Trustworthy, Secure, and Privacy-aware Food Monitoring Enabled by Blockchains and the IoT

1st Christoph Stach
University of Stuttgart
Stuttgart, Germany
stachch@ipvs.uni-stuttgart.de

2nd Clémentine Gritti
NTNU
Trondheim, Norway
clementine.gritti@ntnu.no

3rd Dennis Przytarski
University of Stuttgart
Stuttgart, Germany
przytads@ipvs.uni-stuttgart.de

4th Bernhard Mitschang
University of Stuttgart
Stuttgart, Germany
mitsch@ipvs.uni-stuttgart.de

***Abstract*—A large number of food scandals (e. g., falsely declared meat or non-compliance with hygiene regulations) are causing considerable concern to consumers. Although *Internet of Things* (*IoT*) technologies are used in the food industry to monitor production (e. g., for tracing the origin of meat or monitoring cold chains), the gathered data are not used to provide full transparency to the consumer. To achieve this, however, three aspects must be considered: *a)* The origin of the data must be verifiable, i. e., it must be ensured that the data originate from calibrated sensors. *b)* The data must be stored tamper-resistant, immutable, and open to all consumers. *c)* Despite this openness, the privacy of affected data subjects (e. g., the carriers) must still be protected. To this end, we introduce the *SHEEPDOG* architecture that *"shepherds"* products from production to purchase to enable a trustworthy, secure, and privacy-aware food monitoring. In SHEEPDOG, *attribute-based credentials* ensure trustworthy data acquisition, *blockchain* technologies provide secure data storage, and *fine-grained access control* enables privacy-aware data provision.**

***Index Terms*—Attribute-based credentials, blockchain, data authentication, IoT, privacy, service utility, transparency, trust**

I. Introduction

The *Internet of Things* (*IoT*) is becoming increasingly relevant. Due to sensors integrated into *Smart Things*, any aspects of daily life can be monitored continuously. By means of *Big Data* analytics, the collected data can be processed productively in manifold application areas.

Also, the food industry has realized the great potential of the IoT. In this sector, limited food resources must be managed responsibly in order to feed as many people as possible. IoT techniques can be used to monitor the entire food supply chain [1]. This enables to monitor livestock via sensors, e. g., to document feed additives or to detect diseases at an early stage [2]. It is also possible to monitor the entire transport chain from butchery to retailer, e. g., to verify that the cold chain has been maintained all the time [3]. Even the meat products can be monitored via cheap disposable sensors in the packaging, e. g., to detect the formation of bacteria and thus the contamination of the product [4]. These data can be used in operational management processes to optimize production processes by reacting dynamically to unexpected incidents [1].

However, these data are also valuable for consumers. Reports about food scandals are unsettling consumers, e. g., the horse meat scandal in 2013 (horse meat was falsely declared as beef) [5] or the recent discovery that meat products were contaminated due to non-compliance with hygiene regulations [6].

The data captured by the food industry can be used to restore consumer trust. If the industry fully publishes these data, consumers are able to trace the products they buy back to their origin and retrieve any information related to food safety [7]. However, in order to provide such an end-to-end traceability towards consumers, three requirements have to be met: *a)* The data sources must be verified so that the data cannot be manipulated before transmission and the integrity of the data is assured. *b)* The data must be stored tamper-resistant and immutable and made available transparently to all customers. *c)* The privacy of affected data subjects (e. g., movement patterns of carriers) and the interests of the producers (e. g., insights into supply chains) must be protected.

To this end, we introduce the *SHEEPDOG* architecture to enable a trustworthy secure privacy-aware food monitoring. We make the following contributions: *A)* We present an *attribute-based credentials* mechanism (*Requirement a*). This mechanism is tailored to the IoT, i. e., it is particularly resource–efficient. Yet, it is capable to verify identifying attributes of Smart Things without revealing any additional information. *B)* We outline a *blockchain-based data storage* (*Requirement b*). In it, data are protected against manipulation and visible to everyone. Yet, unlike conventional blockchain-based data stores, it enables the efficient execution of Big Data analytics on the stored data. *C)* We introduce a *pattern-based access control* (*Requirement c*). By means of a privacy-aware attribute-based access control, it determines dynamically which data are available to an entity and applies privacy techniques to them if necessary. Yet, data utility is still upheld. That way, the SHEEPDOG *"shepherds"* food data. Yet, the food industry is only an application example. SHEEPDOG can be applied to any scenario where trustworthy, secure, and privacy-aware data management is required.

The remainder of this paper is as follows: In Section II, we discuss related work. Then, we present our SHEEPDOG architecture in Section III. Finally, Section IV concludes this paper and gives a brief outlook on future work.

II. Related Work

To the best of our knowledge there is no approach that meets all three of the aforementioned requirements. Thus, we discuss

This work is partially funded by the DFG project DiStOPT (252975529).

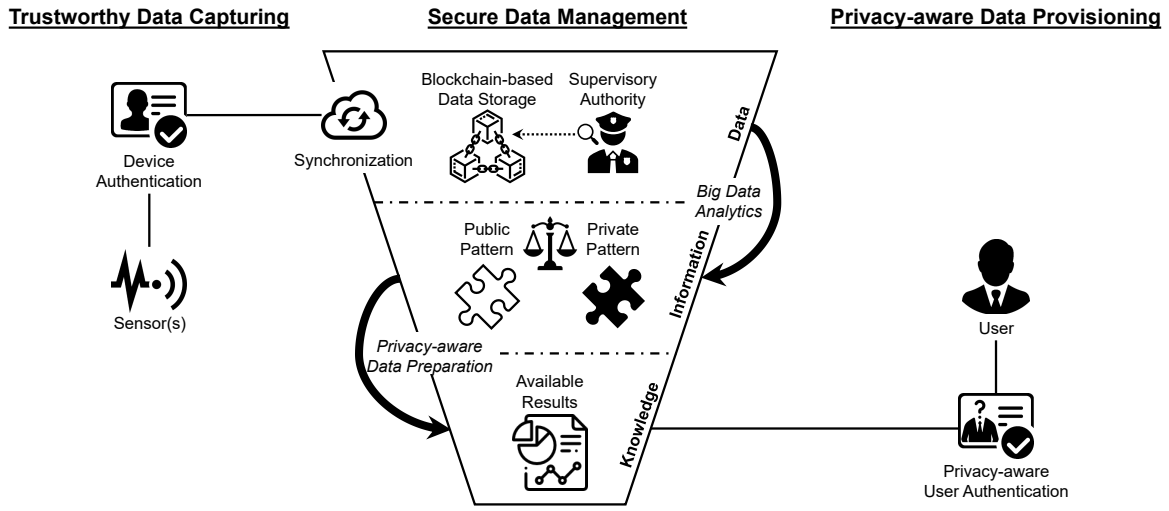

Figure 1. Simplified Architecture and Fundamental Modus Operandi of SHEEPDOG.

related work separately in the categories *device authentication*, *secure data storage*, and *privacy-aware data provisioning*.

a) Device Authentication: Attribute-based authentication is well suited for dynamic environments such as the IoT. Characteristic attributes of an entity are used for authentication (e. g., the firmware version of a Smart Thing). Yet, these algorithms are complex, i. e., they require much computing power and consume a lot of energy [8]. Huang and Wang [9] introduce a resource-saving authentication procedure via *physical unclonable functions* (*PUFs*) integrated into Smart Things. Yet, this is expensive and thus unsuitable for widespread use.

b) Secure Data Storage: Public blockchains can be used for a tamper-evident, decentralized, and transparent publication of food data [7]. Yet, this openness contradicts the privacy of the data subjects. *Permissioned blockchains* therefore support fine-grained access control [10]. As access rights for each data record can be revoked, transparency is no longer ensured. Also, Big Data analytics on blockchains are complex and expensive.

c) Privacy-aware Data Provisioning: AVARE [11] enables users to apply privacy filters to data sources. These filters are tailored to the respective sources, so that the data quality and thus the privacy threat can be reduced. Yet, data utility is not considered, i. e., the usability of applications is severely impaired. Approaches based on *differential privacy* ensure a high data utility but are only suitable for statistical analyses [12].

Since none of these approaches is adequate in our application scenario, we now introduce our approach called SHEEPDOG.

III. The SHEEPDOG Architecture

The basic idea of SHEEPDOG is shown in Fig. 1. First, SHEEPDOG ensures *trustworthy data capturing*. To this end, the sensors must authenticate before transmitting data to SHEEPDOG. A resource-friendly attribute-based credentials mechanism is used for this purpose (see Section III-A). The attributes prove, e. g., that the hardware and software of the data source are not manipulated and that the transmitted data are therefore legit. As IoT devices do not have a permanent Internet connectivity, SHEEPDOG uses *ECHOES* [13], a lightweight synchronization mechanism for Smart Things. This ensures that even if the connection is interrupted, the data is transmitted correctly as soon as the Internet connection is re-established.

For *secure data management*, the transmitted data is stored in a blockchain-based data storage. Yet, the public blockchain holds only proof that the data are authentic. Only supervisory authorities (e. g., consumer protection agencies) have unrestricted access to the payload data (see Section III-B). Other users have just access to data required for their use cases. This approach is similar to the use of *Smart Contracts* on a permissioned blockchain. Yet, our approach is more lightweight and enables Big Data analytics on the data. Such analyses are necessary to derive information patterns from the data. We differentiate between *private patterns*, which must not be published (e. g., movement profiles), and *public patterns*, which are required for certain use cases (e. g., temperature progression to verify the cold chain). In addition, use case tailored *privacy filters* can be applied to these public patterns to make the available results even more privacy-friendly.

For *privacy-aware data provisioning*, a user must authenticate. A similar approach is used as for the authentication of sensors. Yet, this approach also considers privacy, as the attributes might reveal a lot of information about a user. After authentication, SHEEPDOG selects which knowledge may be made available in terms of maximizing data utility while still protecting the privacy of the data subjects (see Section III-C).

A. Attribute-based Credentials

The process applied in SHEEPDOG for attribute-based credentials is shown in Fig. 2 in accordance with Gritti *et al.* [14]. Initially, a trusted authority (e. g., a certification authority) generates a key pair for each sensor (e. g., the thermometer in a refrigerated vehicle). These keys include the attribute values of a sensor (e. g., its firmware version). If the sensor has been manipulated, i. e., its attribute values have changed, the key cannot be used. Beside a *full key* for

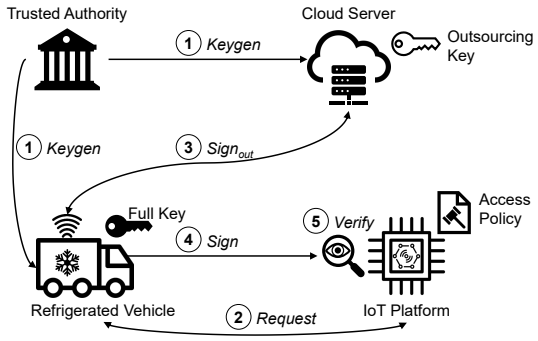

Figure 2. Authentication Process Applied in SHEEPDOG (cf. [14]).

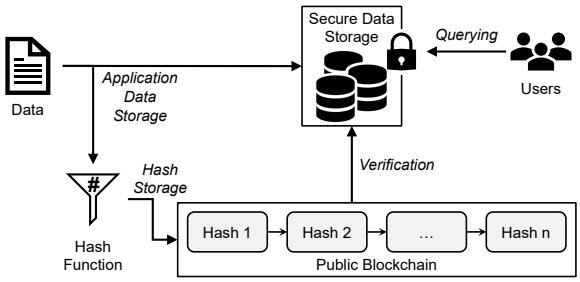

Figure 3. Blockchain-based Data Storage Applied in SHEEPDOG.

the sensor, an *outsourcing key* is generated, which contains only hashes of the attributes[1]. This key is forwarded to a Cloud server, which carries out the computation-intensive steps ①. While asymmetric cryptography imposes greater loads on processors, memory, and electrical resources than symmetric cryptography, its main advantage is effective key management, a must in an IoT environment with many actors. Yet, our asymmetric approach remains resource-efficient since the majority of computations can be delegated to the Cloud.

When a sensor requests to send data to an IoT platform (e. g., the SHEEPDOG data storage), it receives an *access policy* back, i. e., a description of the required attributes ②. The sensor sends a hashed version of this access policy to the Cloud server. It signs the sensor data with its outsourcing key ③. Since the Cloud server receives only hashed data and not the actual data, it cannot derive any secrets. That way, even if assuming an honest-but-curious Cloud server, the sensor data is still protected. In order to prevent the Cloud server from sending data to SHEEPDOG without the sensor's knowledge, the sensor must additionally sign each message with its full key ④. This step, however, is lightweight, as the messages are already pre-signed and therefore requires little resources.

Finally, the message is sent to SHEEPDOG. There, the signature can be used to uniquely verify from which sensor the data originates and whether it corresponds to the configuration specified by the certification authority, i. e., whether the hardware and software of the sensor is not manipulated ⑤.

B. Blockchain-based Data Storage

If a sensor was successfully authenticated, its data is stored persistently, immutably, and transparently. A public blockchain is used in SHEEPDOG for this purpose. However, this entails three issues: I_1 Storing data in a blockchain causes huge transaction fees. I_2 Complex operations on the data cannot be performed efficiently. I_3 All data is accessible by all users. In SHEEPDOG, we therefore apply a different approach.

This approach is shown in Fig. 3. The basic idea is to minimize the amount of data stored in the blockchain. For this purpose, a hash is calculated for each transmitted data record. Instead of storing the complete record in the blockchain, only the much smaller hash is stored in it. The actual data can be stored in any kind of data store. This solves Issue I_1. By choosing a data store type suitable for the intended data, the execution of the expected operations can be facilitated. For instance, in the context of heterogeneous IoT data, *triplestores* are well suited for this purpose. This solves Issue I_2. By applying data security techniques such as access restriction and encryption to this data storage, Issue I_3 is solved as well.

That way, data management in SHEEPDOG is efficient and secure. Nevertheless, any time a user is in doubt about the data provided to him or her, SHEEPDOG can recalculate the hash of the data stored in the data store and compare it to the hash in the assigned transaction in the blockchain.

Even if some of the data must be open to all users, a mixed strategy is also possible, i. e., a subset of the data can be stored in the blockchain in addition to the hashes. We also consider approaches for storing triples in the blockchain so that Big Data analyses can also be carried out efficiently on the blockchain. For more information on this topic, see Przytarski [15].

C. Pattern-based Access Control

To enable users to access the payload in the secure data store, we introduce a privacy-aware data access mechanism for SHEEPDOG. This mechanism is shown in Fig. 4. First, a user has to be authenticated. We apply a method from Gritti *et al.* [16] that is similar to the attribute-based credentials process described above. To this end, a user transmits all identifying attributes to a trustworthy IoT gateway ❶. The gateway verifies the attributes and forwards a strict subset of these attributes to the IoT platform—i. e., SHEEPDOG—for authentication ❷. This means, no privacy-critical data is shared with the platform.

After successful authentication, SHEEPDOG checks which access rights this user has ❸. Access rights are specified in SHEEPDOG as public patterns. Such a pattern describes the knowledge that may be shared with the user (e. g., the production chain of a certain product). This knowledge is mapped to data sources, i. e., based on the patterns SHEEPDOG can identify all relevant data records [17].

Before these data are forwarded to the user, it must be checked whether they reveal any private patterns, i. e., whether they compromise the privacy of a data subject or interests of the producers. For this, a *utility metric* is applied which maximizes the detection of public patterns and prevents the detection of private patterns [18]. This enables a fine-grained and need-based

[1]We use a keyed hash function with a symmetric key only known to the sensor and the IoT platform, i. e., the Cloud server cannot read the attributes.

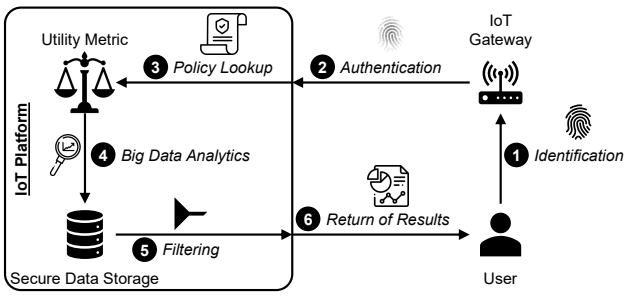

Figure 4. Privacy-aware Data Access in SHEEPDOG (cf. [19]).

data access control. According to this metric, the corresponding patterns are then extracted from the data storage by means of Big Data analytics ❹. In addition, privacy filters can be applied to these patterns ❺. Stach [19] discusses such privacy techniques that preserve data utility. The privacy-aware results are then forwarded to the user ❻. By applying homomorphic hashing, the hashes of the results can still be verified against the public blockchain despite the privacy filters [20].

Reflections

With SHEEPDOG, we strive for a trustworthy, secure, and privacy-aware food monitoring. In order to achieve this, three requirements have to be met: *a)* The sensors must be verified to ensure data integrity. *b)* The data must be made available persistently, immutably, and transparently. *c)* The privacy of affected data subjects must be protected.

The attribute-based credentials used in SHEEPDOG (see Section III-A) ensure the authenticity and integrity of the sensor data. The therefore applied procedure is tailored to the IoT.

The blockchain-based data storage used in SHEEPDOG (see Section III-B) enables transparent verification of stored data in terms of tamper-resistance and immutability. For efficiency reasons, we waive tamper-resistant and immutable storage.

The access control mechanism used in SHEEPDOG (see Section III-C) protects the privacy of both, data subjects and users by enabling a need-based data provisioning.

Thus, the SHEEPDOG architecture is able to *"shepherd"* products from production to purchase and enables a trustworthy, secure, and privacy-aware food monitoring.

IV. Conclusion and Future Work

Food scandals are unsettling consumers. The IoT enables an end-to-end monitoring of the food supply chain. Such monitoring data could restore customer trust in the food industry. To that end, we present SHEEPDOG. In SHEEPDOG, attribute-based credentials allow the verification of sensors and ensure data integrity. A blockchain-based data storage ensures the tamper-resistant and immutable archiving of the data. A pattern-based access control protects the privacy of affected data subjects without impairing transparent data provision.

SHEEPDOG is work in progress. That is, although the usability and capability of the individual concepts have already been demonstrated in practice, the efficiency of SHEEPDOG as a whole must be evaluated as part of future work. In particular, it must be determined which storage infrastructure is best suited for the secure data storage and how Big Data analytics can be carried out efficiently even with a distributed data management.

References

[1] S. Parvin *et al.*, "Smart Food Security System Using IoT and Big Data Analytics," in *ITNG '19*, 2019.

[2] T. Khan, "Internet of Things: The Potentialities for Sustainable Agriculture," in *International Business, Trade and Institutional Sustainability*. Springer, 2020, pp. 291–302.

[3] H. Zhang *et al.*, "Security and Trust Issues on Digital Supply Chain," in *CyberSciTech '19*, 2019.

[4] A. Popa *et al.*, "An Intelligent IoT-Based Food Quality Monitoring Approach Using Low-Cost Sensors," *Symmetry*, vol. 11, no. 3, pp. 1–18, 2019.

[5] F. Lawrence, "Horsemeat scandal: The essential guide," The Guardian, Press Release, Feb. 15, 2013.

[6] F. Schulz, "Contaminated meat scandal exposes Germany's food safety flaws," EURACTIV, Press Release, Oct. 15, 2019.

[7] J. Lin *et al.*, "Blockchain and IoT Based Food Traceability for Smart Agriculture," in *ICCSE '18*, 2018.

[8] S. Belguith *et al.*, "Analysis of attribute-based cryptographic techniques and their application to protect cloud services," *Transactions on Emerging Telecommunications Technologies*, vol. June, pp. 1–20, 2019.

[9] Z. Huang and Q. Wang, "A PUF-based unified identity verification framework for secure IoT hardware via device authentication," *World Wide Web*, vol. April, pp. 1–32, 2019.

[10] A. D. Dwivedi *et al.*, "A Decentralized Privacy-Preserving Healthcare Blockchain for IoT," *Sensors*, vol. 19, no. 2, pp. 1–17, 2019.

[11] S. Alpers *et al.*, "PRIVACY-AVARE: An approach to manage and distribute privacy settings," in *ICCC '17*, 2017.

[12] M. U. Hassan *et al.*, "Differential Privacy Techniques for Cyber Physical Systems: A Survey," *IEEE Communications Surveys Tutorials*, vol. October, pp. 1–46, 2019.

[13] C. Stach and B. Mitschang, "ECHOES: A Fail-safe, Conflict Handling, and Scalable Data Management Mechanism for the Internet of Things," in *ADBIS '19*, 2019.

[14] C. Gritti *et al.*, "CHARIOT: Cloud-Assisted Access Control for the Internet of Things," in *PST '18*, 2018.

[15] D. Przytarski, "Using Triples as the Data Model for Blockchain Systems," in *BlockSW '19*, 2019.

[16] C. Gritti *et al.*, "Privacy-preserving Delegable Authentication in the Internet of Things," in *SAC '19*, 2019.

[17] C. Stach and B. Mitschang, "Recommender-based Privacy Requirements Elicitation – EPICUREAN: An Approach to Simplify Privacy Settings in IoT Applications with Respect to the GDPR," in *SAC '19*, 2019.

[18] C. Stach *et al.*, "How a Pattern-based Privacy System Contributes to Improve Context Recognition," in *CoMoRea '18*, 2018.

[19] C. Stach, "VAULT: A Privacy Approach towards High-Utility Time Series Data," in *SECURWARE '19*, 2019.

[20] K. Lewi *et al.*, "Securing Update Propagation with Homomorphic Hashing," IACR Cryptology ePrint Archive, Technical Report, 2019.