Tackling the Cloud Forensic Problem while Keeping your Eye on the GDPR

Magnus Westerlund* and Martin Gilje Jaatun†‡
*Arcada University of Applied Sciences, Finland
†University of Stavanger, Norway
‡SINTEF Digital, Trondheim, Norway

***Abstract*—If the cloud is just someone else's computer, then securing forensic evidence can be tricky. The GDPR requires providers to deploy appropriate measures to safeguard their users' personal information, also in the case of onward transfer along cloud provider chains.**

In this position paper we delimit the processing architectures from a top-down organizational control point of view in order to define an initial taxonomy for an accountability-based approach that aims to improve both compliance and forensic readiness.

Through our initial analysis, we find that achieving compliance and forensic readiness may become even more difficult as we move towards distributed architectures, such as blockchain technology. Our conclusion is that the forensic challenge requires a renewed research focus, and we highlight how an accountability-based approach will be instrumental to the overall acceptability of the solution.

***Index Terms*—Cloud computing, forensics, accountability, security, privacy, blockchain, DLT**

I. Introduction

Accountability supports governance and control of corporate and private data processed by cloud providers [1]. Such data is currently typically stored and transferred in cloud provision chains, and it is thus imperative that accountability obligations of cloud service providers and organisations that use cloud services are enforceable along the entire provider chain.

Personal data can only be used by a third party if there is a legal basis for processing it. This implies that a system handling personal data cannot make a decision about the individual based on the personal data, unless under contract or through consent. This is not new, but the increased focus due to GDPR [3] should prompt service providers to question whether they really *need* to process personal data? Many things will be easier if this can be avoided altogether.

This paper analyzes the cloud forensic problem for systems processing personal data. Zhao and Duncan [4] advocate using blockchain-based cryptocurrencies for this purpose, and while we remain cautiously optimistic about the future of blockchain, we are aware that it is not all plain sailing ahead.

The remainder of this paper is organized as follows: In section II, requirements for defining an accountable system owner are presented. We then introduce the forensic problem in Section III, which is followed in section IV by an initial taxonomy for different system architectures for processing and their role in increasing cloud forensic complexity. Section V discusses our findings, and Section VI summarizes the paper.

II. Requirements for Accountable Organisations

We posit that any solution to the cloud forensic problem must involve accountable organisations. An accountable organisation must commit to responsible stewardship of other people's data [1]; it must:

- define what it does,
- monitor how it acts,
- remedy any discrepancies between the definition of what should occur and what is actually occurring
- explain and justify any action.

We will return to how this is relevant to the forensic problem, but in the meantime these elements can be elaborated as follows:

1) **An accountable organisation must demonstrate willingness and capacity to be responsible and answerable for its data practices.**
   Data practices refer to the processing and storing of data; this primarily concerns personal data as defined in the GDPR [3], but may extend to types of confidential information that do not involve personal data.
2) **An accountable organisation must define policies regarding their data practices.**
   Aspects of the data practices that need to be defined include:
   - the entities involved in the processing of data and their responsibilities
   - the scope and context of processing data
   - the purposes and means of processing
   - data handling and data access policies
   - risk monitoring and risk mitigation
   - relevant external legal obligations (such as what legal obligations the organisation has in disclosing data to third parties (e.g., in the context of law enforcement)

   These items include information obligations as defined in the data protection legal framework, but extend those to include all elements that are relevant for customers to make informed choices about the organisation's offering and that allow checking compliance later on (in the monitoring stage) and will also be based on business considerations related to the service provider's services.

TABLE I
CLOUD COMPUTING FEATURES AND RELATED PRIVACY ISSUES [2]

Cloud feature	Privacy Issue
Multi-tenancy	Data of co-tenants may be revealed in an investigation of another tenant, isolation failure, improper deletion of data
Complex, dynamically changing environment and data flows	Ensuring appropriate data protection, overlapping responsibilities, unauthorized secondary usage of data, vendor demise, lack of transparency
Data duplication and proliferation; unknown geographical location	The context of protecting data stored and transferred in cloud provision chains, and thus accountability obligations of cloud service providers and organisations that use cloud services, to data subjects and data protection regulators.
Convenient and enhanced data access from multiple locations	Data access from remote geographic locations subject to different legislative regimes, subpoenas, access by foreign governments; employees may unilaterally decide to use Cloud services for enterprise purposes without due regard to organizational policies or risk assessment

3) **Accountable organisations must monitor their data practices**
   Accountable organisations outline how they process data and have to be able to prove that they acted according to their policies and hence have to monitor the actual data practices and keep records of the monitoring and its results.
4) **Accountable organisations must correct policy violations**
   If discrepancies between the stated policies and actual (system) behaviour are detected, several things need to be done about it. First of all the effects of the violation need to be addressed. Errors need to be corrected, damages need to be compensated (financially or otherwise). Second, the causes of the violation need to be addressed. If the violation is the result of a faulty process, the process needs to be repaired, or improved. If the violation results from a data breach or (other) cybercrime, the security needs to be improved, etc. Third, the appropriate stakeholders need to be informed. In some cases the authorities (such as the Data Protection Authorities) need to be informed; in other cases the customer or affected data subjects may need to be informed.
5) **Accountable organisations must demonstrate policy compliance**
   The final element of the accountability loop is demonstration of compliance with the adopted policies. Furthermore, the organisation should be able to demonstrate that the controls that are selected and used within the service provision chain are appropriate for the context and provide evidence that the operational environment is satisfying the policies (cf point 3 above).

As mentioned above, there is also a need for accountability requirements across the cloud service provision and governance chains, and not just in isolation for organizational cloud consumers or cloud service providers.

III. THE FORENSIC PROBLEM

In the previous section we defined the requirements for an accountable organization, both in light of the GDPR [3] and in regard to improving the trust-relationship between organisations that handle personal data and users of said services. Still, accountability may include rather different aspects to consider when bearing in mind the type of IT systems used. The chosen software architecture of an IT system may often affect the approach to accountability.

In this section, we highlight four different processing approaches that require distinctive software solutions. We encapsulate a categorization from a top-down perspective of ownership control, and postulate specific cloud software architectures for each category. Traditional cloud-native application analyses tend to consider software and network architectures from a bottom-up approach [5]. Our motivation for a top-down analysis is primarily to understand the ownership perspective, i.e. organization accountability. Abbasi et al. [6] promote a similar perspective, and state that the key objective for Software-Defined Cloud Computing (SDCC) is to promote open and interoperable solutions, in our opinion thus requiring a top-down analysis approach.

In Fig. 2 we visualize the processing approaches as an industry evolution (1-4), as a continuation of the work by Westerlund and Kratzke [7]. The first included category (1) is centralised processing, that makes use of infrastructure under direct control of the service provider (owner). This control can either be direct through hardware access, or indirect through virtualization. The second category (2), decentralised processing, entails the use of containers and Application Programming Interfaces (API), as to increase the level of service abstraction. Implementation of these types of software architectures often avoid dealing with the underlying infrastructure and simply assume its flawlessness. Compared to (1), the monitoring and handling of security attacks against the infrastructure is often delegated to a cloud provider. The third category (3) is distributed processing that is based on an ad-hoc configuration of nodes and that often communicate peer-to-peer.

A distinguishing trait of (3) compared to (1) is that it may provide improved processing efficiency and is less sensitive to certain attacks, such as denial of service attacks. The final category (4), autonomous processing refers to implementations that are not only distributed but that are also self-contained and self-governed. Category (4) type of architectures are still to become widely used but may offer Turing complete software environments that are dispersed on utility computing-based infrastructure (cf. smart contracts on blockchain [8]).

IV. A Taxonomy of Processing Architectures

The centralised processing category (1) embraces the traditional architectures based on local installations, virtual machines, or bare metal cloud provisions, where the provider is in control of the operating system and hence can create a security audit trail of both system and service. The audit trail is often stored in an append only database in a separate system, but is still often dependent on a monitoring process instance running on the same system that is being monitored. Hence, a weakness can occur if the monitoring process instance is disabled by an attacker. From a forensic perspective, access to full system logs may offer the ability to create a complete view of the system at a desired time. Provided that functionality of monitoring processes can be guaranteed, this type of a system may offer the best architecture for handling forensic activities and the forensic problem is more of a question of resources.

Decentralised processing (2) refers to architectures such as microservices, Function as a Service (FaaS), and serverless. Typical for these types of architectures are that they function on an abstraction layer above category 1, by often being without permanent states and not offering access to the operating system logs. Decentralization can be achieved as the service operator is free to implement the provisioning of the service in various layers and automatically serve content from any number of cloud service providers or cloud regions. There is a difference between microservice and FaaS/Serverless architectures in that microservices often retain control of container management functions, whereas the later architectures are based on a time-share principle and therefore are often unaware of infrastructure where they reside. We see though an increase in difficulty to adhere to the forensic requirements as presented in Section II, due to the loss of access to system data when using public cloud computing resources and the lack of permanent states for instances.

Distributed processing (3) architectures have advanced over the last two decades but many architectures are still in an early or immature development phase. These architectures include Peer-to-Peer (P2P) networks, blockchain ledgers, and IoT sensor networks. They share some characteristics with decentralised processing architectures but require direct node-to-node communication using a purpose-built communication protocol. A requirement for such a protocol is that it is self-maintained, e.g., for P2P file sharing it maintains various file segments, offer node discovery, and allows a client to determine file cohesion once a file is downloaded. Public distributed ledger implementations have cryptographic security measures, so called consensus achieving proofs, built into the communication protocol. The aim of these proofs, e.g. Nakamoto consensus and its implementation Proof of Work [9], are often to enable so called trustless processing. This entails that peers do not need to trust or know each other in order to validate transactions in the network. The forensic problem for distributed architectures is further exacerbated as each node in the network should to some extent implement its own audit trail functionality. The collection activities in case of forensic investigations is limited to each node's reluctance to share its respective audit trail. The solution for distributed ledger technology (DLT) has been to only record verified transactions in the ledger and to then rely on the cryptographic measures as a guarantee for network honesty. A solution that for Proof of Work consensus is known to be vulnerable to a majority attack by participating nodes in the network. For forensic activities it often means that client vulnerabilities are difficult to detect unless some reporting function is built into the client. Also, any such reporting function would be sensitive to tampering with and may thus be an unreliable source for an audit trail.

Considering (payload) data that may be processed in distributed systems, several problems can occur that are extremely hard to mitigate without a de facto controller, including infringement on intellectual property and violation of data protection rights. Considering the GDPR [3], the right to be forgotten hinders any processing of personal data on a distributed immutable ledger. In case anyone stores personal data on a ledger, the ecosystem may in the future be held in contempt of the law. For distributed systems it may often be unclear who holds data controller obligations, and as in "The Pirate Bay case" regrading copyright infringement, the developers and system maintainers were held accountable [10].

Autonomous processing (4), as indicated, is still mostly in the "further work" category.

V. Discussion

An accountable cloud service provider must provide its users with more control than they currently have in cloud service situations [1]. This implies more opportunities for (dynamic) negotiation of security SLAs, including such aspects of who may do what with the customer's data. Furthermore, cloud providers must provide evidence that the negotiated obligations are also met downstream throughout the service provision chain.

Cloud providers must prove that the procedures and mechanisms they employ are appropriate to the context. When using blockchain for forensic purposes, this also implies that providers need to argue why a conventional solution is not suitable. This is important to avoid pitfalls seen, e.g., in supply chains [11], where some players evidently have implemented blockchain solutions merely as a path to digitization, without asking themselves whether they really need the properties offered by the blockchain (as opposed to simply using an inter-enterprise Public Key Infrastructure).

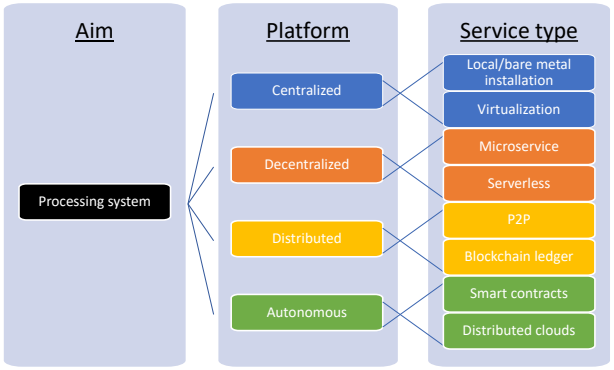

Fig. 1. Taxonomy for processing architectures

Integration of continuous monitoring contributes to a proactive approach to compliance monitoring. The cloud providers also need to satisfy external criteria such as relations to law enforcement agencies, and this needs to made explicit to the customers. Using DLTs in conjunction with continuous monitoring clearly contributes to transparency, but introduces additional privacy challenges as discussed below.

Privacy (in the context of personal information) is one of the reasons why we need accountability in the cloud. Warren & Brandeis' [12] definition of privacy – the right to be let alone – seems to have lost much of its validity, since we base so much of our existence on interaction and communication with others over the internet. Businesses have come to assume that we will use the internet for both purchases and user support.

In theory, privacy can be achieved by maintaining absolute confidentiality of all our personal information, but this would preclude us taking part in things like social networks, online shopping, and electronic filing of tax returns. Eventually, filing a paper tax return will be the equivalent to paying your back taxes in stacks of dried fish, and even the most paranoid will be left with no choice. This has led many individuals to despair, resorting to apathy as a privacy coping strategy [13].

An accountability-based approach [1] increases transparency for cloud users, and enables cloud providers to "do the right thing" with respect to accountability along the provider chain. It is only a question of time before providers will be required to justify their practices and mechanisms for handling customers' data to external parties [14].

The blockchain has some attractive features with respect to transparency; it is distributed, with no central authority, and it is also immutable – whatever happens on the blockchain, stays on the blockchain. The immutability does however pose a problem for one of the elements of the GDPR – the right to be forgotten, as mentioned above. This implies that no privacy sensitive data must be written to the blockchain, which poses a limitation on what can be recorded in the audit trail. This particular conundrum is not discussed by Zhao and Duncan [4]. Traditional logging in enterprises has not been subject to this restriction, and the current consensus is that GDPR does not pose an obstruction to conventional audit trail, as this falls under "legal grounds for processing".

With reference to Zhao and Duncan [4], we also wonder whether using cryptocurrencies as a basis for a cloud forensic solution is sustainable; whereas it seems to solve one problem (who will pay for the DLT infrastructure), it creates a disincentive for use (you effectively need to pay a transaction fee for every record). Although there are alternative proposals for bitcoin-like cryptocurrencies with lower resource consumptions [15], we believe a more community-oriented approach will be more appropriate.

VI. Conclusion

In this paper, we have presented how an accountability-based approach to service provisioning can contribute to improved forensic readiness. Furthermore, we have sketched how blockchain technology and smart contracts apply to various processing models, and indicated how the processing models relate to different forensic challenges. An initial top-down processing taxonomy for explicating organisational control differences has been put forward.

The analysis reveals that to enforce the concept of accountable organisations the forensic challenge will become harder to solve while moving towards distributed technologies and require further research to understand what this entails for autonomous processing. For further work, we are particularly interested in exploring the autonomous processing paradigm with respect to accountability and forensic capability.

Acknowledgements

This paper is inspired by a panel session at the SPCloud workshop of the 2018 HPCS conference. Thanks to Louise Spellacy for live-tweeting the panel session. This paper is in part based on results from the EU FP7 A4Cloud project.

References

[1] M. G. Jaatun, S. Pearson, F. Gittler, R. Leenes, and M. Niezen, "Enhancing accountability in the cloud," *International Journal of Information Management*, 2016. [Online]. Available: //www.sciencedirect.com/science/article/pii/S0268401216301475

[2] D. Catteddu and G. Hogben. (2009) Cloud Computing – Benefits, risks and recommendations for information security. ENISA Report. [Online]. Available: https://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport

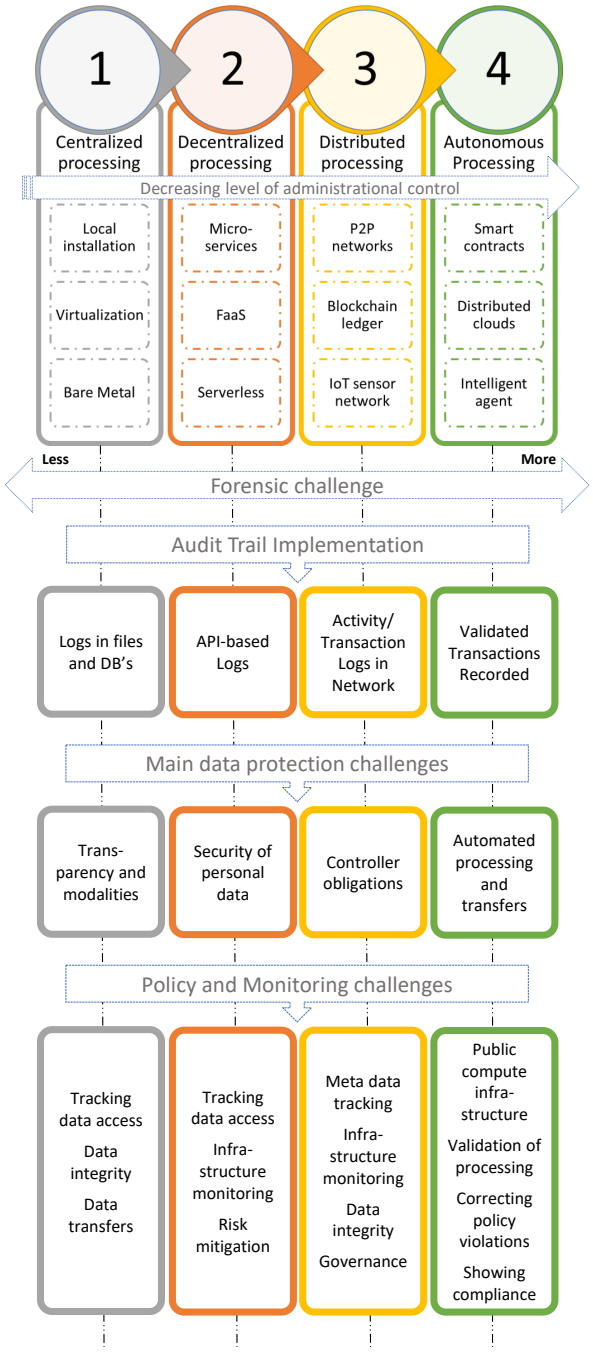

Fig. 2. Forensic challenge for various processing architectures

[3] EU, "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, no. 119, 2016.

[4] Y. Zhao and B. Duncan, "Fixing the cloud forensic problem with blockchain," *International Journal on Advances in Security*, vol. 11, no. 3 & 4, 2018.

[5] N. Kratzke and P.-C. Quint, "Understanding cloud-native applications after 10 years of cloud computing-a systematic mapping study," *Journal of Systems and Software*, vol. 126, pp. 1–16, 2017.

[6] A. A. Abbasi, A. Abbasi, S. Shamshirband, A. T. Chronopoulos, V. Persico, and A. Pescapè, "Software-defined cloud computing: A systematic review on latest trends and developments," *IEEE Access*, vol. 7, pp. 93 294–93 314, 2019.

[7] M. Westerlund and N. Kratzke, "Towards distributed clouds: A review about the evolution of centralized cloud computing, distributed ledger technologies, and a foresight on unifying opportunities and security implications," in *2018 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE, 2018, pp. 655–663.

[8] M. Bartoletti and L. Pompianu, "An empirical analysis of smart contracts: platforms, applications, and design patterns," in *International conference on financial cryptography and data security*. Springer, 2017, pp. 494–509.

[9] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system."

[Online]. Available: https://bitcoin.org/bitcoin.pdf

[10] "Fråga om straffrättsligt medverkansansvar för upphovsrättsintrång genom tillhandahållande av en fildelningstjänst (the pirate bay) inom ett datornätverk och om de medverkandes skyldighet att för intrånget betala ersättning enligt upphovsrättslagen (1960:729)." Court Decision - Swedish Court of Appeal (Svea hovrätt) RH 2013:27, 2013, (Swedish only). [Online]. Available: https://lagen.nu/dom/rh/2013:27

[11] M. M. Queiroz and S. F. Wamba, "Blockchain adoption challenges in supply chain: An empirical investigation of the main drivers in india and the usa," *International Journal of Information Management*, vol. 46, pp. 70 – 82, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0268401218309447

[12] S. D. Warren and L. D. Brandeis, "The Right to Privacy," *Harvard Law Review*, pp. 193–220, 1890.

[13] P. Sprenger, "Sun on Privacy: 'Get Over It'," *Wired*, 1999. [Online]. Available: http://www.wired.com/politics/law/news/1999/01/17538

[14] S. Pearson, "On the relationship between the different methods to address privacy issues in the cloud," in *On the Move to Meaningful Internet Systems: OTM 2013 Conferences*, ser. Lecture Notes in Computer Science, R. Meersman, H. Panetto, T. Dillon, J. Eder, Z. Bellahsene, N. Ritter, P. Leenheer, and D. Dou, Eds. Springer Berlin Heidelberg, 2013, vol. 8185, pp. 414–433. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-41030-7_30

[15] C. Li, C. Rong, and M. G. Jaatun, "A cost-efficient protocol for open blockchains," in *Proceedings of Cyber Science 2019*, 2019, pp. 74–80.