MODULE

Data Protection

Overview

Focus Areas

1. Protecting commercially sensitive data
2. Technology solutions for data protection
3. Case study: a manufacturing conglomerate

Tools and Resources

4. Data protection strategies

Overview

The perceived loss of control over data is one of the biggest obstacles to blockchain adoption that many supply-chain organisations face. With good project planning and communication, however, this issue can be greatly mitigated.

Blockchain technology never requires an organisation to reveal more data than it is comfortable with. On-chain data can also be encrypted so that it is only usable by permissioned parties. Thus, in the course of selecting and deploying a blockchain solution, a supply-chain organisation has real flexibility to ensure it addresses both its data protection and privacy concerns and those of other supply-chain partners.

Recommended reading - Inclusive Deployment of Blockchain for Supply Chains: Part 4 – Protecting Your Data[94]

FOCUS AREAS

1. Protecting commercially sensitive data

What are the top action items to consider for protecting the confidentiality of sensitive data shared on a blockchain network?

There are two fundamental questions to answer when building a data protection structure for a blockchain network:

- Which supply chain partners need to have access to which pieces of information?
- Who, external to the system, should have access to what information?

The typical requirements baseline among supply-chain organisations for sharing data include the following four dimensions as shown in Figure 8.1.

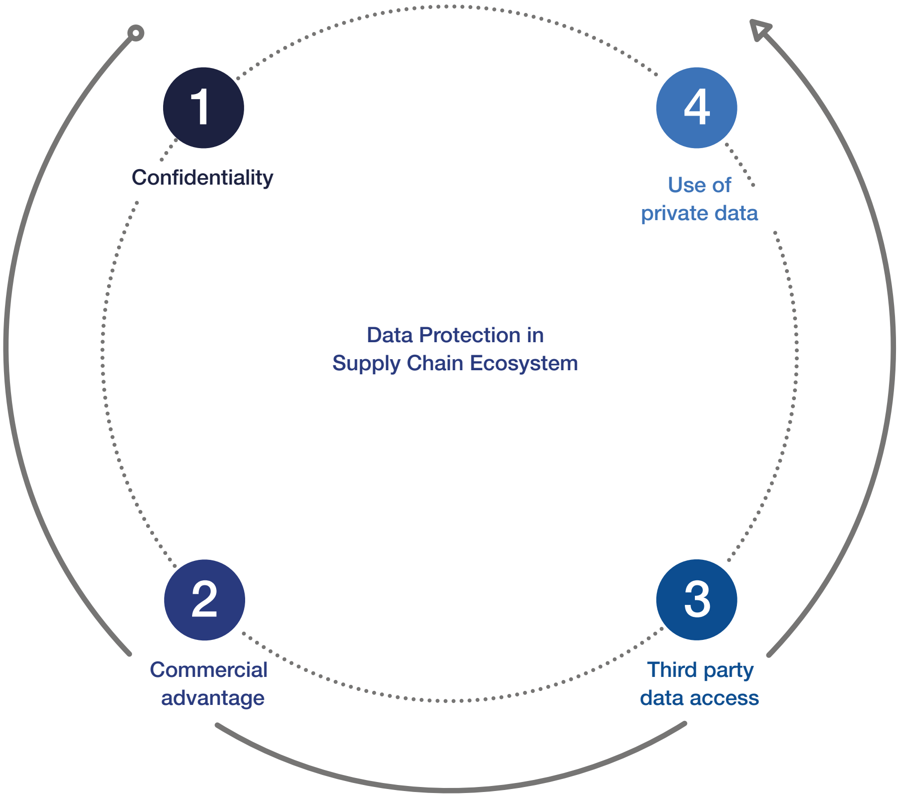

Figure 8.1 – Key points to investigate in protecting data

Confidentiality

At a fundamental level, transactions in a supply chain cannot be transparent to all participants in a blockchain network. Confidentiality requirements for some information must be maintained regardless of which permitted parties it is shared with.

To put it another way, supply-chain partners transacting with one another may be logging information onto the blockchain, but still need to keep the information from each other. Two reasons why that happens:

- They believe there is value to having the blockchain serve as a single source of truth for authenticated supply-chain data so that participants can extract the particular data they need.

- The practical challenges of understanding what should be obfuscated and what can be revealed during a one-to-one integration process are too immense.

Example: An electronics contract manufacturer (CM) provides vendor-managed inventory services to its buyer, a large electronics original equipment manufacturer (OEM). The CM would now like to obtain supply-chain finance on the blockchain, which will entail revealing to the OEM some of the CM's current financing costs without revealing other operating costs such as storage and insurance. The financier providing the capital will want to know all of this information and is willing to offer more competitive financing precisely because of this visibility. Thus the CM needs to be able to share information on a secure, need-to-know, and one-to-many basis with any counterparty – a good use case for blockchain. No traditional solution presents a practical way of meeting the CM's requirements.

Commercial advantage

Organisations want to use supply-chain data in forecasting and planning. At the same time, however, there may be a number of reasons why supply-chain partners are naturally resistant to providing the raw datapoints needed to conduct such analysis. For instance, they may feel they're not adequately compensated for the information, that it reveals strategically sensitive information about their own business, or that raw data they provide might be cross-referenced with other information to generate insights that could be used by competitors.

Example: When forecasting demand, buyers are incentivised to either inflate expected demand to ensure adequate supply or secure a volume discount. In anticipation of this, a supplier will therefore underproduce and "adjust" its reported inventory depending on whether it needs to create scarcity or meet outsized demand. This cat-and-mouse game creates inefficiency in the supply chain as a whole. A blockchain solution could allow suppliers and buyers to take a more collaborative approach, reporting data more truthfully without giving away control entirely or compromising competitive advantage.

Third party data access

To illustrate the challenges in this area, let's say an organisation needs to use a critical piece of pricing information in a blockchain transaction, but that information cannot be known to certain parties with access to the chain.

This is an instance where value can be unlocked by hiding certain information from parties even when those parties need to use that information in a transaction. This case is slightly different from the one in which parties acknowledge that information treated as confidential in the status quo must preserve the same level of confidentiality after a blockchain network is put in place.

In this example, the information was not confidential when only two parties were involved. However, by bringing it into a blockchain solution, that data must now stay hidden to certain participants on the blockchain, even where such information might be integral to the activities of the blockchain.

Example: A commodities producer would like to get its inventory off its balance sheet as soon as possible and recognise revenue. It can sell this inventory to a trading company or third-party financier on the blockchain, who can then sell to the end buyer at the appropriate time. However, the sensitivity of commodities prices is such that, while all parties would benefit from this financing structure, it would be commercially unacceptable to the producers for the financier in the middle to know the actual price.

Use of private data

Organisations only need to verify information authenticated on the blockchain, but they have to do so without seeing the data itself.

This situation is similar to the one described above. Whereas that case can be solved with blockchain-enabled computation, this particular problem requires matching and verification of large volumes of data without ever revealing the information itself. Once hashed or encrypted, data must remain in this state even when functions are performed on it. Zero-knowledge proofs (ZKP), although still in research and development phases, can be a useful tool in this endeavour for certain computationally intensive cases. (Zero-knowledge proofs are covered at greater length in the module Personal Data Handling and focus area Technology approaches to GDPR compliance).

For further details on this subject, see focus area Intellectual property considerations in the module Consortium Governance, focus area Legal and regulatory risks in the module Risk Factors, and the module Legal and Regulatory Compliance.

> *For audit purposes, it is important when planning a new blockchain solution not to put requirements solely on the technology being used. Instead, it's better to co-design technology and business processes together to ensure sound audit results. For example, handling confidential data on a public blockchain may be technically resolved by encryption, but that is not sufficient from an audit perspective. There must also be a design for the audit process itself for things such as security level of encryption and healthy key management.*
>
> **Takayuki Suzuki,** Financial Information Systems Sales Management Division, Hitachi

2. Technology solutions for data protection

What are some of the current technologies that establish data protection on a blockchain supply chain?

Design options available for data protection

There is no single blockchain solution or set of solutions to solve for data protection needs. The solutions adopted depend on the technological capabilities of a particular blockchain platform and the specific privacy and performance factors that a supply chain is trying to optimise. There may also be contractual relationships to consider between the blockchain's users and network participants.

In most cases, a blockchain solution is built to be a core component within a larger system in which it works in conjunction with other technologies. For example, tools like application logic to implement access controls may be employed to supplement the capabilities of the blockchain itself. These additional technologies are required components of the technology stack in a blockchain solution to achieve the data protection or integrity often incorrectly assumed to be a core feature of blockchain.

Once an organisation determines there is confidential information that must be protected in a blockchain solution, there are several security design options.

Keep in mind that these technologies are best-in-class as of the writing of this toolkit. However, blockchain is a fast-developing space that may offer new technology options within a span of months.

On-chain/off-chain configurations and hashing	Role-based access controls (RBAC)	Zero-knowledge proof (ZKP)	Homomorphic encryption
Basic protections, such as on-chain/off- chain configurations, and only storing hashed data on the blockchain	Enable selective obfuscation of data depending upon the identity of a particular participant	Allows users to prove their knowledge of a value without revealing the value itself	An approach in which data is encrypted before being shared on-chain. It can then be analysed without decryption

Figure 8.2 – Major design options for data confidentiality of a blockchain solution

Each of these approaches has trade-offs with respect to sophistication, complexity, cost, and technology readiness.

The more complex the technology becomes, the more potential drawbacks in usability, including:

- Limited transaction speed
- Necessity of a trusted blockchain operator
- Higher transaction costs (in terms of computing power)
- Risk of irrelevant data being included in the payload, or any supplemental data area

The methods outlined above must be understood in the context of a broader blockchain solution architecture in order to effectively put them into practice. These architectures can include additional databases or storage mechanisms that communicate with the blockchain. See Figure 8.3 for different blockchain configurations for data confidentiality. Data may still be kept entirely on-chain or stored in an off-chain database. In the latter case, data is stored as a hash on the blockchain, and the raw information is securely placed in an off-chain database.

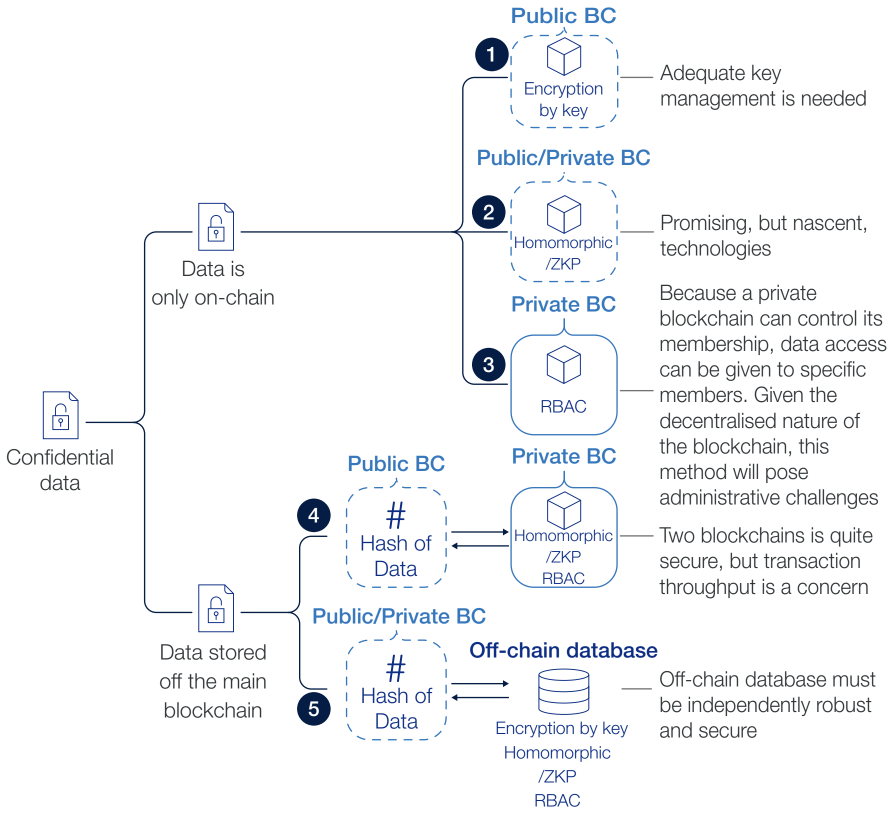
Public BC
1
Encryption by key
Adequate key management is needed
Public/Private BC
2
Homomorphic /ZKP
Promising, but nascent, technologies
Data is only on-chain
Private BC
3
RBAC
Because a private blockchain can control its membership, data access can be given to specific members. Given the decentralised nature of the blockchain, this method will pose administrative challenges
Confidential data
Public BC
4
# Hash of Data
Private BC
Homomorphic /ZKP RBAC
Two blockchains is quite secure, but transaction throughput is a concern
Data stored off the main blockchain
Public/Private BC
5
# Hash of Data
Off-chain database
Encryption by key Homomorphic /ZKP RBAC
Off-chain database must be independently robust and secure

Figure 8.3 – Blockchain configurations for data confidentiality

- **Option 1:** Public or private blockchain with encryption. When confidential information is stored in raw form on the blockchain, it should be encrypted. Decryption keys are then shared through another secure channel.
- **Option 2:** Public blockchain with information cryptographically hidden, but mathematically usable by itself via methods like ZKP and homomorphic encryption.
- **Option 3:** Private blockchain with necessary permissions and role-based access controls are sufficient to provide required confidentiality. Data can be recorded as raw data.
- **Option 4:** Private blockchain is paired with a public blockchain to store the raw data or documents while the public blockchain only stores hashes. The private blockchain is configured to provide required confidentiality.
- **Option 5:** An ordinary database is paired with a public or private blockchain to store the raw data or documents while the public blockchain only stores hashes. The database and microservices that publish from the blockchain to the database are configured to provide required confidentiality.

Considerations for each design option

- **Option 1:** Functionally realises both confidentiality and data utility. Its drawback is that the basic key generation function of a blockchain is insufficient to implement meaningful access controls. Rather, a solution must also consider secure key storage, status monitoring of key confidentiality, key revocation, and key deletion. To determine which keys can decrypt information shared on the blockchain over multiple types of data payloads, a network needs sophisticated key management and coordination among data policies that are often dictated by contract at the data field level. Group key management can remove some of the complexity of this task, but the overall process of key management simply requires investment and clear communication among participants prior to network membership, as well as constant monitoring and upkeep once participation is established.

- **Option 2:** Leverages promising technologies that are not yet easily scalable. Zero-knowledge proofs add several seconds of latency to each transaction it is applied to. Fully Homomorphic Encryption (FHE) is turning into the most powerful and useful encryption technology for blockchain and supply chain, but other technologies may be better equipped to provide value today.
- **Option 3 and 4:** Control the admittance of membership in a blockchain network more closely. These options therefore have greater ability to manage identities. If not coordinated through key management, then access to data will have to be run against a list of permissions before that access is granted. That list, in turns needs to be maintained across blockchain nodes and freely auditable by the owners of the nodes to ensure trust. The administrative effort involved may make these approaches cost-prohibitive.
- **Option 5:** Pairs a public or private blockchain with an off-chain database that is part of a member's node. The blockchain will only save the hash of the data. The system needs to be architected such that when the hashed data is queried, it can be pulled from the off-chain database and verified against a key management system that is part of the blockchain. Those who are meant to have access to the data will then be able to take actions upon it.

> *When implementing any data protection measures, it is important that the participants in a blockchain network be able to audit and control access rights to their data, regardless of the overall network governance structure. For those reasons, Role Based Access Controls and key management are far more trustable than FHE and Zero Knowledge Proofs, which still require faith in algorithms.*
>
> **Rebecca Liao,** Co-Founder and Executive Vice President, Skuchain

3. Case study: a manufacturing conglomerate

How should data protection technologies be applied in a real-world use case?

Challenges

To see how technologies may be adopted on a curve or combined to achieve optimal results, let's consider through a hypothetical use case in collaborative planning. This is one of the most fertile grounds for efficiency gains in supply chain and one of the hardest to achieve due to privacy reasons.

Let's say a major heavy manufacturing company has historically overstated its forecast to its plastics supplier to account for potential emergency orders. The supplier has become aware of this practice after years of building up excess inventory because the manufacturing company ultimately does not buy anywhere near the levels of its forecast.

One year, the supplier decides to significantly cut the procurement of resin from its supplier, a Tier 2 supplier to the manufacturing company. The supplier cut too much and could not meet demand for the manufacturing company that year.

In an effort to avoid supply outages, the manufacturing company would like to access data about the plastic supplier's, even the resin supplier's, inventory on hand and production rate on a more frequent basis. The plastic supplier would like to know the manufacturing company's inventory level, consumption rate and demand forecast as often as possible. None of the parties have any incentive to share this information with one another given how it will affect pricing and negotiation leverage. The question therefore arises of what can be done.

If the manufacturing company simply knew the schedule of delivery, in real time, of resin to the plastic supplier and of the plastic supplier to them, there can be an incremental improvement in planning. Perhaps the resin supplier is not ready to share other information at this time, so the logistics information goes onto the blockchain, but other data stays off-chain.

Applying the toolkit

On the other hand, another solution may be one in which all parties are comfortable placing just-in-time (JIT) inventory data on the blockchain, but only their immediate counterparty has access to the information. In addition, the counterparty may have access for purposes of executing smart contracts or algorithms with the data, but the counterparty may not see the underlying data itself. Employing encryption together with a key management is one of the solutions. This native approach is scrutinised and achieves good maturity.

Otherwise, with role-based access controls (RBAC) on the blockchain, the parties are able to accomplish this. They can then engage in collaborative planning with data that is obfuscated but usable for valuable data analysis. With both of these technologies, sensitive data can stay hidden, but it is not exactly encrypted.

If the companies are unsatisfied with RBAC and key management but still want to use the cryptographic technique, then more sophisticated means will have to come into play. If the manufacturing company wants to control the level of resin inventory at the plastic supplier, then when the level falls below 5,000 litres, the manufacturing company will ask the plastic supplier to order more.

A zero-knowledge proof can certify to the manufacturing company that this threshold has indeed been crossed without revealing exactly how much resin remains at the plastic supplier. Otherwise, fully homomorphic encryption allows all parties to place their data on the blockchain, keep it encrypted, and simply run any planning algorithms on the encrypted data.

As is noted as a drawback, zero-knowledge proof or fully homomorphic encryption will incorporate complex software which requires further engineering for maturity. Also, these techniques may consume more computational power, and this can be a bottleneck if the system is expected to handle a large amount of data.

TOOLS AND RESOURCES

4. Data protection strategies

The following checklist is an overview of high-level considerations your organisation will need to address to approach data protection concerns. It collects together the key points presented in this module and the more detailed overviews in the focus areas can be referenced while going through this checklist.

Since data protection considerations will have a pervasive impact on the final implementation of the project, these questions should be considered early in the timeline of a blockchain deployment, in the later portions of the design phase, after the core value proposition and mechanics of the use case have been determined but before the use case begins code development. Data protection considerations should be revisited ongoing in an organisation as external and internal requirements, rules and regulations change.

- ☐ Which supply-chain partners need access to certain information to execute their roles on the network?
    - Who has write permissions? Who has read permissions? How are these permissions established (who would determine who has access to the blockchain)? What level of access are users granted?
- ☐ How will the protocol, framework or platform protect data privacy and confidentiality?
- ☐ Which approaches for data protection are best fit?
- ☐ When the approach for confidentiality is taken, what are potential drawbacks, barriers, and risks?
    - Is encryption-based protection adequate for safeguarding data, or is there sensitivity around even sharing encrypted data with unauthorised parties?
    - How do you overcome such drawbacks, barriers, and risks?
- ☐ How is identity managed to meet data protection needs?
- ☐ What sets of policies are needed for governance and control of the blockchain network?
    - How would these policies interact with individual contractual arrangements among the network participants for distribution and use of data?
- ☐ How frequently do data standards change, and what level of flux does it cause for the data stored on-chain?
- ☐ How long does data need to be available for, and how does this affect any archival and obsolescence processes?
- ☐ What data access audit requirements need to be built into the system?
- ☐ Protocols are rarely deployed without middleware or an application layer sitting on top of them, each of which will likely have its own data privacy functionality. Where will the data privacy features sit?
- ☐ Is overall system security engineering designed to achieve data confidentiality? Many, if not most, of the purported features and capabilities of blockchain are design- and implementation-specific. Assumptions should not be made that because one design implementation includes a particular feature, that others will share that feature as well.