily imply immutability…)



CASE 1: Permissioned Blockchains

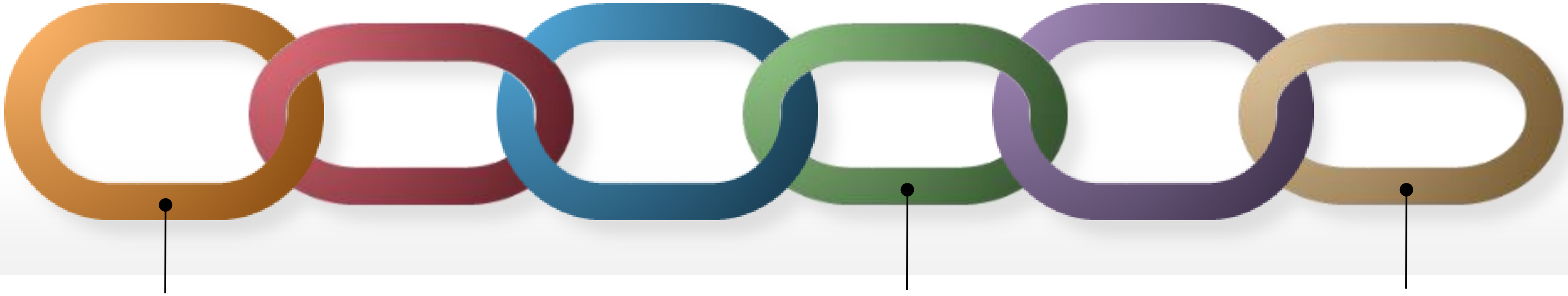
can we remove data?

YES, of course! and it is **trivial**! it’s enough that actors in charge are willing to do it…. and it does not require any fancy cryptography…



CASE 1: Permissioned - TRIVIAL

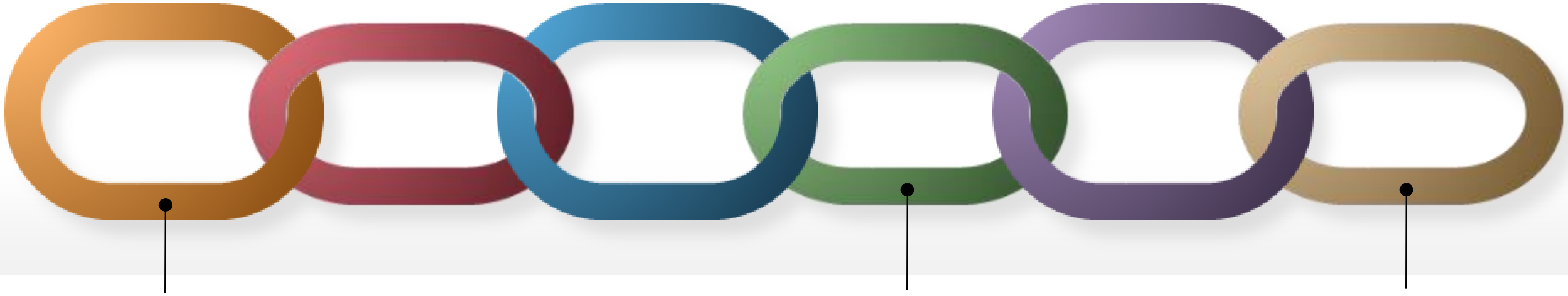
1) no cryptographic hash is required to link blocks; block N includes number N, transactions, timestamp and endorsing signatures



CASE 1: Permissioned - TRIVIAL

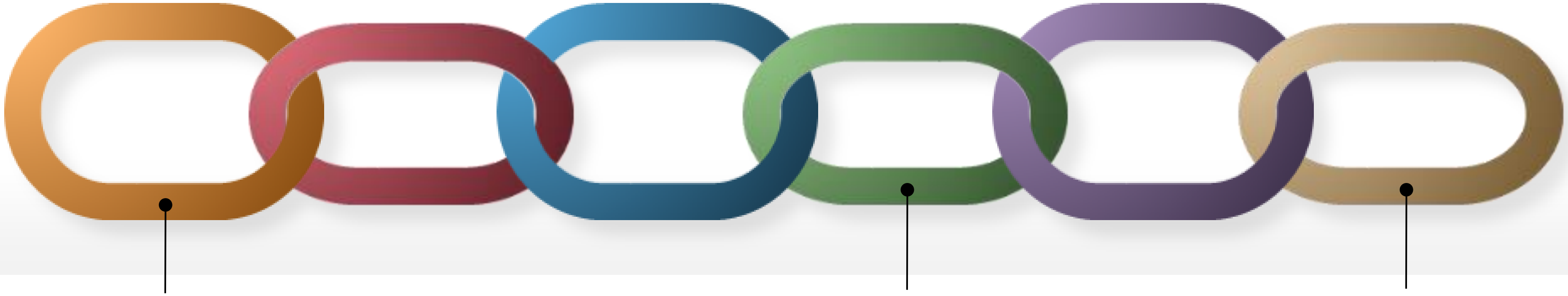
1) no cryptographic hash is required to link blocks; block N includes number N, transactions, timestamp and endorsing signatures

2) delete block N by endorsing a new block N with a fresher timestamp (plus some other potential updates)



CASE 1: Permissioned - TRIVIAL

1) no cryptographic hash is required to link blocks; block N includes number N, transactions, timestamp and endorsing signatures
2) delete block N by endorsing a new block N with a fresher timestamp (plus some other potential updates)
3) use threshold signatures [Crypto 91] if you want compactness

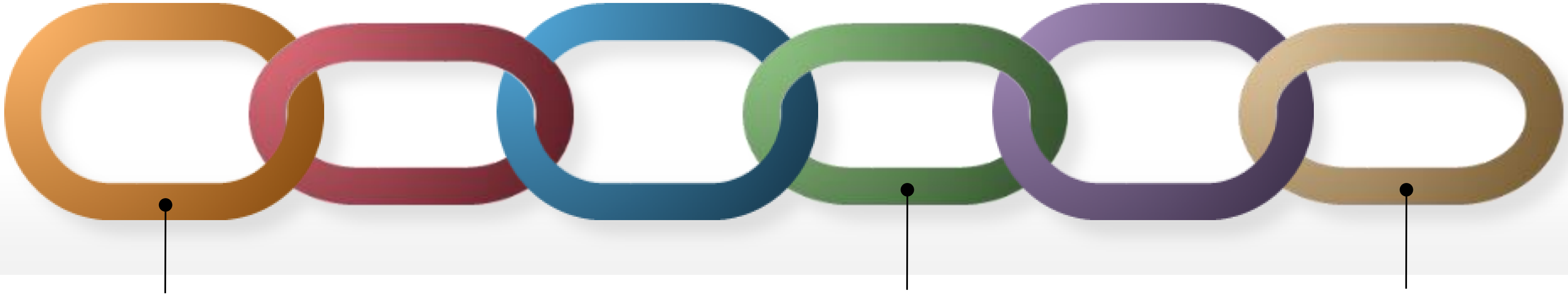


CASE 1: Permissioned - TRIVIAL

1) no cryptographic hash is required to link blocks; block N includes number N, transactions, timestamp and endorsing signatures

2) delete block N by endorsing a new block N with a fresher timestamp (plus some other potential updates)

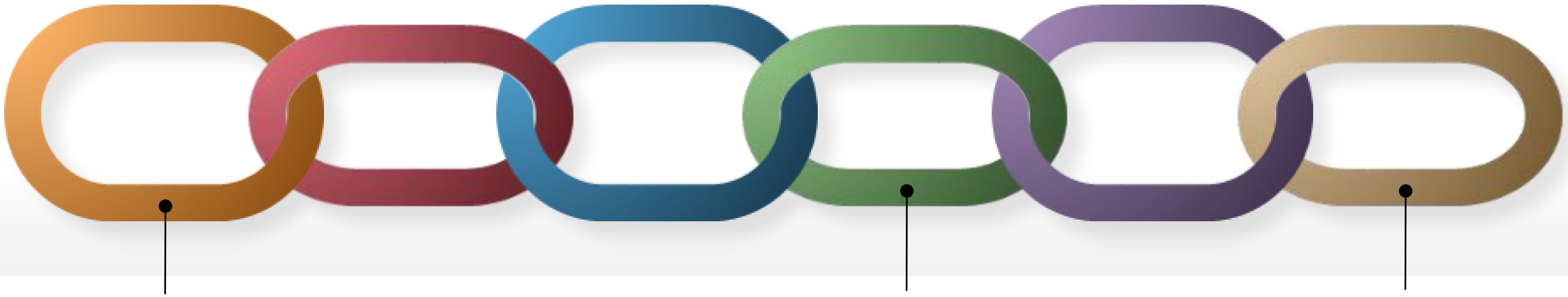
3) use threshold signatures [Crypto 91] if you want compactness



Recall that

if you want to use a Blockchain for applications involving confidential data then:

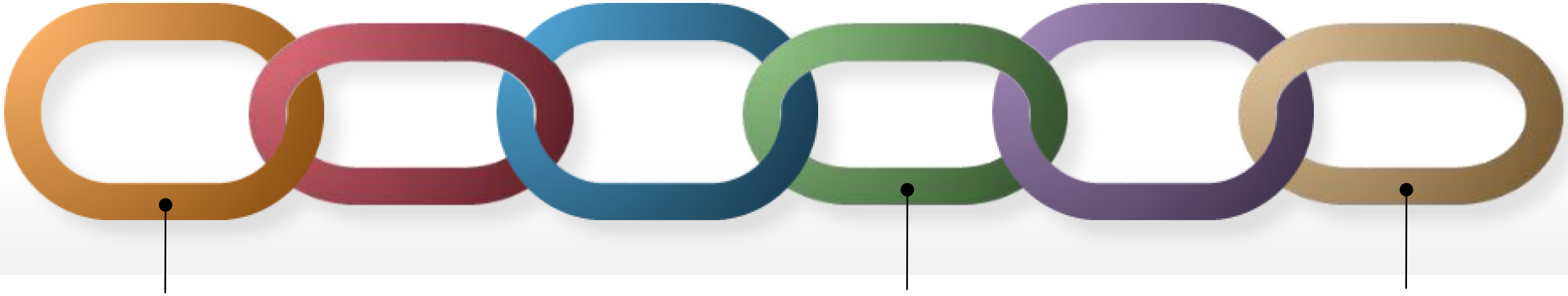
- keep in mind that there is no much trust around
- keep in mind that it must be efficient
- keep in mind that it must be secure
- keep in mind that you must obey to regulations



CASE 2: Permissionless Blockchains

let’s go straight to the point:

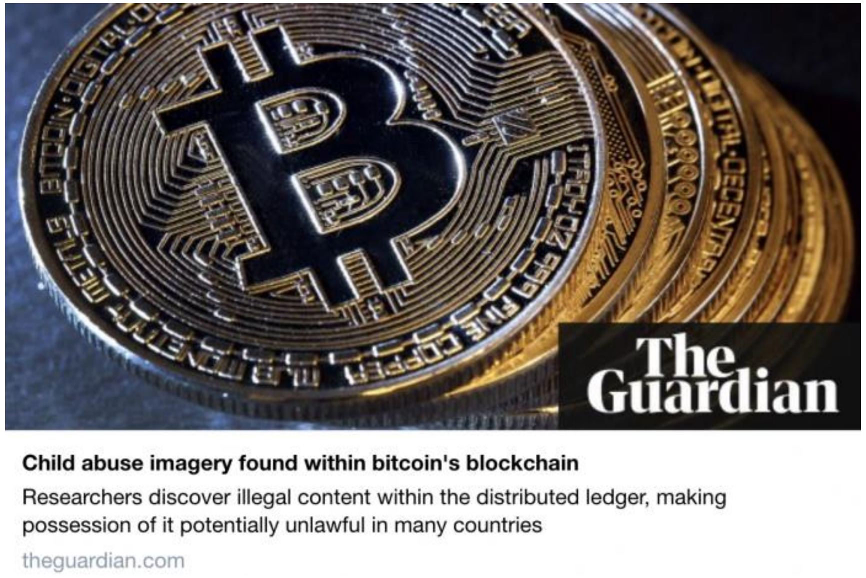
Bitcoin, THE BLOCKCHAIN



Bitcoin Blockchain

in some transactions of the Bitcoin blockchain there are links to child pornography

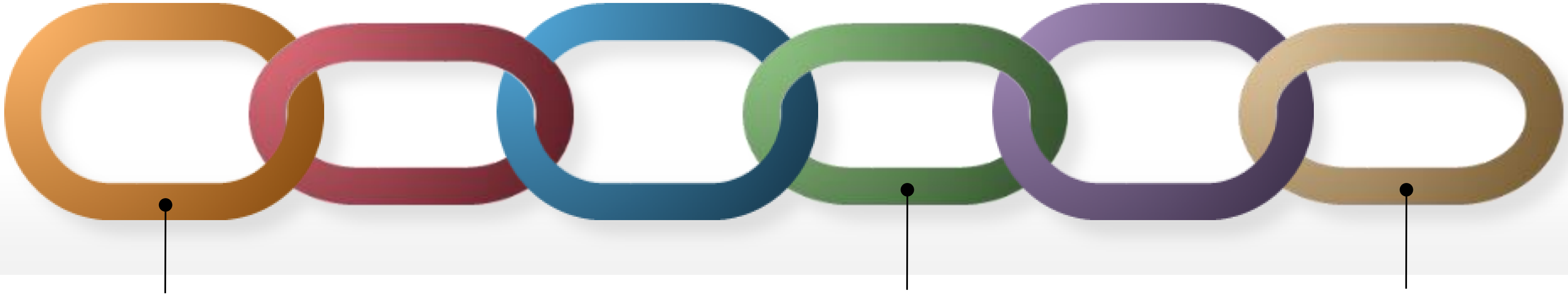
[Financial Cryptography 2018]



Child abuse imagery found within bitcoin's blockchain

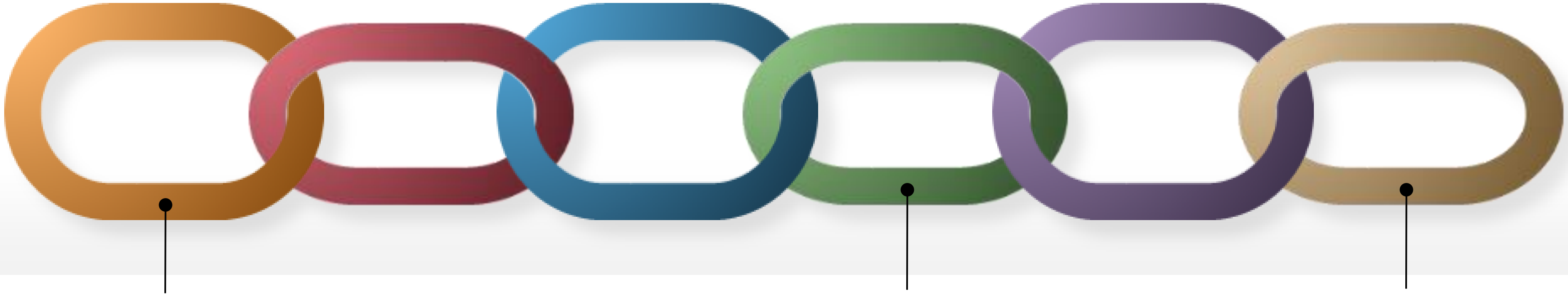
Researchers discover illegal content within the distributed ledger, making possession of it potentially unlawful in many countries

theguardian.com



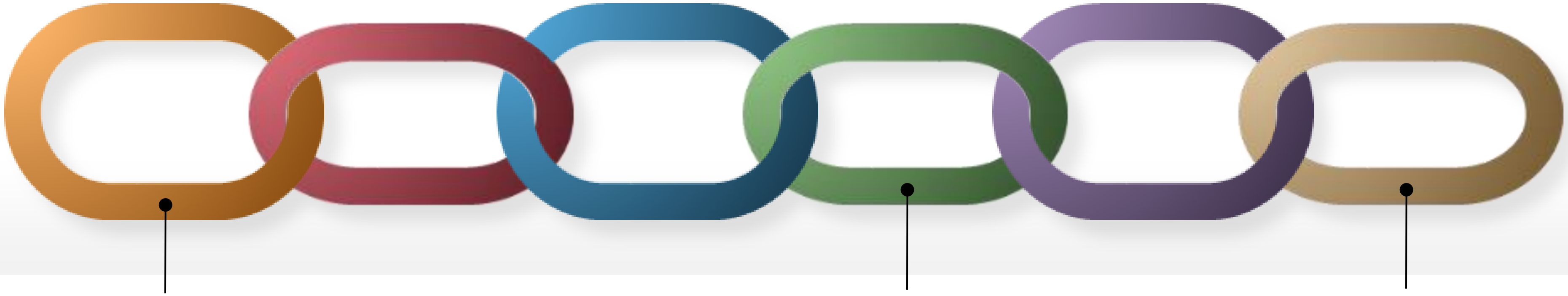
~~Public Verifiability~~

by removing a single transaction, an entire process becomes untrusted



Bitcoin Blockchain

can we remove illicit data? we all know that Bitcoin is secure because the history is untouchable!

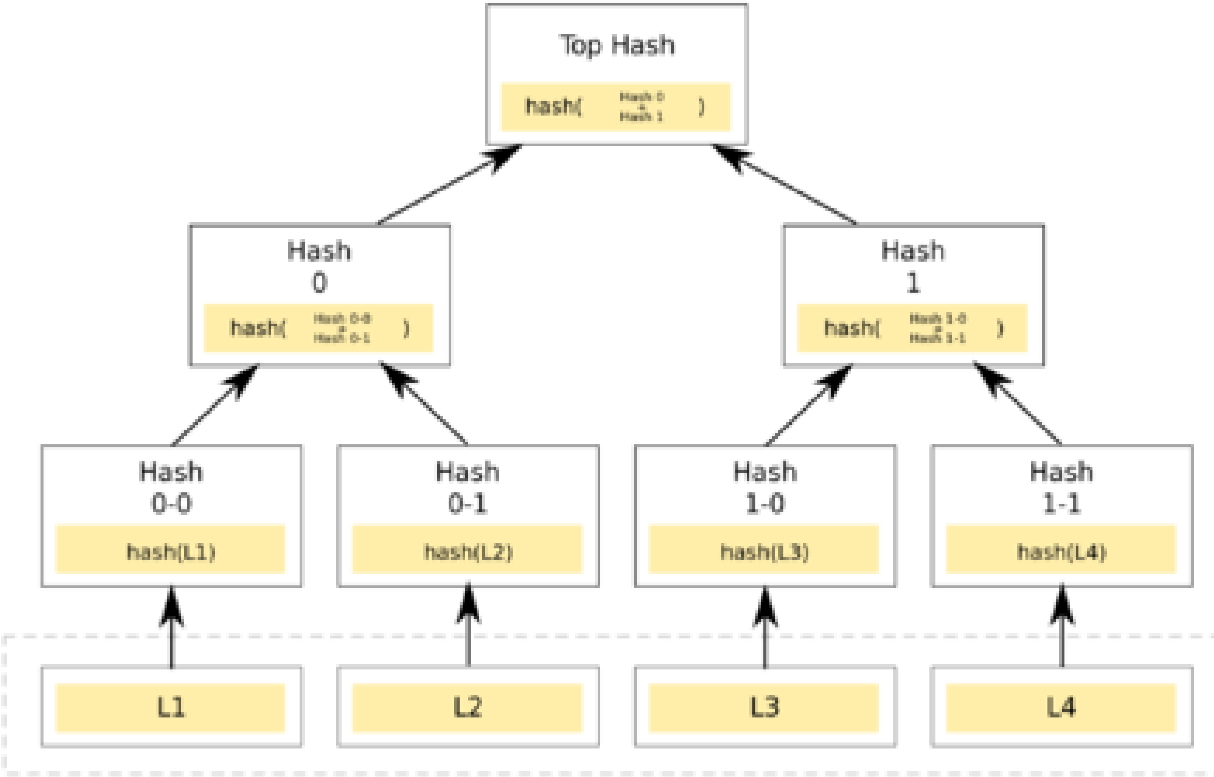


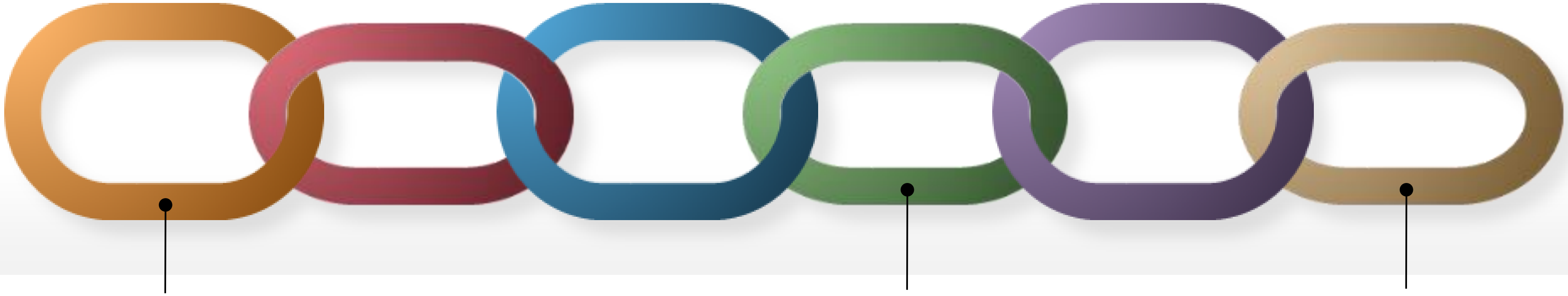
Bitcoin Blockchain

can we remove illicit data? we all know that Bitcoin is secure because the history is untouchable!

let’s investigate (after all we are scientists)

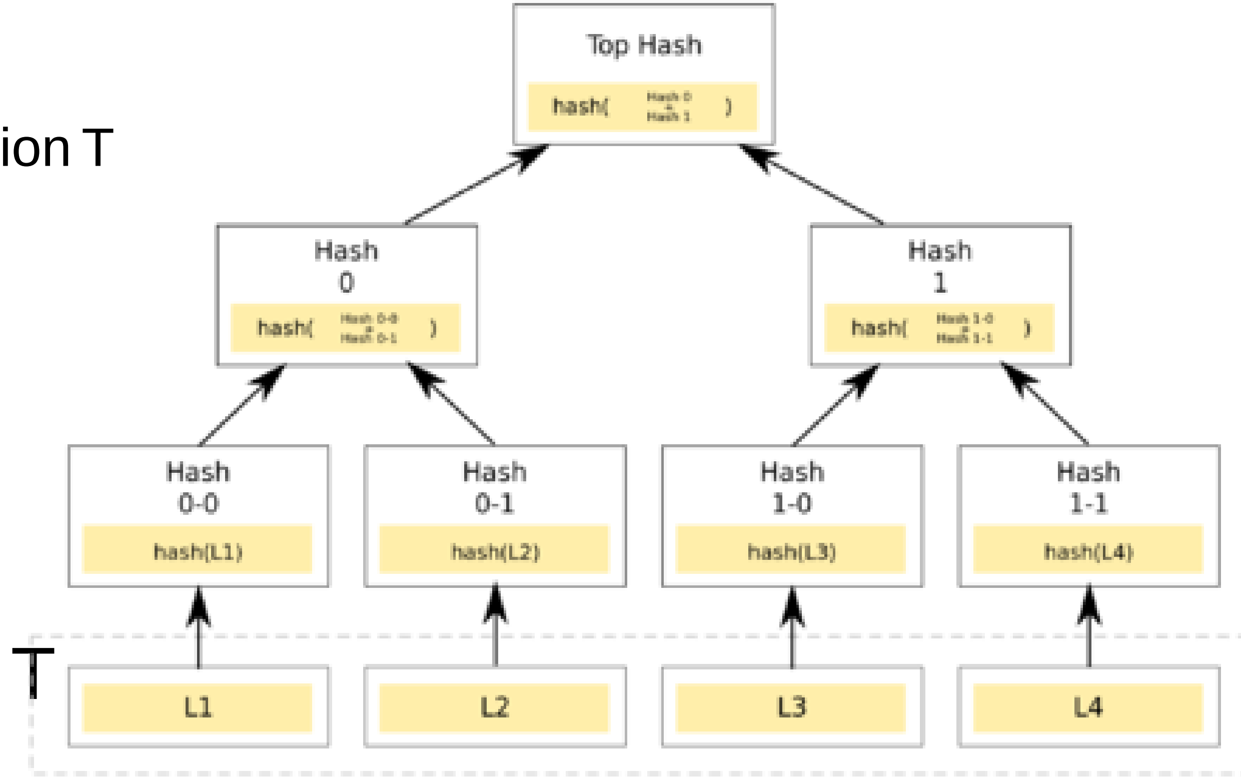
Removing Data from Bitcoin Blockchain: HowTo





Removing Data from Bitcoin Blockchain: HowTo

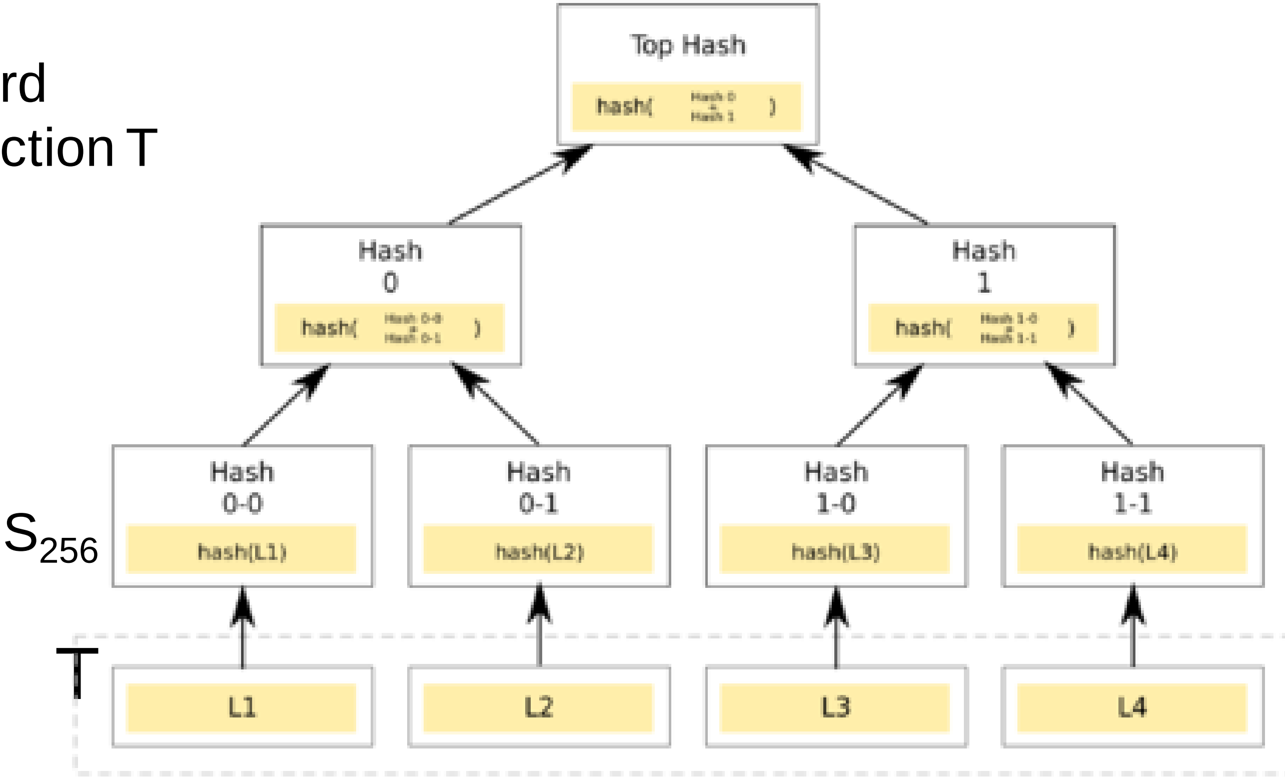
- notice that illicit data are stored after the keyword OP_RETURN of a Bitcoin script inside a transaction T



Removing Data from Bitcoin Blockchain: HowTo

- notice that illicit data are stored after the keyword OP_RETURN of a Bitcoin script inside a transaction T

- the SHA256 S_{256} of T is a leaf of a Merkle tree;

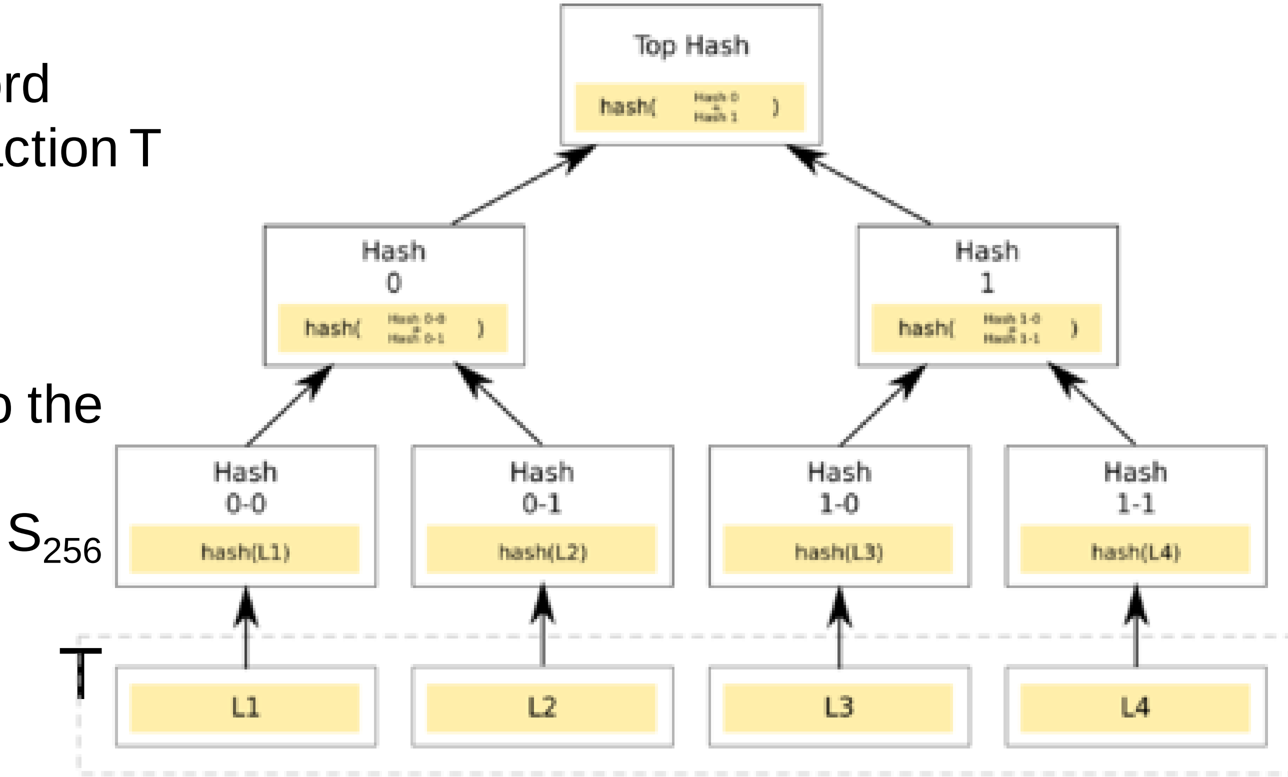


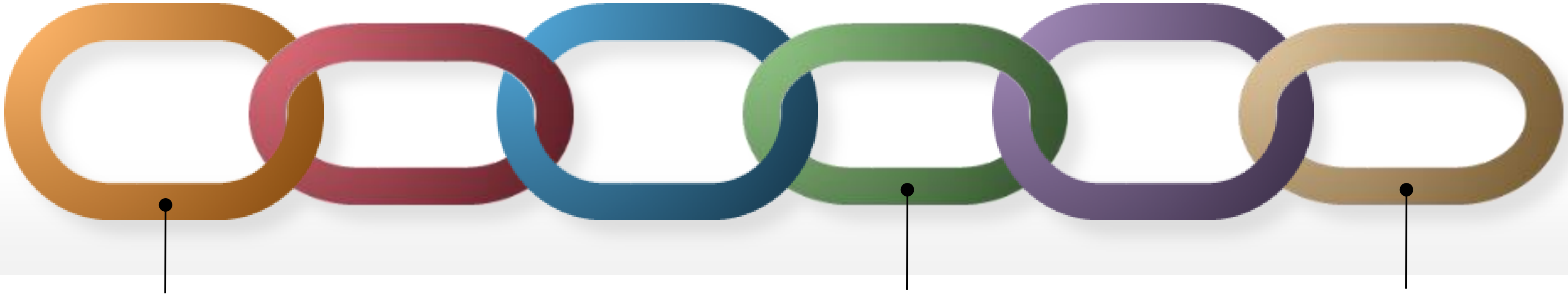
Removing Data from Bitcoin Blockchain: HowTo

- notice that illicit data are stored after the keyword OP_RETURN of a Bitcoin script inside a transaction T

- the SHA256 S_{256} of T is a leaf of a Merkle tree;

- if we update T then the update will propagate to the root and to the next blocks





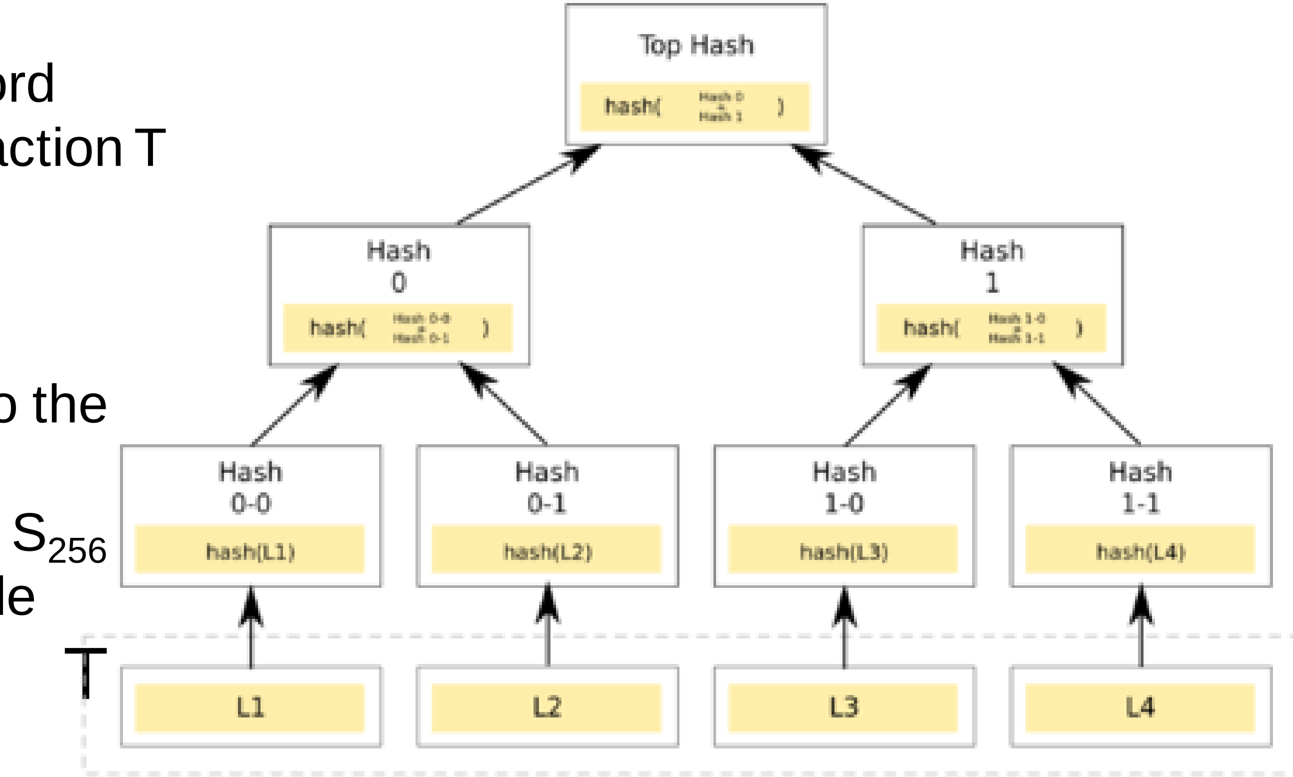
Removing Data from Bitcoin Blockchain: HowTo

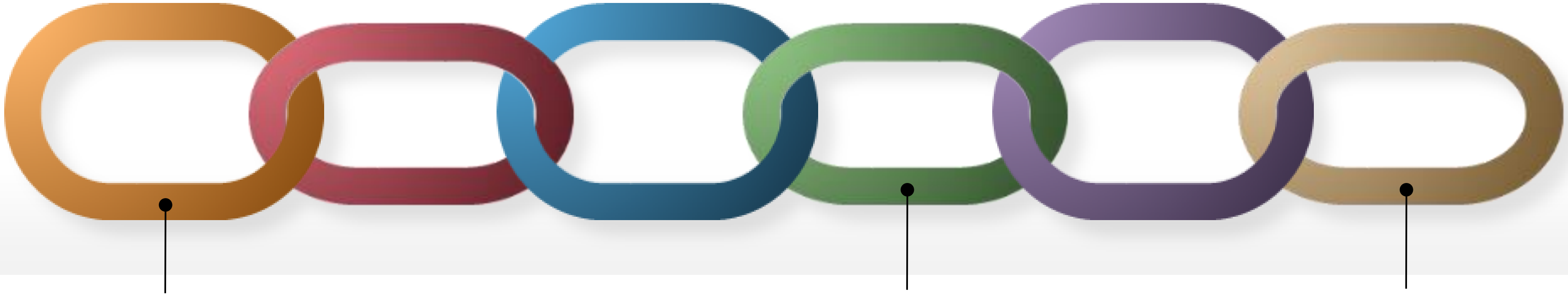
- notice that illicit data are stored after the keyword OP_RETURN of a Bitcoin script inside a transaction T

- the SHA256 S_{256} of T is a leaf of a Merkle tree;

- if we update T then the update will propagate to the root and to the next blocks

- if we remove T then double spending is possible

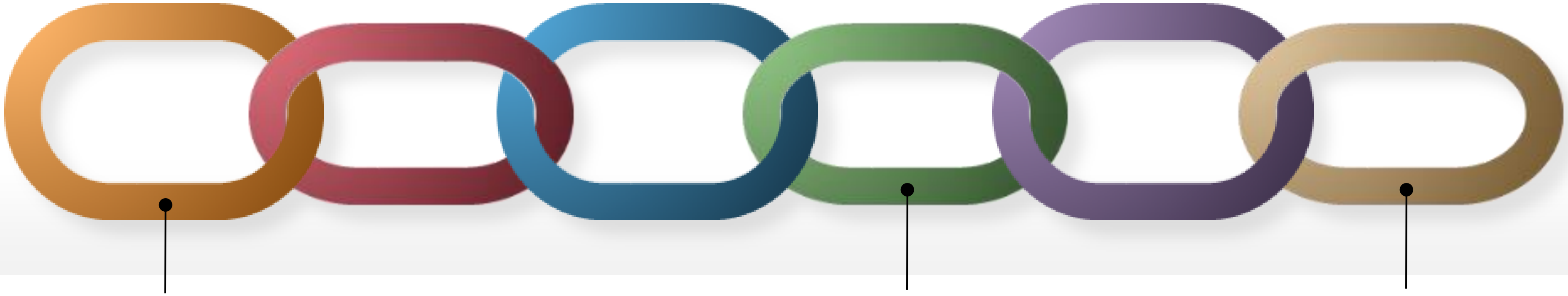




Removing Data from Bitcoin Blockchain

what if privacy regulations require to remove data?

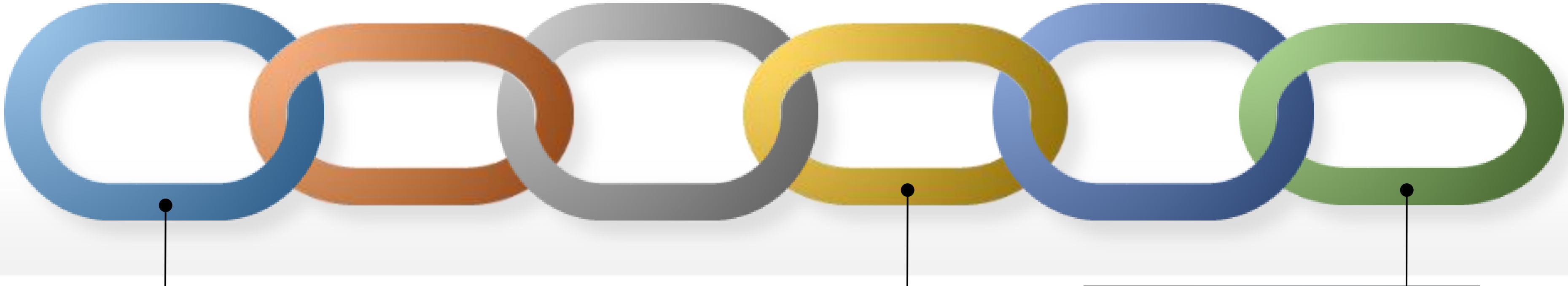
are we really in trouble?



Star Wars: The Force Unleashed developer LucasArts publisher LucasArts
© GameWallpapers.com hosted by JTLnet.com

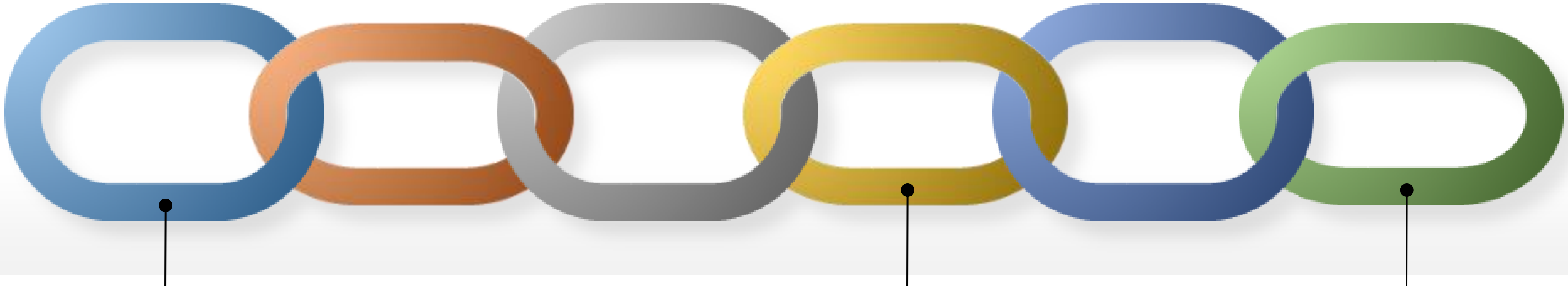
Zero-Knowledge Proofs [GMR85]

Prove that something is true without revealing any other information



ZK Proofs

it’s not just theory, they are very efficient for several useful claims



ZK Proofs

it’s not just theory, they are very efficient for several useful claims

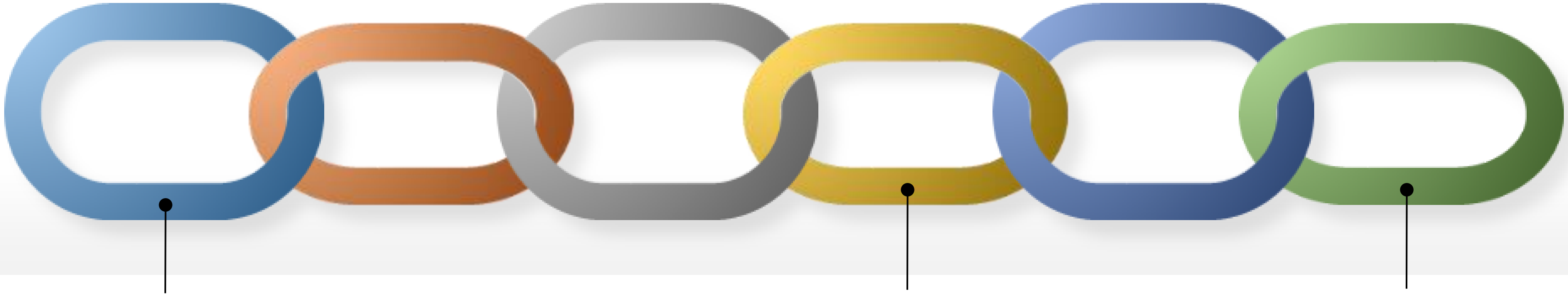
they can even be **non-interactive** (i.e., NIZK) and succinct (i.e., small length for proving a claim about the entire blockchain!);

ZK Proofs

it’s not just theory, they are very efficient for several useful claims

they can even be **non-interactive** (i.e., NIZK) and succinct (i.e., small length for proving a claim about the entire blockchain!);

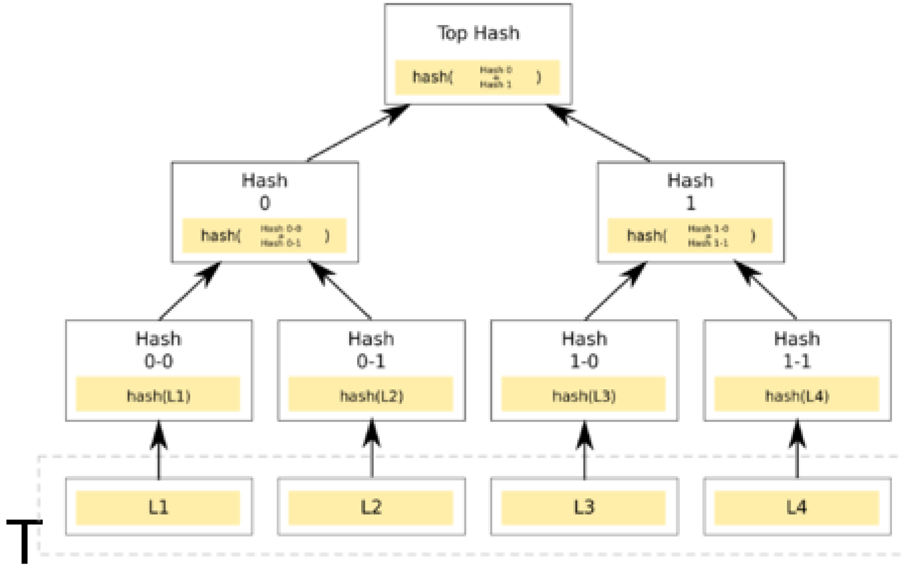
some names getting popular: zk-SNARKS (used in ZCash), zk-STARK….make sure you don’t play too much with fire…

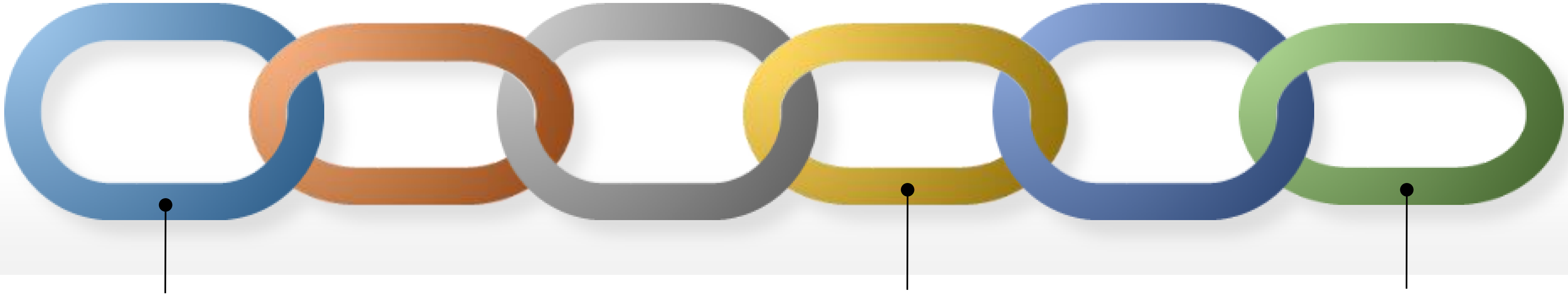


Removing Data from Bitcoin Blockchain: HowTo

Solution:

- replace illicit data in T with ZEROes obtaining T’

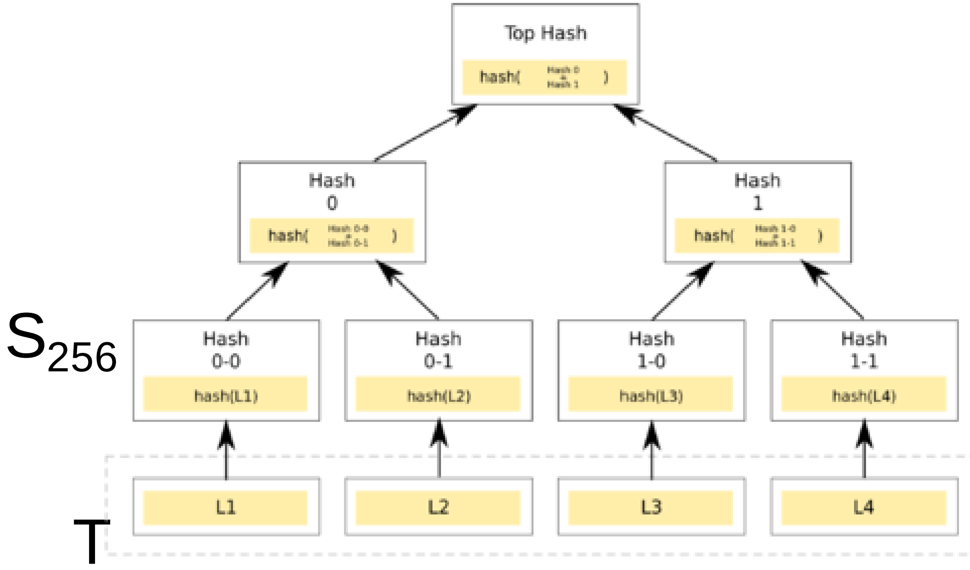


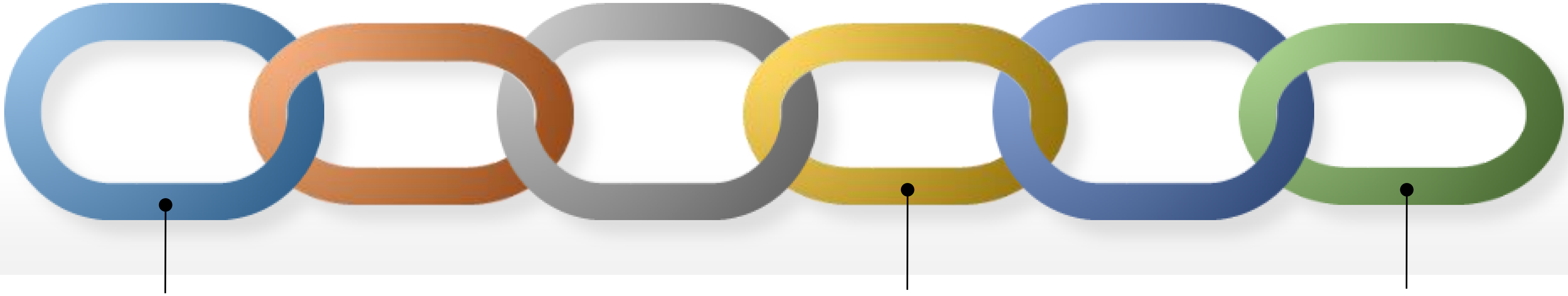


Removing Data from Bitcoin Blockchain: HowTo

Solution:

- replace illicit data in T with ZEROes obtaining T’
- compute a **NIZK proof** that there exists some bits that replaced in T’ after OP_RETURN would produce S_{256} as output of SHA256



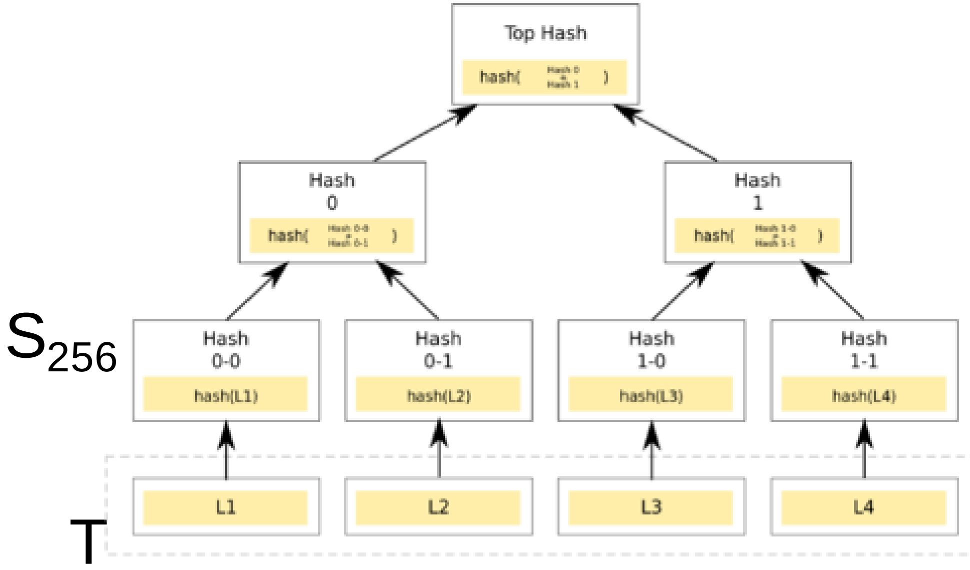


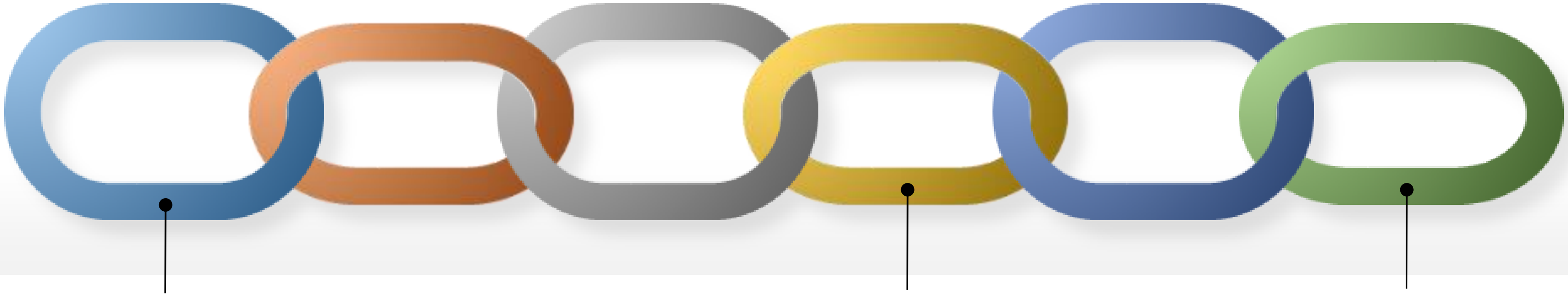
Removing Data from Bitcoin Blockchain: HowTo

Solution:

- replace illicit data in T with ZEROes obtaining T'
- compute a **NIZK proof** that there exists some bits that replaced in T' after OP_RETURN would produce S_{256} as output of SHA256
- essentially we cut out undesired values and apply a patch

[Vincenzo Botta, Vincenzo Iovino, Ivan Visconti 2020]





ZKProof for SHA256: is it efficient?

- Nakamoto did not have ZK proofs in mind… SHA256 is not ZK friendly

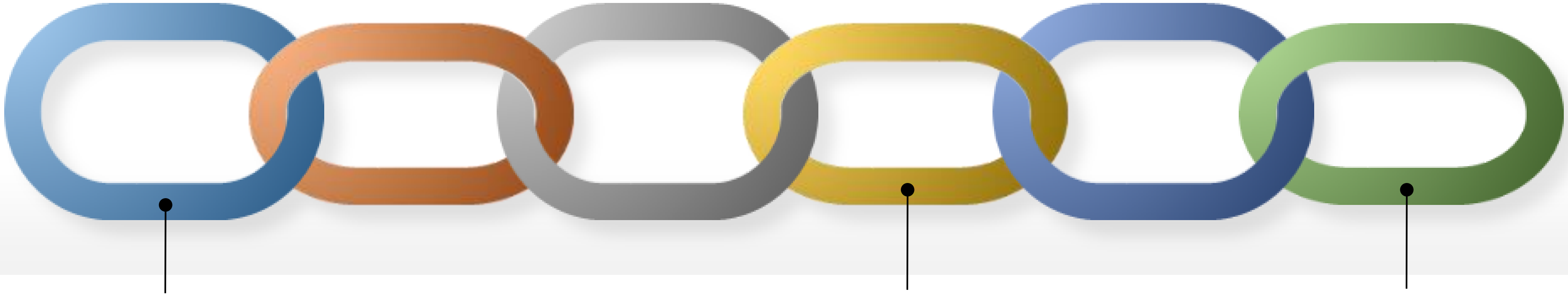
- Nevertheless, computing an efficient ZK proof for SHA256 is doable

- If you resort to ZCash technology (ZK-SNARKs), you end up with adding a trusted parameter (and in turn a possible trapdoor) to Bitcoin! That's not acceptable… Recall about playing with fire…

ZKProof for SHA256: is it efficient?

- ZK-SNARKs are very very succinct, but require (huge) trusted parameters

- a more suitable solution is to instead consider systems with a transparent setup, (**Ligero/Aurora/ZK-STARK)**, only very succinct but the underlying security is based on assumptions already used in Bitcoin (random oracle model)



Bitcoin Blockchain

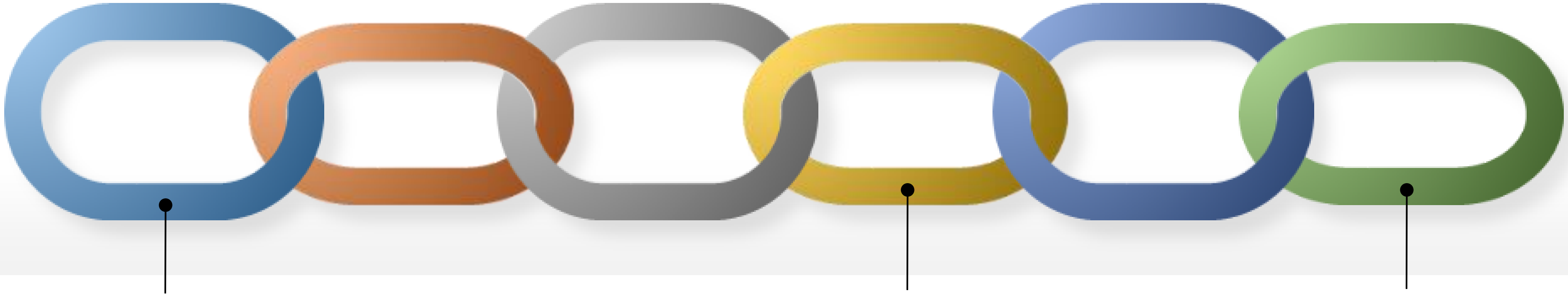
can we remove illicit data?

yes, we can sanitize completely the Bitcoin Blockchain removing data arbitrarily added after OP_RETURN and COINBASE transactions

we can pull out the famous sentence “The Times 03/Jan/2009 Chancellor on brink of second bailout for banks.” from the genesis block

Still not convinced?

Have a look at Coda



Still not convinced?

Have a look at Coda

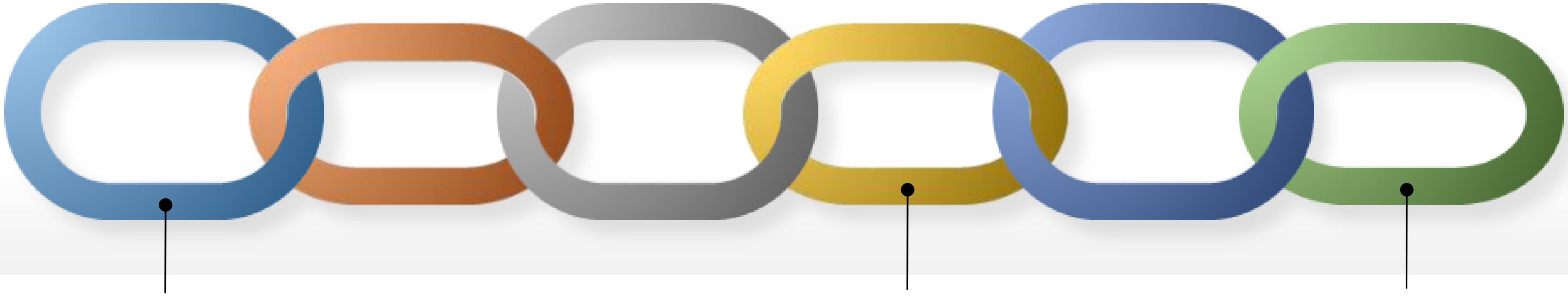
“Coda is a cryptocurrency (built by O(1) Labs) which has a succinct blockchain, meaning users can sync with the network by obtaining a constant amount of data and performing a constant amount of computation”

Still not convinced?

Have a look at Coda

“Coda is a cryptocurrency (built by O(1) Labs) which has a succinct blockchain, meaning users can sync with the network by obtaining a constant amount of data and performing a constant amount of computation”

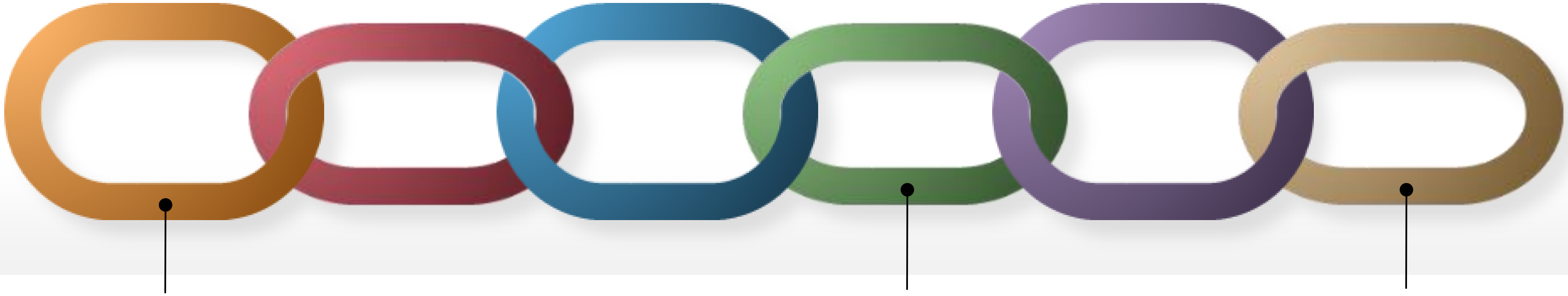
They use even more powerful tools (i.e., recursive ZK-SNARKs), but on the other hand they designed their blockchain with this goal in mind.



Recall that

if you want to use a Blockchain for applications involving confidential data then:

- keep in mind that there is no much trust around
- keep in mind that it must be efficient
- keep in mind that it is must be secure
- keep in mind that you must obey to regulations

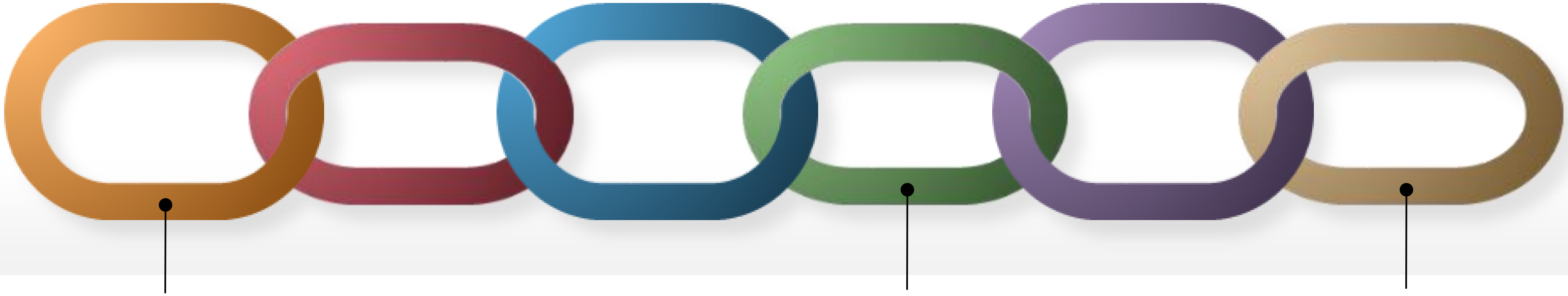


Challenge number 2: private data on a blockchain

trivial approach: do nothing

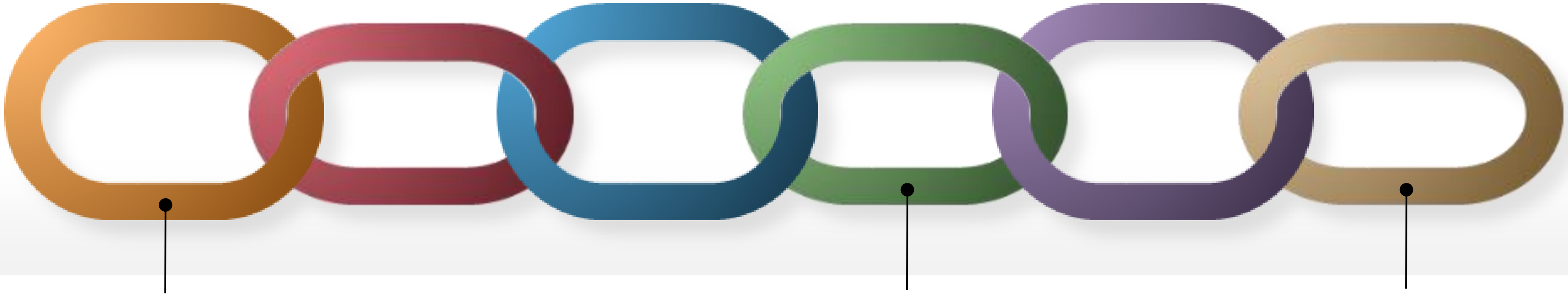
certainly you maintain security and privacy…

but obviously we would like to realize interesting applications and doing nothing does not help



Challenge number 2: private data on a blockchain

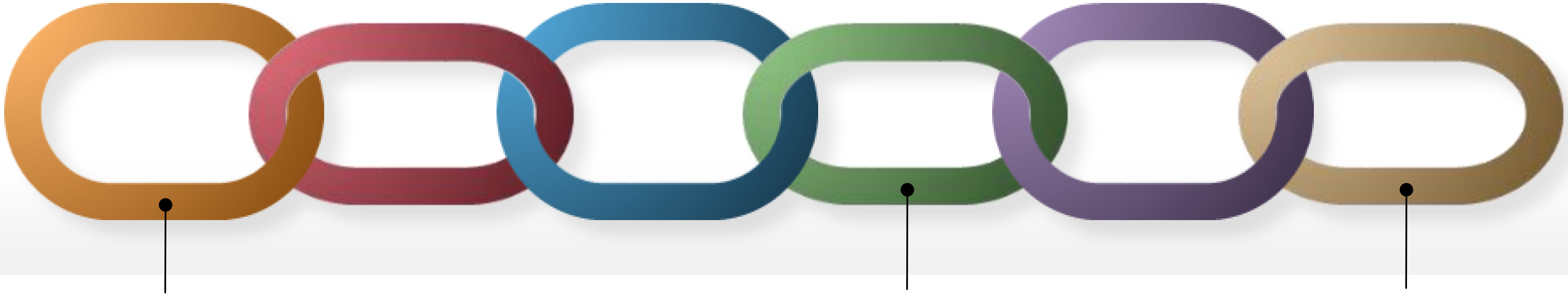
Folklore approach (Blockchain for notarization):



Challenge number 2: private data on a blockchain

Folklore approach (Blockchain for notarization):

The notary service takes your confidential value m, computes s=SHA256(m) and stores s on the Blockchain in block j

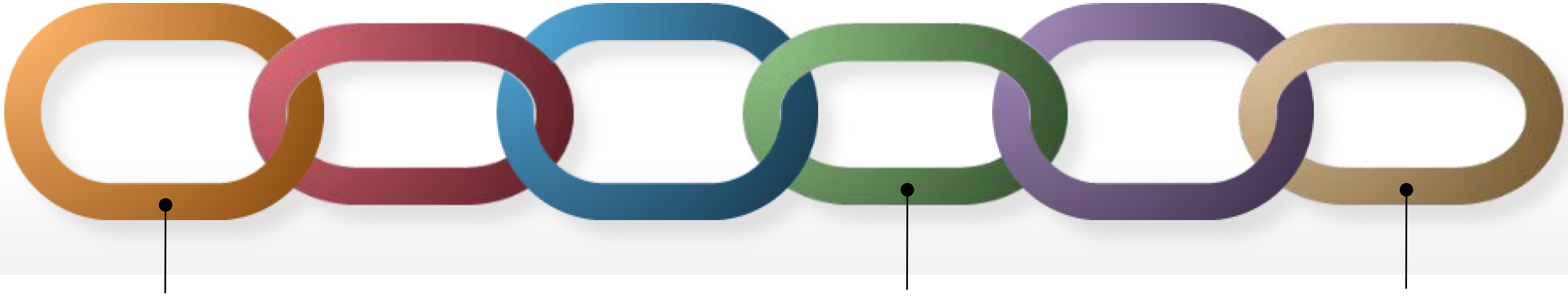


Challenge number 2: private data on a blockchain

Folklore approach (Blockchain for notarization):

The notary service takes your confidential value m, computes s=SHA256(m) and stores s on the Blockchain in block j

Later on, you can send (m,s,j) to anyone you would like to allow verification of the notarized m



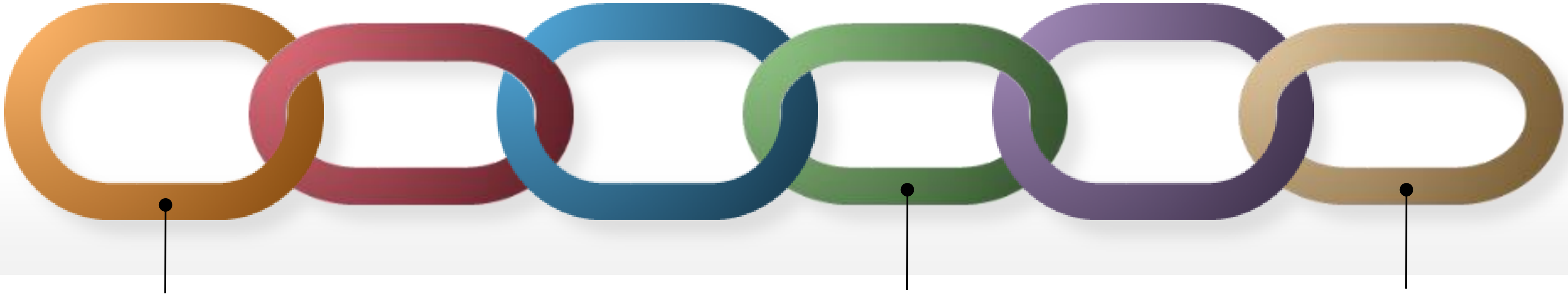
Challenge number 2: private data on a blockchain

WRONG!!!

Folklore approach (Blockchain for notarization):

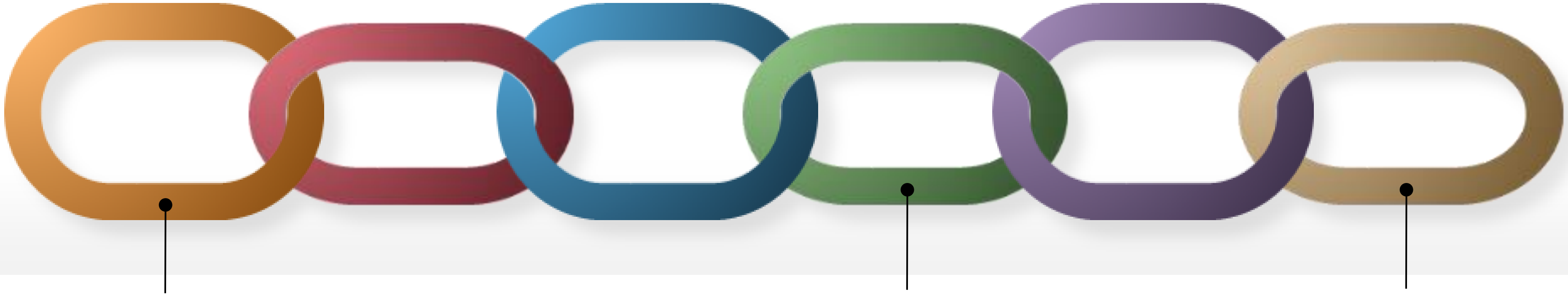
The notary service takes your confidential value m, computes s=SHA256(m) and stores s on the Blockchain in block j

Later on, you can send (m,s,j) to anyone you would like to allow verification of the notarized m



Challenge number 2: private data on a blockchain

1st of all: if m is not unpredictable
(more technically, it is not a string with high min-entropy)
anyone reading s on the blockchain can brute force SHA256 quickly finding m (i.e., finding m such that s=SHA256(m) is easy if the set of possible values for m in the context in which it is used is small)



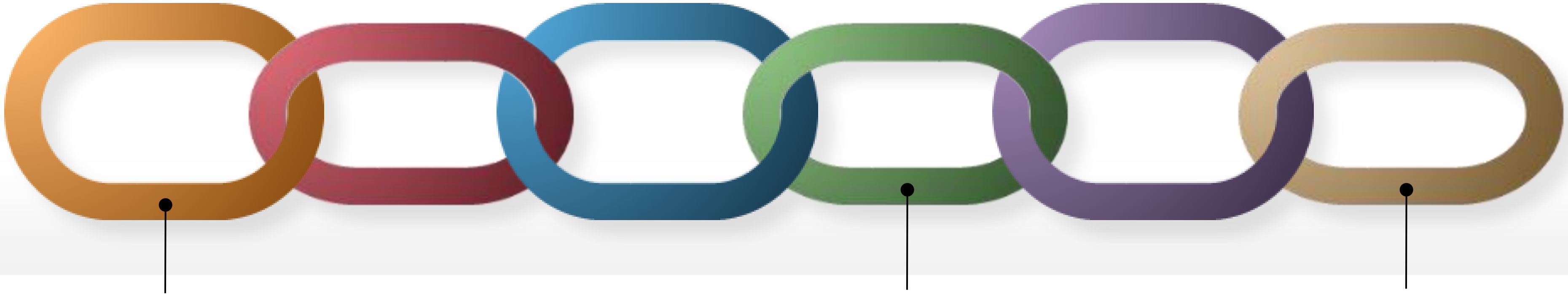
Challenge number 2: private data on a blockchain

1st of all: if m is not unpredictable
(more technically, it is not a string with high min-entropy)
anyone reading s on the blockchain can brute force SHA256 quickly finding m (i.e., finding m such that s=SHA256(m) is easy if the set of possible values for m in the context in which it is used is small)

you’re probably looking for a “commitment scheme”;
it’s a fully understood cryptographic primitive admitting plenty of constructions and implementations… don’t do homemade cryptography

Challenge number 2: private data on a blockchain

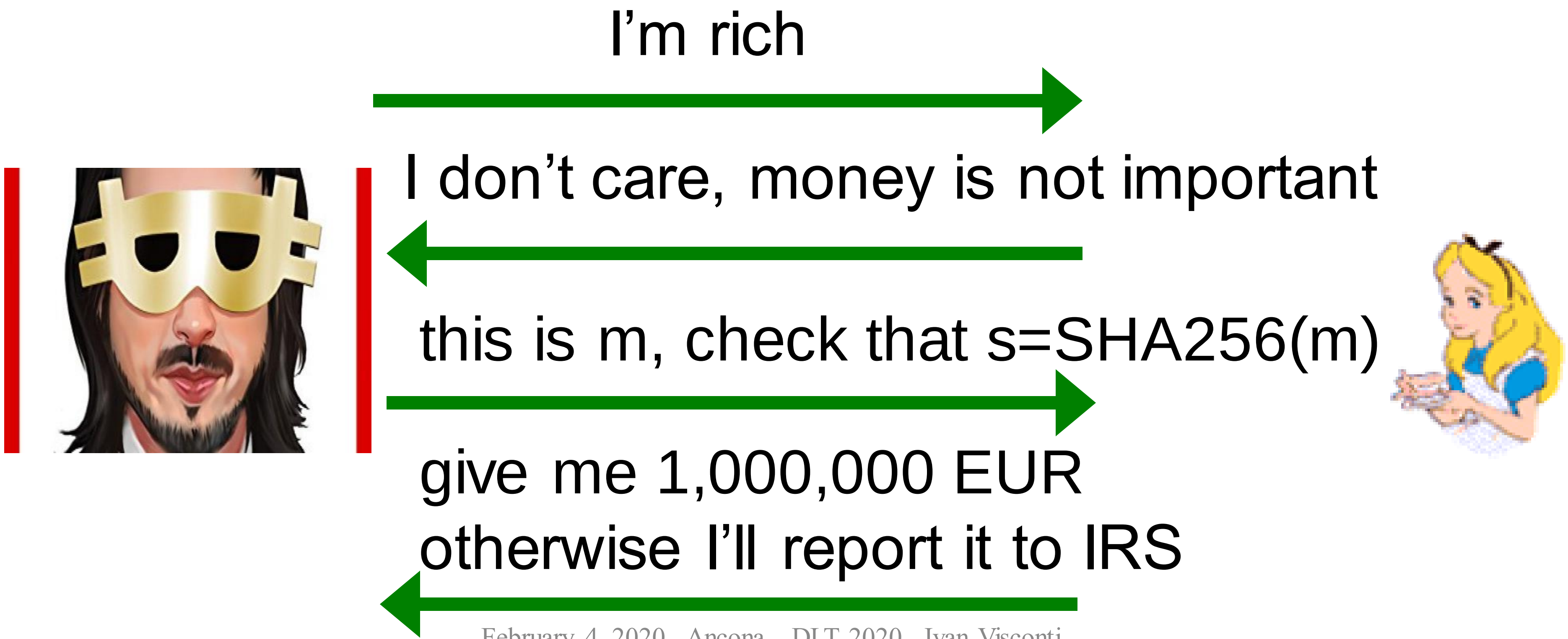
let’s ignore the previous problem and let’s assume that you “commit” to m properly with SHA256
there is even more…



Challenge number 2: private data on a blockchain

let’s ignore the previous problem and let’s assume that you “commit” to m properly with SHA256
there is even more…
when later on you send (m,s,j) to give evidence of the notarized m, you’re giving away the confidentiality of m to everyone

the message hashed in s in block j of the blockchain by Bank of Cayman Islands is a bank statement with a balance of 1,000,000,000 EUR



I’m rich

I don’t care, money is not important

this is m, check that s=SHA256(m)

give me 1,000,000 EUR otherwise I’ll report it to IRS

the message hashed in s in block j of the blockchain by Bank of Cayman Islands is a bank statement with a balance of 1,000,000,000 EUR

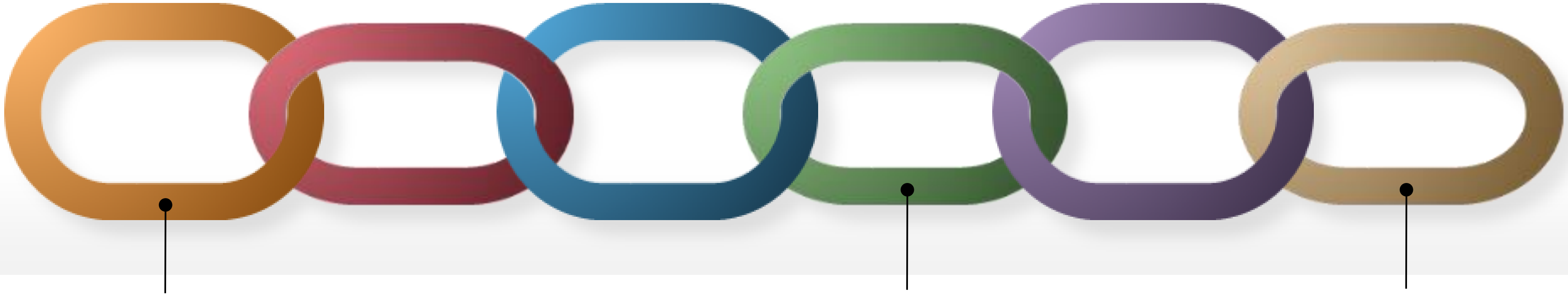
WRONG!!!

I'm rich

I don't care, money is not important

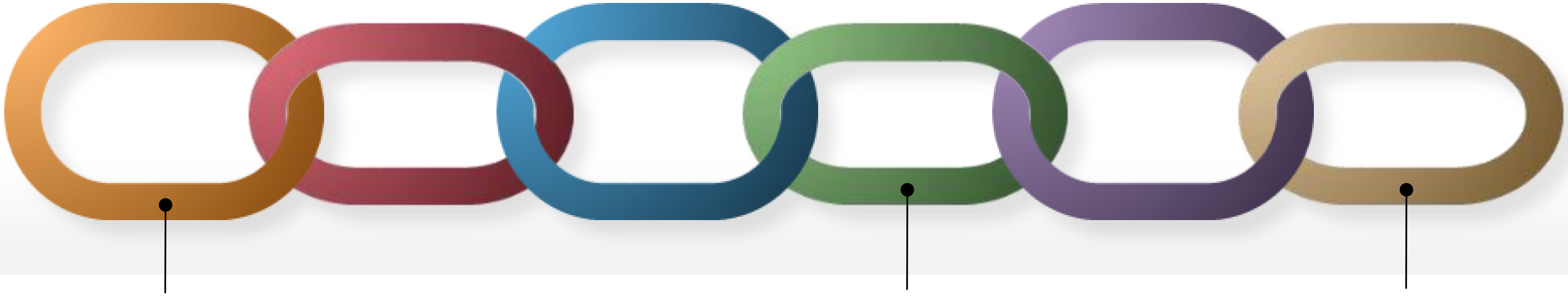
this is m, check that s=SHA256(m)

give me 1,000,000 EUR
otherwise I'll report it to IRS



Challenge number 2: private data on a blockchain

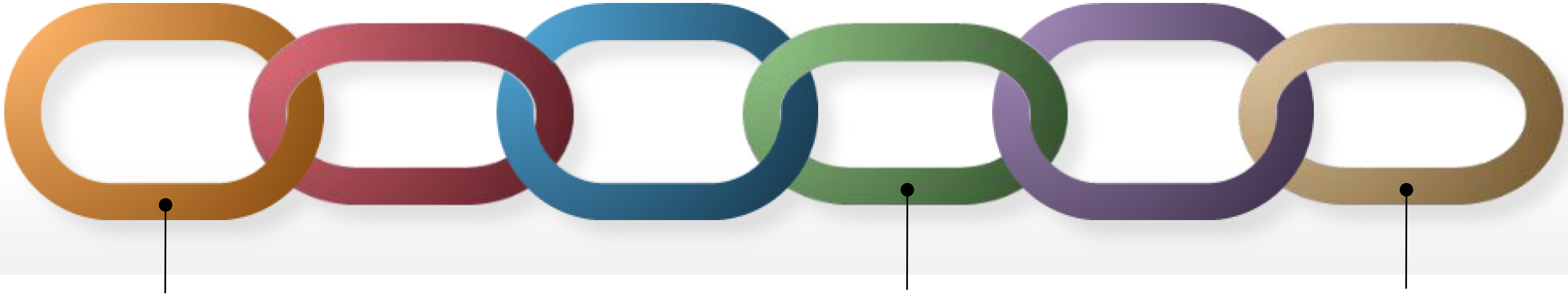
the lesson is that one should be extremely careful even with the trivial use of a Blockchain to notarize private information



Challenge number 2: private data on a blockchain

the lesson is that one should be extremely careful even with the trivial use of a Blockchain to notarize private information

this simple task is actually a cryptographic protocol and proving the correctness requires experience

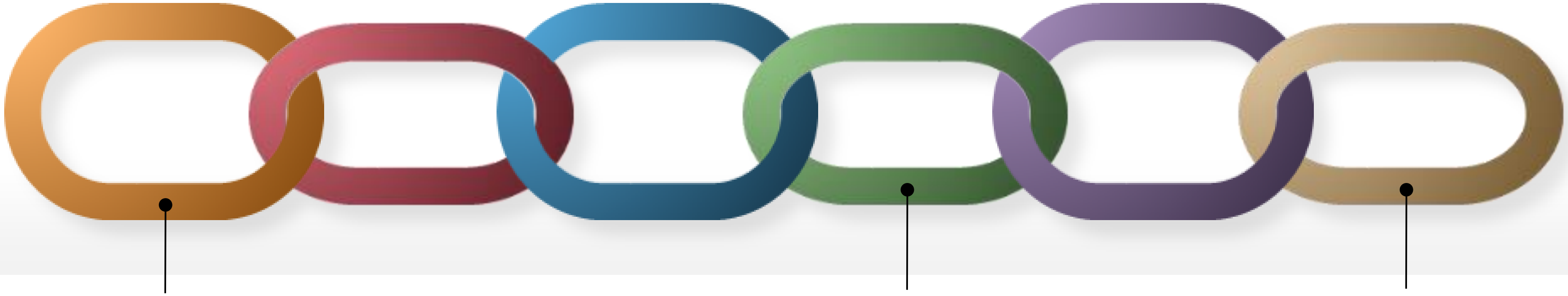


Challenge number 2: private data on a blockchain

the lesson is that one should be extremely careful even with the trivial use of a Blockchain to notarize private information

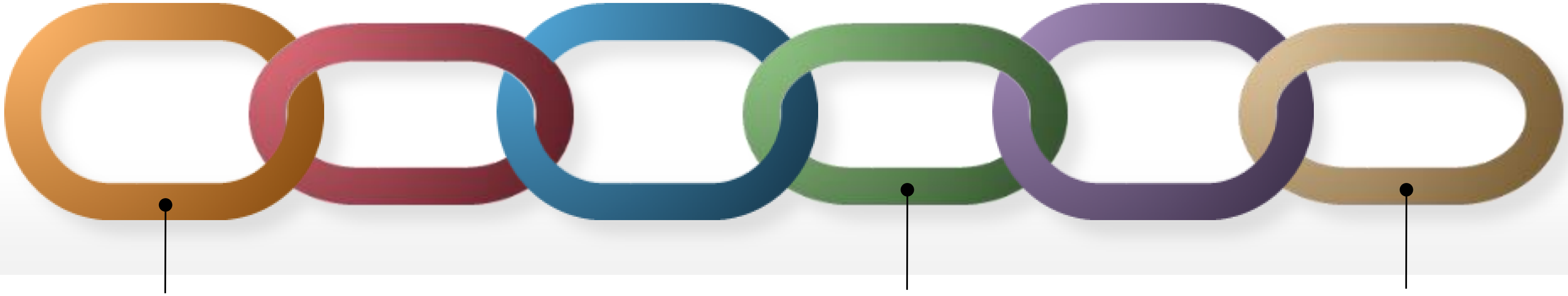
this simple task is actually a cryptographic protocol and proving the correctness requires experience

security and privacy by design is beautiful, but it should not be just claimed, it should be rigorously proven



Challenge number 2: private data on a blockchain

guess what? we might fix it using zero-knowledge proofs

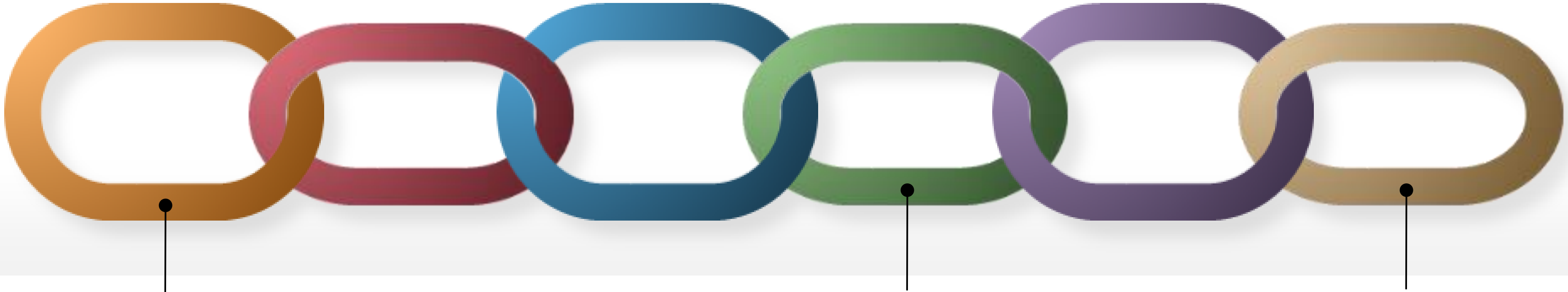


Challenge number 2: private data on a blockchain

The solution:

the notary service uses a ZK-friendly commitment scheme to encode the message m to be notarized

When giving evidence of the encoded data use a ***“Deniable ZK proof”*** (i.e., a proof that convinces a specific verifier but that can not be transferred to convince others).



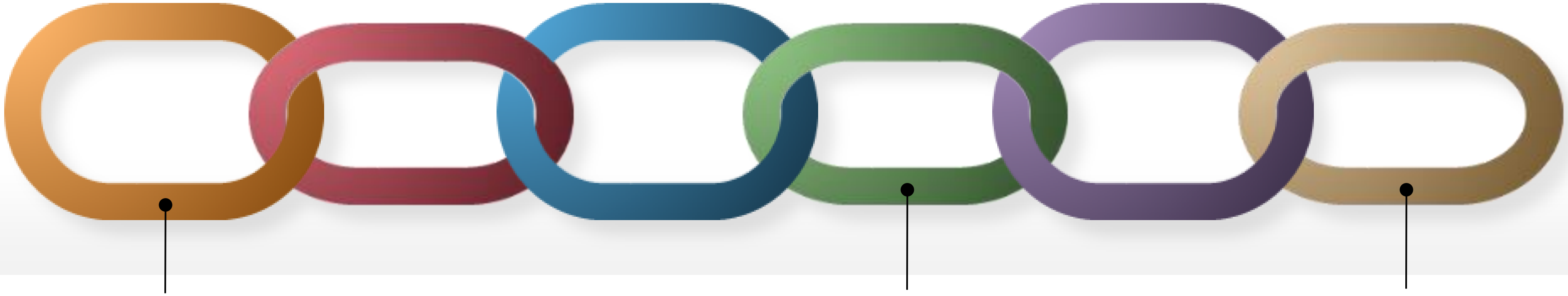
the message hashed in s in block j of the blockchain by Bank of Cayman Islands is a bank statement with a balance of 1,000,000,000 EUR

I'm rich

I don't care, money is not important

NIZK proving the claim OR knowledge of Alice's secret

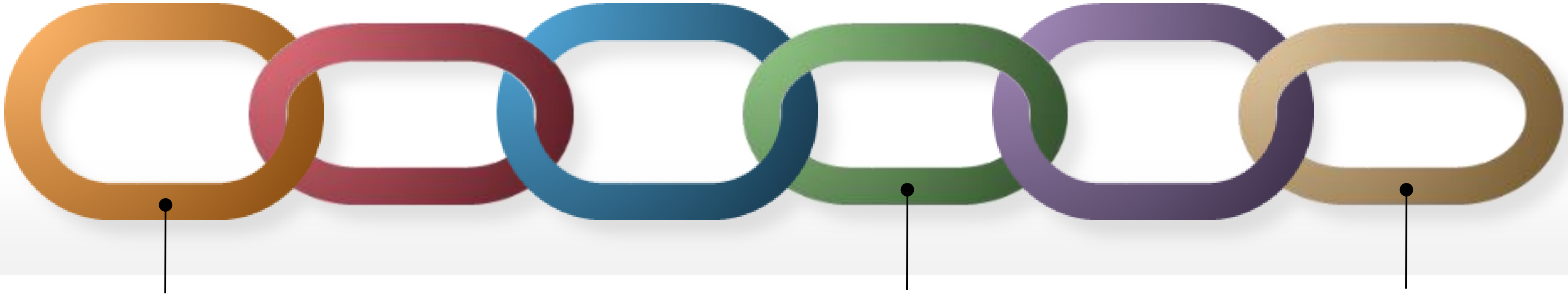
Towards Secure E-Voting with Everlasting Privacy
Avitabile, Heiberg, Lipmaa, Siim and Visconti, tomorrow 11:45am



Challenge number 2: private data on a blockchain

what if the computation involves private data of multiple players?

ZK proofs seem to help only when one player owns a secret

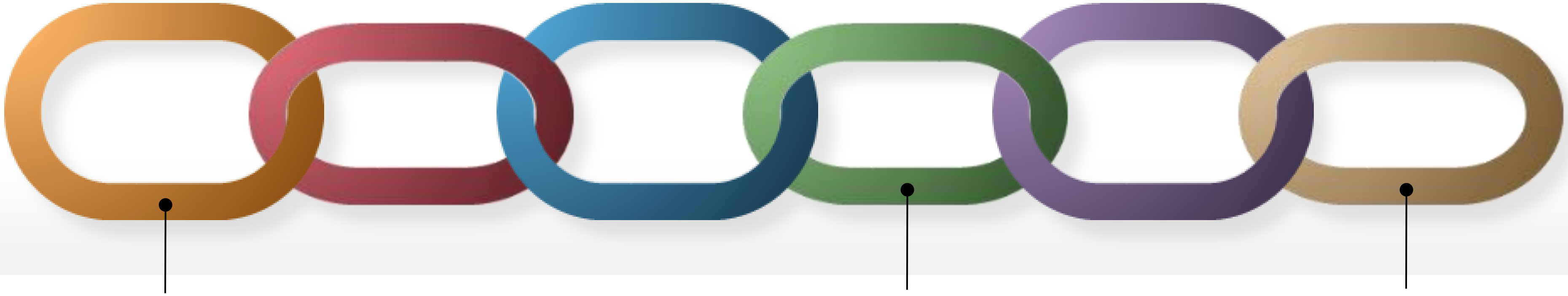


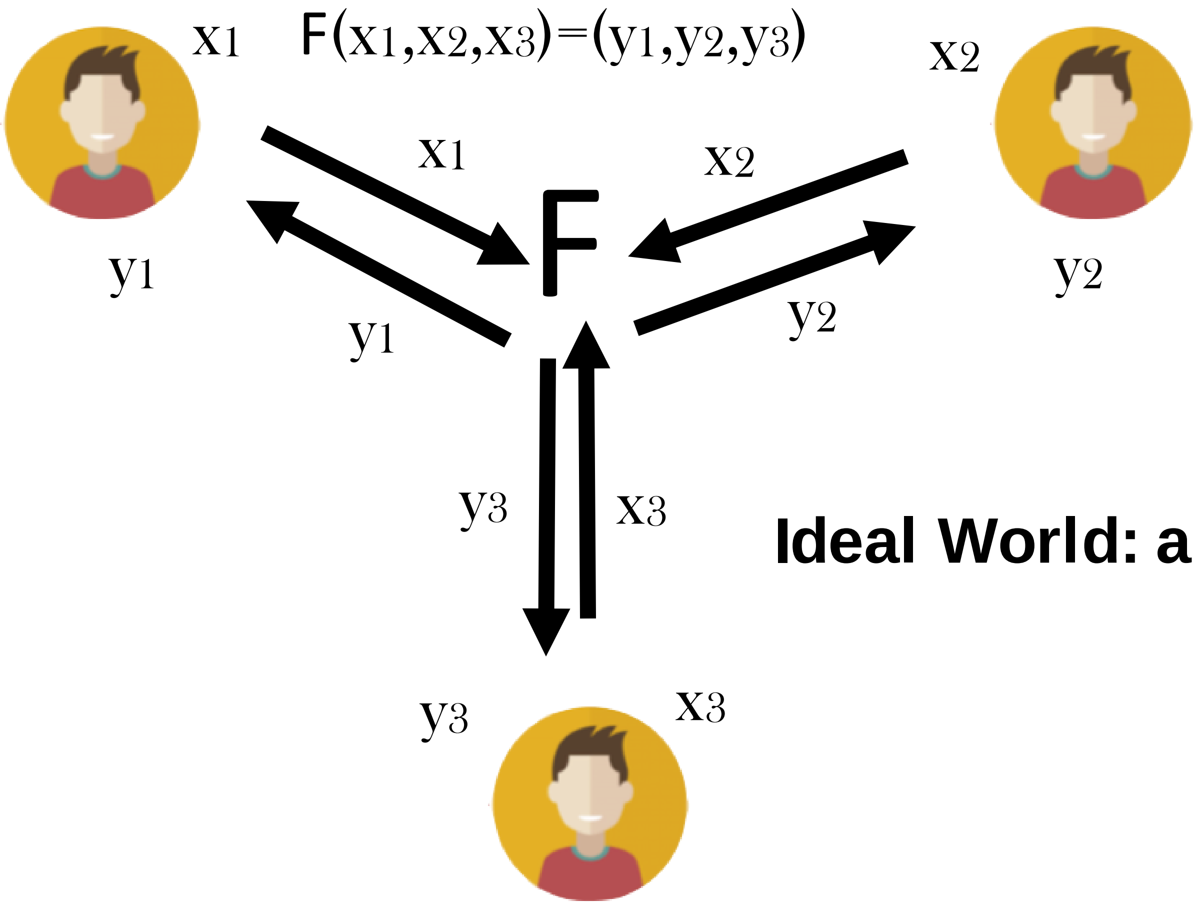
Challenge number 3:
joint computation over private data on a blockchain

what if the computation involves private data of multiple players?
(e.g., “The Danish Sugar Beet Auction”)

ZK proofs seem to help only when one player owns a secret

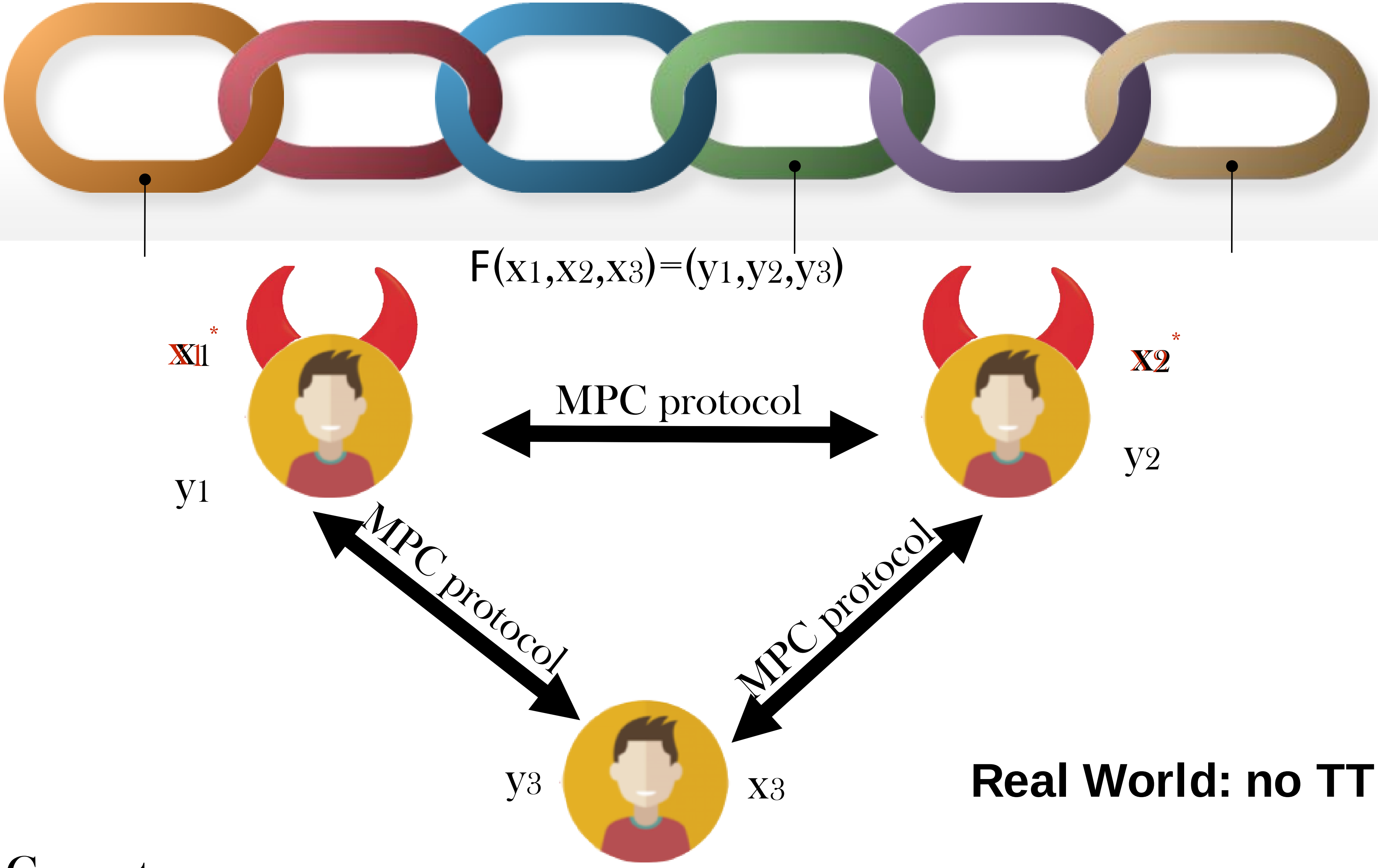
the natural next step consists of using protocols for
“secure multi-party computation”





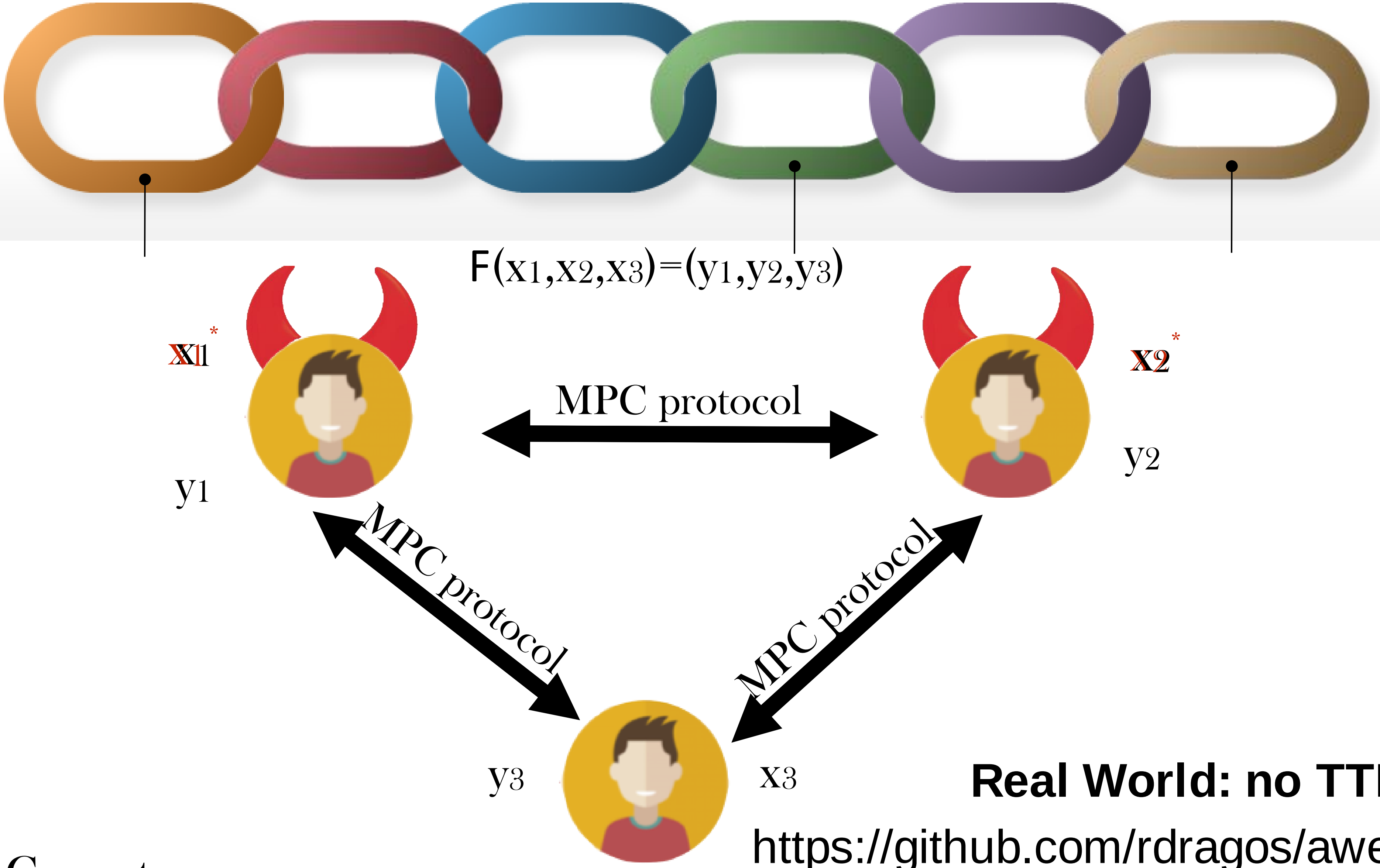
$F(x_1,x_2,x_3)=(y_1,y_2,y_3)$

Ideal World: a TTP working for us



Real World: no TTP

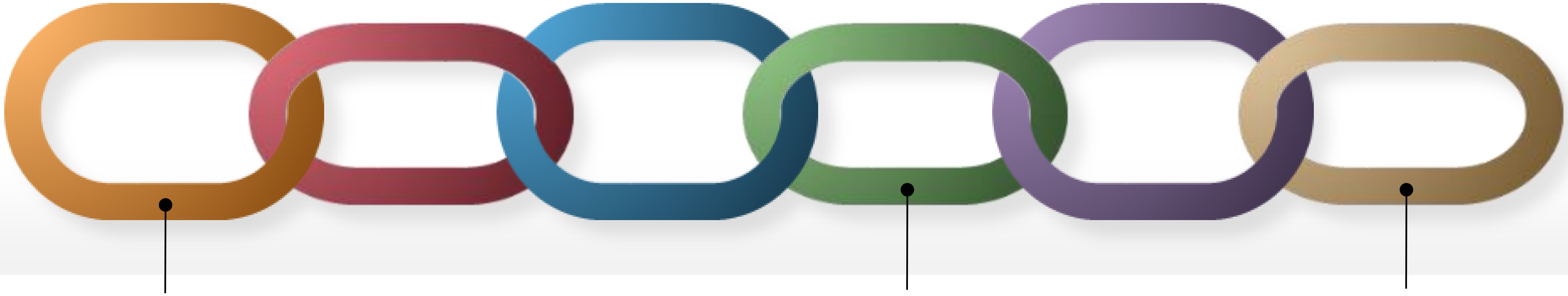
- Correctness
- Security (input-output privacy is preserved)



Real World: no TTP

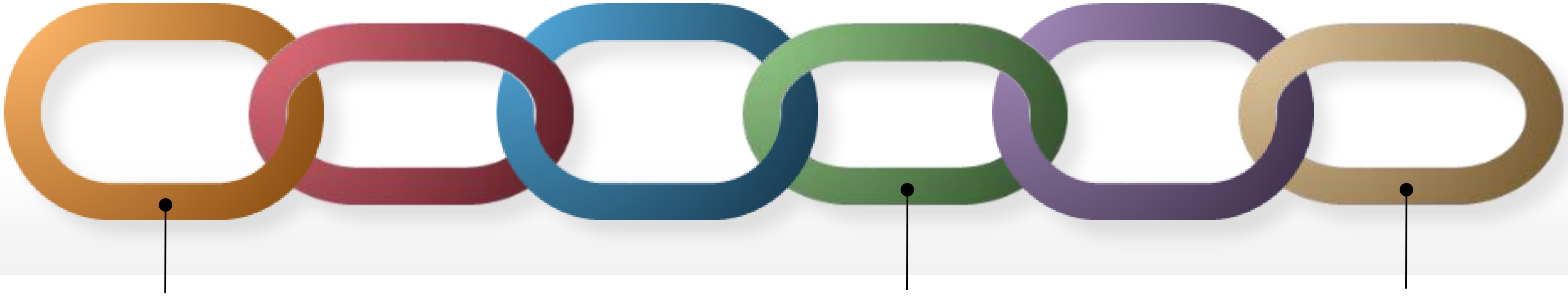
https://github.com/rdragos/awesome-mpc

- Correctness
- Security (input-output privacy is preserved)



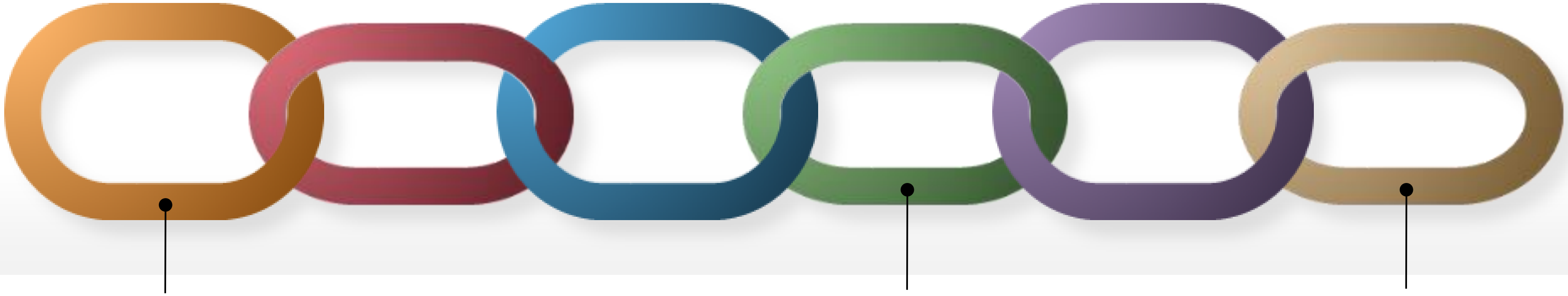
Conclusion

Blockchain technology is a powerful tool against counterfeiting, it allows to relax the need of trusted third parties and can have a strong impact on our societies.



Conclusion

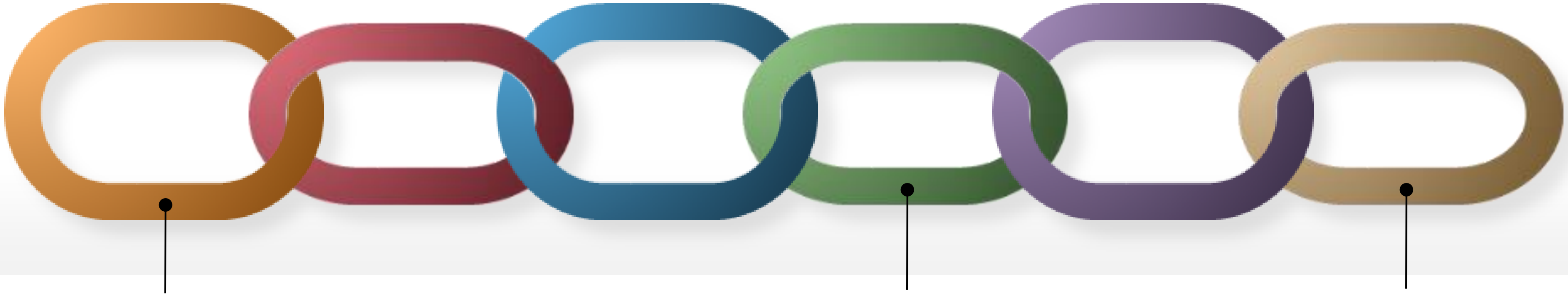
There are natural issues related to confidential data but often we can address them using advanced cryptography.



Conclusion

There are natural issues related to confidential data but often we can address them using advanced cryptography.

Don’t’ play with fire…



Thank you for your attention!

visconti@unisa.it