Addressing Data Governance in Cloud Storage: Survey, Techniques and Trends

Majid Al-Ruithe[1], Elhadj Benkhelifa[1], Yaser Jararweh[2], Chirine Ghedira[3]

[1] Cloud Computing and Applications Research Lab, Staffordshire University, UK
[2] Jordan University of Science and Technology, Jordan
[3] Université de Lyon, France
mrowathi@gmail.com, E.Benkhelifa@staffs.ac.uk, yijararweh@just.edu.jo, chirine.ghedira-guegan@univ-lyon3.fr

Abstract

Cloud Computing is becoming a popular paradigm that has emerged to deliver IT services to consumers as a utility service over the Internet in recent years. The adoption rate of cloud computing is still in the beginning stage in some developing countries. One of the most important services that cloud computing offers is called, 'cloud storage'. Cloud storage is associated with many challenges because customers are still skeptical about trusting it wholeheartedly. There are many aspects related to user's concerns to adopt cloud storage. Data governance is a helpful solution to manage these concerns and make users trust cloud storage. Therefore, the main research contribution of this paper is addressing and considering data governance for reducing the risk on data when it is stored in the cloud. Also, this paper will present the data governance trends in light in new emerging technologies defined by different cloud models, including mobile cloud computing, cloudlet, edge computing and IOT-Cloud Converged architectures. Software defined solutions, more specifically, that for Storage, is also discussed in the context of this paper to potentially be a crosscutting solution for future developments of cloud storage in its different models, though underpinned by a predefined data governance strategy.

Keywords: Cloud storage, Data governance in cloud storage, Software-defined storage, Mobile cloud, Edge computing

1 Introduction

Data volumes in the world have increased and this growth leads to an increased demand for online services. Therefore, Cloud Computing has become one of the most significantly debated issues of information technology (IT) aspects to support data volume growth. Cloud Computing is an emerging trend and a serious adoption in both public and private sector organizations [1].

Cloud Computing has motivated the research on related technologies by academia and the industry [2]. It also is composed of various elements from other computational models, such as autonomic computing, grid computing and utility computing, to form one of the most innovative computational deployment architecture in the world today. Hence, for scientists, clouds promise to be an alternative to supercomputers, clusters and grids**.** Cloud computing definition commonly used today is that expressed by the National Institute of Standards and Technology (NIST) [3]. The NIST defines cloud computing as: "*a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*" [3]. In addition, Cloud Computing includes some essential characteristics, four cloud deployment models and three cloud service delivery models [2]. The essential characteristics of cloud computing are: on-demand self-service, broad network access, resource pooling, rapid elasticity, measured service, multi-tenancy and virtualization. The cloud deployment models are public, private, hybrid and community clouds [4]. Also, the main cloud service delivery models are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [5]. Figure 1 shows the cloud computing architecture.

In addition, storage is considered one of the most important services provided by cloud computing to consumers. This service allows users to store their data online over a network where data is remotely maintained, managed and backed up; this service allows users to access their data and information from any location and any time [6]. The core business of the storage service in cloud computing is the combination of application software with a storage device, to achieve changing from a storage device to a storage service using the application software [7]. Thus, cloud

*Corresponding Author: Yaser Jararweh; E-mail: yijararweh@just.edu.jo
DOI: 10.3966/160792642018111906013

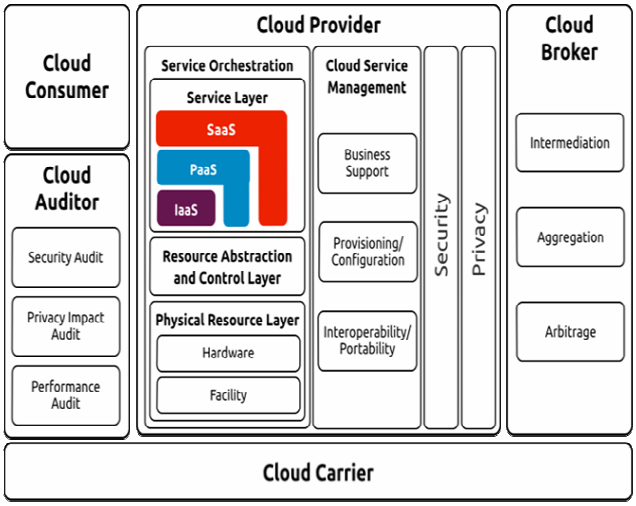


Figure 1. Cloud computing architecture [3]

computing provides direct data storage services for end-users and indirect data access in the application system, and many other services in the form of a network hard drive, online storage, online backup and online archive storage service. Cloud computing vendors provide storage services to clients with high availability, low cost and a pay-as-you-go option, such as Google, IBM and Amazon [8].

Moreover, storage service models and strategies in cloud computing are still in the early stages and the research and development in this arena is yet to be satisfactory [9]. Nowadays, many researchers and IT professionals are concerned with finding enough storage space to hold their data [8]. Although giant cloud storage providers have made successful services, such as Amazon and Google, many enterprises and scientists are still unable to make the transition into the cloud environment due to issues related to privacy, security, data protection and vendor lock-in [8]. Thus, to adopt cloud storage as a solution, the organization should standardize its service levels, data access methods, operational and security processes, and emergency plans for data migration if the enterprise wishes to change vendors and improve its performance in the future [9].

This paper is structured as follows. Section 1 presents the background regarding cloud computing and storage. Section 2 presents data governance in cloud storage with the barriers and challenges to adopt cloud storage. Section 3 describes the data governance techniques and policies for storing data in the cloud. Section 4 presents cloud storage and the data governance trend, while Section 5 provides concluding remarks.

2 Data Governance in Cloud Storage

In this study, we classify the cloud storage barriers into two parts: the barriers that are related to the technology and the barriers that are related to the data. The barriers related to the data that consider the most aspects which data governance is focused on when it is implementing for cloud storage. The technology barriers refer to barriers for the utility of cloud storage [11-14]. Figure 2 shows the of barriers adopting the cloud storage.

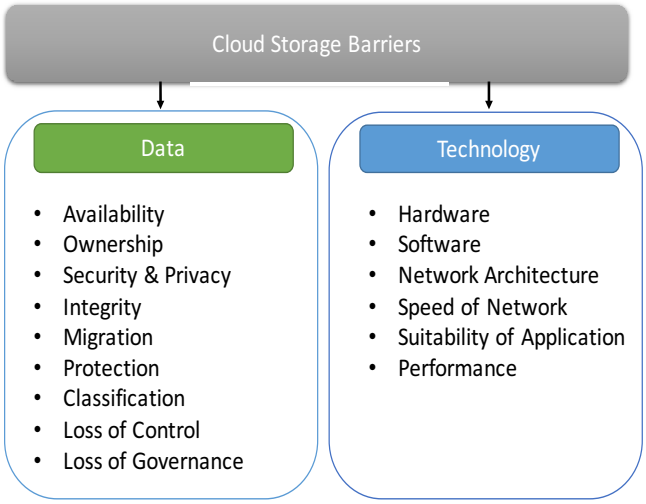


Figure 2. Barriers of adopting cloud storage

Due to recent advances in data storage and processing data sharing in the cloud computing environment, most organisations require flexibility for many data processes, such as: data representations, secure and consistent data production and automated data quality checking. These processes also require a strong governance methodology to achieve their goals. In addition, in the last few years some organizations have become aware of the increasing significance of governing their data in cloud computing and focus on data governance aspects to guarantee high data quality and better data management [15, 29, 32, 58].

Data governance refers to, "*the overall management of the availability, usability, integrity, and security of the data employed in an enterprise*". The data governance program includes a governing body or council, a defined set of policies, a defined set of procedures and a plan to execute those procedures. The academic and industry research on data governance for non-cloud and cloud computing is still in its infancy.

In our previous work, a systematic review approach was undertaken to review existing work from academia and industry on data governance [40]. The systematic review showed the absence of published academic research data governance in cloud computing and non-cloud environments. Figure 3 illustrates the number of published research in data governance in the last 10 years from both academia and industry literature. In the industry field, Security Cloud Alliance may be seen as a leader in the literature about data governance in the cloud [16]. In the literature, within the MeriTalk report in 2014, only 44% of IT professionals in the federal government believe their agency has mature data governance practices in the cloud, and about 56% of

agencies are currently in the process of implementing data stewardship or a data governance program. Cloud storage is radically changing the scale of organizing systems and the accessibility of the data and information they contain; therefore, effective data governance procedures can ensure that changes in cloud storage are systematically evaluated and implemented in a way that balances the potential against the risk [17].

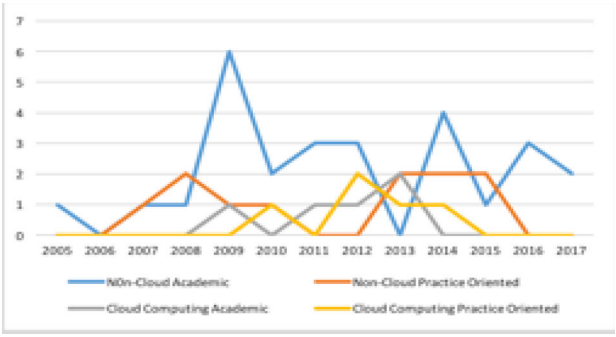


Figure 3. Number of published research and industrial contributions in data governance in the last 10 years [40]

In recent years, storage technology has moved through the development course from tape and disk to storage networking systems and the application demand for massive data storage is growing, which is directly contributed to the development of high-performance storage technology**.** The change in storage technology forces many organizations to apply data governance to gain control of their data when it moves to the cloud. Security, data loss prevention and data loss protection on distributed servers are essential important aspects of data governance in cloud computing technology [18]. Classifying data that is stored in cloud computing is very important and useful for organizations. In this study, we classify these data into three types: data in motion, data at rest and data in use. Therefore, we should consider these types of data when implementing a data governance program for cloud computing, and organizations should use these types when they determine the roles and responsibilities between cloud actors. Figure 4 displays the data types used in cloud storage.

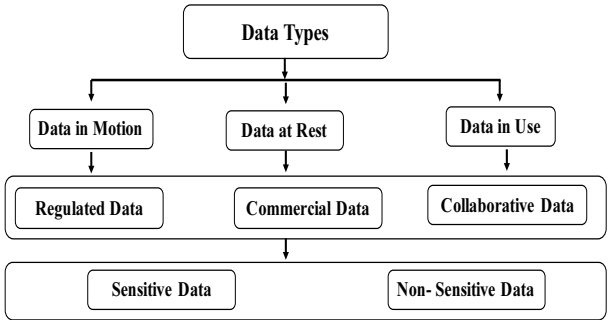


Figure 4. Data types used in cloud storage

However, there are several challenges that affect the implementation of the data governance program for cloud computing, namely regulations, security, privacy, data migration, data recovery, data classification and interoperability [16]. Therefore, data governance is one of the most important functions that removes barriers to enterprise adoption of cloud storage. Also, it has various barriers and challenges that are related to cloud storage and these need to be considered when implementing policies and procedures of a data governance program for cloud storage. Hence, this solution will reduce the cloud consumers concerns of many barriers which include Figure 5: Data Availability [19], Data Ownership [10], Data Security [20], Data Privacy [21], Loss Control [20], Data Integrity: [20], Data Protection [22]. Data Classification [16, 23], Data Migration [24], Data Integration [25], Regulatory Compliance [26].

Figure 5. Data governance barriers in cloud storage

Several schemes have been proposed to get the major attributes of data governance which is data security storage in the cloud environment. In the next section, we will present various existing storage techniques that make data security available in the cloud and the top practical tips to help optimize data loss prevention.

3 Data Governance Techniques and Policies for Cloud Storage

In this section, we present the various techniques and policies proposed in the literature to make data secure and data loss governance in cloud storage. Security and data loss prevention are terms often used interchangeably to describe the controls put in place by an organization to ensure that data of value remains under authorized use and care. Security and data loss prevention on distributed servers are of high importance in cloud computing technology, like any technology [18]. However, data loss prevention and protection in cloud storage are two characteristics that call for different solutions using different controls. Moreover, security and data loss prevention are

important functions which are considered when data governance is increasing in cloud computing. Therefore, the organizations should follow strong techniques and policies to achieve security functions when implementing a data governance program for their data in cloud storage environments. This study describes these policies based on policy name in Table 1 and practical tips in Table 2.

Table 1. Cloud data storage techniques

Storage Scheme	Authority By	Proposed Approach	Advantages	Limitations
Implicit Storage Security to Online Data	[18, 33-34]	Data partitioning scheme for online data storage.	Partitioned data pieces cannot bring out any user information.	If a user forgot where the data is stored, it will become difficult for users.
Identity Based on Authentication and Encryption	[18, 34, 36-37]	New authentication protocol based on identity which is derived from the hierarchical model	Weightless and more expeditious.	Only certificate communication is considered.
Efficient Third-Party Auditing	[18, 34, 38-39]	Novel and uniform security structure. Storage security is accomplished utilizing the BLS algorithm.	Auditor performs auditing jobs for different users at the same.	Unable to support both public verification and dynamic data correctness.
Dynamically Store in the Cloud	[34-35]	New protocol system using the data reading protocol algorithm. Multi-server data comparison algorithm to recover data.	Integrity can be verified before and after data insertion.	Third-party auditing is not considered for the integrity checking process.
Optimal Cloud Storage Systems	[34-35]	Taxonomic approach for achieving cloud storage service optimality. Proposed a new NubiSave prototype.	Proposed generic architecture served as a blueprint for optimal storage controller. NubiSave is freely available.	NubiSave needs to integrate with frontends for future research.
Accessing Outsourced Data Efficiently	[34, 37]	An owner-write-user-read scenario for accessing data.	Original data owner is only able to update/ modify their data.	Combination of multiple policies is not supported.

Table 2. Policies for governing data in cloud storage

Data Governance Policies	Practical Tips	Description
Identify and Classify Data	Identify and classify your data that is stored in the Cloud.	Using data classification to identify, organize and secure all sensitive data and get a low privacy and data security risk when migrating data into the cloud storage.
Access Control	Be concerned about view-only access.	Building security controls with the idea that view-only or read-only access is low-risk.
Data Management Lifecycle	Implement a data management lifecycle in the Cloud.	If data is properly defined, classified and stored appropriately from the beginning, it can propagate throughout the organization that can make protecting it later an even less challenge.
Network Authorisation	Do not allow unauthorized devices on your network.	This will help lessen risk organizations by preventing access to internal network resources by other people who do not work in the organization.
Sensitive Data Location	Do not permit the copying of sensitive data to removable media and do not keep it in the public Cloud domain.	This will assist to protect sensitive data and prevent spreading sensitive data out of the organization.
Understand Data Types	Understand data usage and flows and your data loss vectors.	Organizations need to understand how data is being used and left in the organization. Tools should be implemented to monitor data traffic flows within your infrastructure and increase the current knowledge of data usage within the organization, and network base.
Authorization and Access Control Measures	Improve authorization and access control measures in the Cloud Computing environment.	Review and tighten data access controls is important to ensure data safety and to ensure employees have access only to the data required to fulfil their responsibilities successfully and nothing more. When identity and access management systems are first deployed, user roles and user access are usually broadly defined.

Table 2. Policies for governing data in cloud storage (continue)

Data Governance Policies	Practical Tips	Description
Risk Approach	Take a risk-based approach.	You should note that not all data is created equally. Therefore, your data governance program should be designed to protect your sensitive data (not all data).
Update policies	Update organization policies.	You should improve your security awareness program and build clear guidance to educate employees to incorporate data loss awareness in line with organization policies to ensure that everyone is aware of the potential data loss risks.
Create Awareness and Audit Compliance	Create awareness and audit your own compliance.	Often awareness programs are not tested correctly to validate an employee's knowledge and compliance with organization policies. Therefore, you should consider conducting phishing tests and social engineering to determine awareness levels.

3.1 Data Governance Techniques and Policies

Data governance must be reasonable to ensure the overall management of the security of the data employed in an enterprise. In this paper, we aim to review all the techniques that have been proposed by existing work in the literature. In the literature, we fund there are many techniques that have been proposed to make data secure and ensure data loss prevention in cloud storage. In this section, we present and discus these techniques. These include the six storage schema:

Implicit storage security to online data. This technique aims to partition the data for online data storage. The partitions are stored on randomly chosen servers on the network and they need to be retrieved to recreate the original data. Data reconstruction requires access to each server, a login password and knowledge of the servers on which the partitions are stored. This scheme may also be used for data security in sensor networks and internet voting protocols.

Identity based on authentication and encryption. This technique is a new authentication protocol based on identity which is derived from the hierarchical model. It consists of the following four algorithms: Setup, Extract, Encrypt and Decrypt

Efficient third-party auditing. A third-party auditor (TPA) is type of checker. There are two categories: private auditability and public auditability. Although private auditability can achieve higher scheme efficiency, public auditability allows anyone, not just the client (data owner), to challenge the cloud server for the correctness of the data storage while maintaining no private information. To reduce the burden of management of data by the data owner, the TPA will audit the data for the client.

Dynamically store in the cloud. This is a new protocol system using the data reading protocol algorithm. A multi-server data comparison algorithm is used to recover the data. The advantages of this schema include the integrity can be verified before and after data insertion.

Optimal cloud storage systems. This considers a novel cloud storage management system which optimally combines storage resources from multiple providers so that redundancy, security and other non-functional properties can be adjusted adequately to the needs of the storage service consumer. It also is a taxonomic approach for achieving cloud storage service optimality.

Accessing outsourced data efficiently. This provides secure and efficient access to large-scale outsourced data, which is an important component of cloud computing. This mechanism has been proposed to solve the problem in owner-write-user-read applications. This mechanism approach is based on encrypting every data block with a different key so that flexible cryptography-based access control can be achieved. In addition, this mechanism is designed to handle both updates to outsourced data and changes in user access rights. Table 1 shows that cloud storage techniques. guidelines for ensuring the proper management of an organization's digital information. Such guidelines can involve policies for enterprise risk planning, security, data quality, privacy, access control, data classification and network authorization.

4 Cloud Storage and Data Governance Trend

Recently, cloud storage has been targeted in multiple real-world applications to simplify IT operations while saving costs. Various government departments in different countries, such as defense, intelligence, and finance are adopting cloud storage services. In this section, we present the state-of-the-art implementations of cloud storage. We first describe the key current SDS. Then, we present the new trend of storage, SDSt and describe how it impacts data governance strategy, as well as, we will present how the data governance will affect SDS, and its types, such as software defined storage. In addition, data governance in the Internet of Things (IoT) and cloud converged environments will be considered in this paper as the trends of data governance.

4.1 Software-defined Systems

Software-defined systems (SDS) will be new trends impacting enterprise data centres in the future, because they are an inevitable result of the paradigm transfer from traditional computing models to the utility-based cloud computing [27]. Software-defined systems refers to an approach for automating the process of optimal cloud configuration by extending the concept of virtualization to all types of resources in a data centre [28]. Virtualization provides the ability to divide physical resources, which allows secure, data governance, efficiency and multi-tenancy upon single machines. Also, it enables the ability to aggregate virtualized resources across multiple hosting providers to provide redundancy and flexibility during resource migration and elasticity through rapid resource provision and cloning [45]. There are many concepts and technologies to enable software-defined cloud computing, namely system virtualization, software-defined networks and software-defined storage [28].

Virtualization is a significant technology to provide the basic building blocks for cloud environments to enhance their flexibility and agility and it is the base for most high-performing clouds. Recently, virtualization has been used in many types of applications including isolation, fault tolerance, intrusion tracking, monitoring, execution replay and software impact analysis. Virtualization components include a virtual memory system (VMs) and a variable module management system (VMMs) [27]. In addition, Software-defined Networking (SDN) was the first software defined resource that allows the management and control planes to be separated from the routing hardware and operated by remote software, providing an increased ability to control and optimize networking [45]. It also allows applications to realize routing, traffic monitoring and access control, and server load balancing tasks [30]. According to Kreutz et al (2014), they stated that "*Software-Defined Networking (SDN) is an emerging networking paradigm that gives hope to change the limitations of current network infrastructures*". The main goal of promotion and adoption of SDN through open standards development, and the SDN breaks the vertical integration by separating the network's control plane from the underlying routers and switches that forward the data plane [31]. OpenFlow is one of most common control protocols in SDN [12].

Software-defined systems will bring new governance issues related to enterprise data centres. These issues require a novel approach to governance and operations management in SDS. This approach aims to enable seamless integration of high-level governance objectives and strategies with concrete operations processes. On the other side, it enables performing operational governance processes for SDS in such a manner that they are feasible in practice. Therefore, to achieve an effective governance approach, the organizations should clearly define the data governance pillars, namely: roles, responsibilities and policies. This will help the organizations to maintain control of their data in the data centre, and achieve security functions for their data.

4.2 Software-defined Storage (SDSt)

Software-defined storage (SDSt) is one of the most important subsystems in SDS, and it is one of the technology trends in the cloud computing industry. It takes the responsibility of a huge data management in storage systems through isolating the data control layer from the data storage layer. The academic research on SDSt is still in its infancy due to the novelty of this area, while there is little work about this subject in industry research, for example IBM, VMWare and Microsoft are seen as the leaders through the literature [27].

Software-defined storage is fundamentally different storage that what is typically used today. VMware defines SDSt as "*the dynamic composition of storage services, aligned on application boundaries, driven by policy*" [27]. Software-defined storage solutions are needed to enable high-level dynamic policies for storage service requirements to be easily enforced, reducing the complexity required by administrators in managing multiple resource paths over multiple layers.

In 2013, Thereska et al. proposed IOFlow, an architecture that enables end-to-end policies in data centers [58]. The policies specify the handling of IOflows from virtual machines to shared storage. Flows are named using a four-tuple comprising human-friendly high-level identifiers, which are VMs, operations, files and shares. IOFlow comprises three components:

- A logically centralized controller that discovers data plane stages and maintains a stage-level data centre topology graph.
- Data plane queues allow for differentiated handling of IO requests, this stage exposes a simple control interface that specifies the low-level identifiers that can be used to direct requests to queues.
- They specify a simple interface between the controller and control applications.

In addition, computer storage architecture is entering its third major epoch, driven by the needs of virtual computing. These changes require a fundamental transformation of data governance methods and strategies, but in ways that will make IT administrator's lives easier with better service level delivery. Software-defined storage offers many benefits for businesses, which include faster time to value, better return on IT spend, no vendor lock-in, IT staff that innovate and efficiencies [53]. Building a SDSt solution also requires taken into consideration some points to overcome the challenges that face data

governance solutions in cloud computing. Some of these challenges include:

- Aggregation of the data,
- Data protection method,
- Data security and availability,
- Ownership,
- Data migration,
- Reliability of the data,
- Data type support,
- Monitor data in the system, and
- Data integration.

Software-defined storage will be changing an organizations strategy to maintain control of their data in cloud computing through bringing new concepts that are data-aware technologies. This technology helps organizations identify which of their files contain elements like personally identifiable information; which folders can be deleted or archived to improve storage management; and how to find misplaced or deleted files instantly. Thus, this probably need to update the data governance strategy in the organisation, and address new factors for data governance to be fits to cloud computing.

In addition, SDSt has a positive impact on data governance for cloud computing, and it will solve some challenges in traditional storage solutions when organizations adopt cloud computing. For example, nearly all traditional storage solutions require a storage administrator to create virtual storage devices for the application to use**.** The storage administrator is deploying data services for the data that is stored on these devices. In many cases, each data service requires its own administration interface. Changing those data services affects all the data stored on those virtual devices and can become problematic. This has a negative effect on the consumer to adopt cloud storage.

On the other hand, SDSt offerings allow applications and data producers to manage the treatment of their data using the storage infrastructure without intervention from storage administrators and with automatic service level management. This will assist an organization to build a strong data governance strategy and methodology for documenting and implementing business rules and controls around data in cloud computing.

We should note that not all organizations will adopt patterns for SDSt at the same rate, or in the same way. Therefore, one helpful way of distinguishing different IT organizations is through their preferred automation model, a good proxy for their progression towards a software-defined environment [27]. However, implementing data governance for all organisations also is not at the same rate, and it will be implemented based on the organisations goals. In SDS, the organisations should consider the characteristics of this technology in its data governance strategy to maintain control of their data in the cloud environment, and to ensure no loss of governance in their data. Implementing the data governance program should be based on a collaboration between the cloud consumer and the provider. Therefore, a data governance program should be implemented automatically to ensure roles, responsibilities and policies of data governance are applied correctly, and ensure no loss of governance in the data. Figure 6 proposes the high-level framework for the data governance path in SDSt. This framework consists of: a user, SDSt, a cloud storage administrator and software developers.

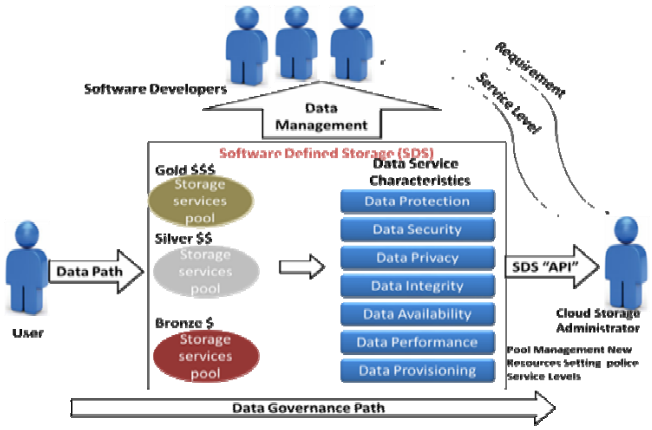


Figure 6. Data governance path in software-defined storage (SDSt)

User. This refers to the cloud consumer who owns the data and it responsible to generate the data and define the data governance requirements.

Software defined storage. This is an approach to data storage in which the programming that controls storage-related tasks are decoupled from the physical storage hardware. In this layer, the user data will be managed and it will consider implementation of data governance requirements for all data service characteristics.

Cloud storage administrator. This is the cloud storage administrator who is responsible for managing user data and implementing the data governance for the user data regarding user requirements and the service level agreement.

Software developers. The software developers are the group responsible for managing and monitoring the user data in SDSt. They also are responsible for monitoring the data governance implementation by the cloud storage administrator to establish an effective data governance program. They also request any update on data governance requirements from the cloud storage administrator, and the cloud storage administrator implements that based on the service level agreement.

Implementation wise, an SDStorage controller unit will mainly be responsible for controlling and managing the storage resources in the system such as local databases and cloud storage. All the configurations of the storage resources are generated

inside this unit. Moreover, this

unit is responsible for monitoring the available storage resources. Furthermore, this unit creates a function table which stores the mandatory information for all storage hosts. More details about this type of controller have been developed and explained in our previous work [53-54], also illustrate in Figure 7 and Figure 8. The abstracted architecture of the SDStorage unit consists of three sub-layers: infrastructure layer, control layer and application layer. The infrastructure layer combines various storage devices storing the raw data. The controller in the control layer is considered the most critical element in SDStorage systems. It is where the storage resources in the infrastructure layer interact with the application layer. The control layer converts different policies to different instructions inside the system. The last layer in this architecture is the application layer which holds different applications and allows the end user to interact with storage devices.

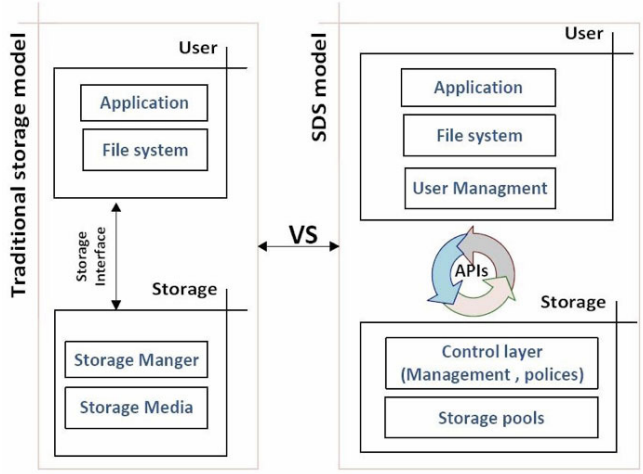


Figure 7. A design comparison between the traditional storage model and the SDStore model [53]

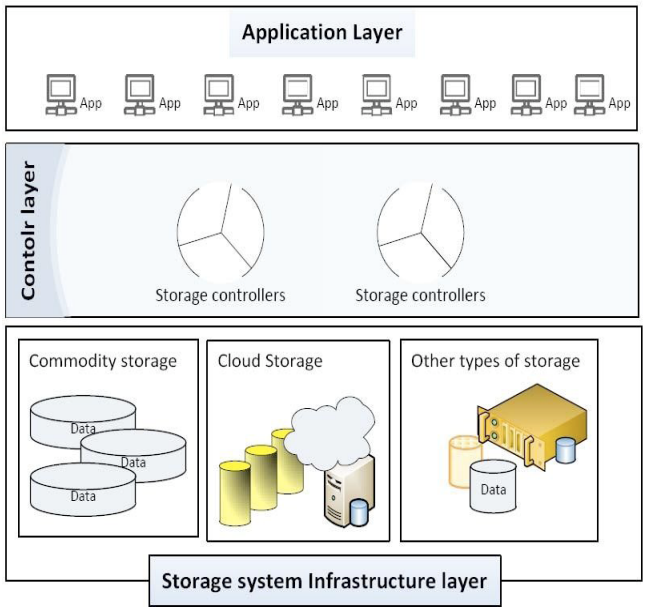


Figure 8. The SDStore system architecture [53]

4.3 Data Governance in the IoT and Cloud Converged Environments

The convergence of the Internet of Things (IoT) with the cloud has been a subject of research interest [40]. Evidence suggests that such a convergence carries huge potential, albeit with some challenges. There is consensus that privacy, security and governance are key concerns. One central issue is the lack of mature governance and security standards for data within the IoT-Cloud converged environments [41].

The IoT possesses huge potential, in particular when considered along with the opportunities ushered in by technologies such as cloud computing. The IoT enables the incorporation of numerous heterogeneous end systems and subsets of data, thus enabling the development of digital services, such as smart cities, smart cars, and the network of things visions [40]. When considered in line with other technologies such as BigData, the IoT results in the exponential growth of data, underpinning growth in productivity and innovation.

Nowadays, one of the greatest difficulties experienced by IT professionals is to anticipate the negative and, or positive impacts that daily operations can cause in the organization. Governance and security are the most challenging issues in the IoT [41]. Therefore, there are many policy issues that relate to governance which need to be addressed if the IoT is to be accepted by society, and wanted to make a difference where it can. The concepts of the IoT governance are also not fully defined and various definitions have been proposed by different industry and research organizations. So, to archive data security in the IoT when interacting with a cloud computing environment, this needs to focus on:

- processes such as policies, responsibilities, and roles as the pillars for governance.
- defining the aforementioned pillars as determinants of the functions and strategies for securing the IoT data within the IoT-Cloud converged environments.

In our previous work, we proposed the framework for data governance and security for the IoT converged cloud computing [42]. The proposed framework aims to create a process layer between the IoT and the cloud computing layers. The authors consider data governance and security as important in ensuring that the IoT data remains secure, private and of an acceptable standard. The IoT-Cloud converged environments involve three different layers as illustrated in Figure 9.

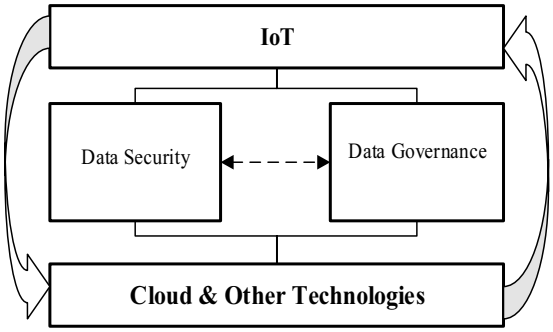


Figure 9. IoT-Cloud converged environment's three layers under consideration [42]

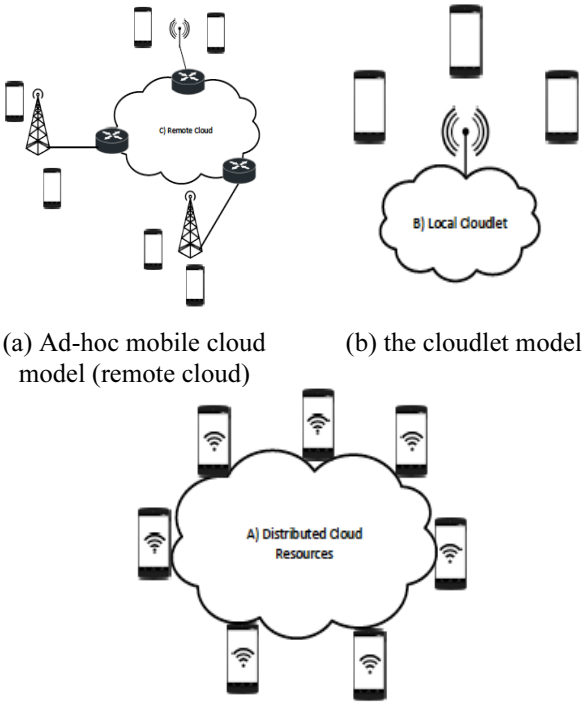
However, we suspect that as more IoT-Cloud converged domains continue to evolve, the roles, responsibilities and policies will remain central to governance and security processes and procedures. Therefore, this framework can be extended in an integrated architecture, to elaborate governance and security rules for stakeholder processes. Nonetheless, we believe that the IoT-Cloud converged environment requires more work from the research community to achieve data governance and security.

4.4 Emergence of Data Governance Policies on Mobile and Social Media Platforms

Today, technology changes rapidly, and generally speaking, organizations are braced for it. Social media technology has become the trending technology used by organisations to engage with their customers [43]. Social media refers to, "web-based tools and services that allow users to create, share, rate and search for content and information without having to log in to any specific portal site or portal destination" [30]. However, mobile phones are increasingly used to access the internet, which has increased usage of social media sites, which no longer need access to a personal computer [44]. Therefore, social media technology is qualitatively different to other technologies for many reasons. First, it has become a force for businesses to reckon with at breath-taking speeds. Second, its effects are far-reaching across the entire spectrum of business activity, from product development to marketing and sales to customer support. Third, organizations do not always have a choice about engaging in social media if their customers are already doing so. So, with the explosive growth and usage of mobile and social platforms, it is necessary for global business operators to now turn their attention to stringent data governance policies on these networks. Without the proper framing and implementation of good governance, mobile and social data can easily get out of control. Again, data stewardship teams can play a key role in monitoring data governance policy violations in these networks. Therefore, the important trends in this technology are that each organisation should build a strong data governance strategy to implement an effective cloud data governance program for mobile and social media platform.

4.5 Mobile Cloud Computing

Whilst mobile computing (e.g smart-phones, tablets) is arguably the end-user computing usage paradigm of choice. Cloud computing is arguably becoming the foremost computing paradigm for service infrastructure and delivery. Mobile Cloud Computing (MCC) is the inevitable aggregation of the aforementioned paradigms. The concept involves creating a cloud-computing environment upon a localized cluster of mobile devices. This is in contrast to cloud environments operating upon a remote data centre. Such a technology has the ability to profoundly change mobile computing usage paradigms. This crossover of paradigms may be found in a number of different forms and models. These models are illustrated in Figure 10. In the traditional cloud model, the negotiation process would be relatively straightforward. As the resources of a typical cloud are seemingly unlimited, there would be minimal contest between clients. However, as these models change, the communication channels employed will vary, the resources available become more contested and the negotiating actor interaction changes, causing the system to become more complex. A cloudlet system such as in [45, 56], or one that operates on another extreme such as in [46-47] will remain similar to the traditional cloud, albeit with slightly more constrained resources. Whereas a mobile cloud is radically different, with the resources being enabled by the devices themselves.



(a) Ad-hoc mobile cloud model (remote cloud)

(b) the cloudlet model

(c) Distributed mobile cloud model (Distributed cloud resources) [57]

Figure 10. Differing mobile cloud models

Although the aforementioned are at the intersection of both fields, they interact in different ways and therefore are posed by different and individual challenges, mainly related to data governance. Therefore, specific data governance strategies appropriate for each mobile cloud model, need to be developed and deployed. To the best of the authors knowledge, this remains a novel area of research with lot of potential on almost all areas of data governance. For instance, storage and processing of data differ from one model to another, and hence appropriate policies and procedures are required to govern these interactions in terms of data.

5.6 Edge Computing and Data Governance

Next generation Mobile Cloud Computing systems will require a paradigm shift in how they are constructed and managed. Current deployment and management platforms are facing considerable challenges regarding flexibility, coverage, dependability and security that next generation systems must handle. Recently, Mobile Edge Computing (MEC) technology emerged as viable solution to provide services to the mobile users within their access range [48-49]. This enables a seamless access to a resources-rich system with high bandwidth network connections. MEC is perceived as a natural evolution to the previously emerged technologies for deploying mobile cloud services such as cloudlet. The increasing demand of the mobile applications for high computing and storage capacities with free user mobility made cloudlet an inefficient solution for end user workload offloading. This is due to the limited resource capabilities of the cloudlet servers and the short coverage range of the cloudlet Wi- Fi connection [49]. The seamless mobile task offloading and execution is crucial to the application latency and the quality of services provided to the users [50]. Many new applications can utilize MEC system such as compute-intensive video encoding [51] and the local streaming service [52]. These applications argue that MEC systems provide the required seamless task execution on the mobile cloud system. MEC marks another trend in addressing data governance for data storage, which is still no researched, but expected to pose many challenges. In our previous work [55, 57], we have proposed a software defined mobile edge computing, which could be a starting stage too to be enhanced by predefining a data governance strategy including necessary policies, standards and procedures.

5 Conclusion

Cloud storage is one of the most important services provided by cloud computing. In the following years, cloud storage will be growing within the private and public sectors in many countries. Cloud storage is more beneficial and has more advantages than the earlier traditional storage system, especially in scalability, cost reduction and portability. Data governance aspects are essential to assist an organization to trust cloud computing when they move their data to cloud storage. Security and data loss prevention are also significant components of data governance in cloud computing. The area of data governance for cloud computing is still under researched and poorly practiced in industry. Mobile cloud computing in its different models and edge computing as well as IoT are all newly emerging technologies, promising to be the evolution from what is now called traditional cloud. These emerging technologies pose real challenges related to data governance; many of them are still not well understood, due to their newly emerging concept which are still under-researched. These new cloud models have been discussed as new trends for addressing future data governance in cloud storage. Software-defined storage has also been discussed in further details as it could be another research trend in addressing future cloud storage solutions. We illustrated how software defined storage could be cross cutting technology which can be implemented for any of the aforementioned cloud models. We emphasized the need for predefining a coherent data governance strategy to be a base line for implementing any cloud-based storage solution. Future work, will attempt to develop a strategic framework for data governance in cloud storage for different models and capitalize on our previous and pioneering research work in software defined systems, including storage, to ensure data governance driven implementation.

References

[1] D. Xu, H. Liu, Reviewing Some Cloud Computing Platforms, *The Second International Symposium on Networking and Network Security (ISNNS 2010)*, Jinggangshan, China, 2010, pp. 161-164.

[2] H. Trivedi, *Cloud Adoption Model for Governments and Large Enterprises*, Unpublished MSc Thesis, Massachusetts Institute of Technology, Citeseer, MA, 2013.

[3] W. Jansen, T. Grance, *Guidelines on Security and Privacy in Public Cloud Computing*, National Institute of Standards and Technology Special Publication 800-144, December, 2011.

[4] S. Sengupta, V. Kaulgud, V. S. Sharma, Cloud Computing Security-Trends and Research Directions, *2011 IEEE World Congress on Services*, Washington, DC, 2011, pp. 524-531.

[5] C. M. Bulla, S. S. Bhojannavar, V. M. Danawade, Cloud Computing: Research Activities and Challenges, *International Journal of Emerging Trends & Technology in Computer Science,* Vol. 2, No. 5, pp. 206-214, September-October, 2013.

[6] A. Rahumed, H. Chen, Y. Tang, P. Lee, J. Lui, A Secure Cloud Backup System with Assured Deletion and Version Control, *2011 40th International Conference on Parallel Processing Workshops*, Taipei, Taiwan, 2011, pp. 160-167.

[7] H. Jeong, J. Park, An Efficient Cloud Storage Model for Cloud Computing Environment, *International Conference on Grid and Pervasive Computing*, Hong Kong, China, 2012, pp. 370-376.

[8] Y. Li, L. Guo, Y. Guo, *CACSS: Towards a Generic Cloud Storage Service*, *CLOSER*, 2012.

[9] R. A. P. Rajan, S. Shanmugapriyaa, Evolution of Cloud Storage as Cloud Computing Infrastructure Service, *IOSR Journal of Computer Engineering (IOSRJCE),* Vol. 1, No. 1, pp. 38-45, May-June, 2012.

[10] I. Arora, A. Gupta, Opportunities, Concerns and Challenges in the Adoption of Cloud Storage, (*IJCSIT) International Journal of Computer Science and Information Technologies,* Vol. 3, No. 3, pp. 4543-4548, May-June, 2012.

[11] M. Sookhak, H. Talebian, E. Ahmed, A. Gani, M. K. Khan, A Review on Remote Data Auditing in Single Cloud Server: Taxonomy and Open Issues, *Journal of Network and Computer Applications,* Vol. 43, pp. 121-141, August, 2014.

[12] W. Braun, M. Menth, Software-Defined Networking using OpenFlow: Protocols, Applications and Architectural Design Choices, *Future Internet*, Vol. 6, No. 2, pp. 302-336, June, 2014.

[13] Norton Internet Security, *Data Loss Prevention*, http://www.nortoninternetsecurity.cc/2011/03/data-loss-prevention.html.

[14] T. Sivashakthi, N. Prabakaran, A Survey on Storage Techniques in Cloud Computing, *International Journal of Emerging Technology and Advanced Engineering,* Vol. 3, No. 12, pp. 125-128, December, 2013.

[15] V. Khatri, C. V. Brown, Designing Data Governance, *Communications of the ACM,* Vol. 53, No. 1, pp. 148-152, January, 2010.

[16] Cloud Security Alliance, *Security Guidance for Critical Areas of Focus in Cloud Computing V3.0*, https://downloads.cloudsecurityalliance.org/initiatives/guidance/csaguide.v3.0.pdf.

[17] M. Manoochehri, R. Glushko, Standards and Governance in Organizing Systems, *Governance: An International Journal of Policy, Administration and Institutions*, Vol. 23, pp. 1-24, 2010.

[18] A. Rajathi, N. Saravanan, A Survey on Secure Storage in Cloud Computing, *Indian Journal of Science and Technology,* Vol. 6, No. 4, pp. 4396-4401, April, 2013.

[19] R. K. L. Ko, P. Jagadpramana, M. Mowbray, S. Pearson, M. Kirchberg, Q. Liang, B. S. Lee, TrustCloud: A Framework for Accountability and Trust in Cloud Computing, *2011 IEEE World Congress on Services,* Washington, DC, 2011, pp. 584-588.

[20] A. S. Weber, Cloud Computing in Education in the Middle East and North Africa (MENA) Region: Can Barriers be Overcome?, *Conference Proceedings of eLearning and Software for Education (eLSE)*, Bucharest, Romania, 2011, pp.565-570.

[21] A. Mathew, *Survey Paper on Security & Privacy Issues in Cloud Storage Systems*, EECE 571B, Term Survey Paper, April, 2012.

[22] S.-H. Kim, I.-Y. Lee, Study on User Authority Management for Safe Data Protection in Cloud Computing Environments, *Symmetry,* Vol. 7, No. 1, pp. 269-283, March, 2015.

[23] S. N. Kumar, Cryptography during Data Sharing and Accessing Over Cloud, *International Transaction of Electrical and Computer Engineers System,* Vol. 3, No. 1, pp. 12-18, March, 2015.

[24] V. S. Kushwah, A. Saxena, A Security Approach for Data Migration in Cloud Computing, *International Journal of Scientific and Research Publications,* Vol. 3, No. 5, pp. 1-8, May, 2013.

[25] F. Liu, J. Tong, J. Mao, R. Bohn, J. Messina, L. Badger, D. Leaf, *NIST Cloud Computing Reference Architecture*, National Institute of Standards and Technology Special Publication 500-292, September, 2011.

[26] J. Sen, Security and Privacy Issues in Cloud Computing, in: A. Ruiz-Martinez, R. Marin-Lopez, F. Pereniguez-Garcia (Eds.), *Architectures and Protocols for Secure Information Technology Infrastructures*, IGI Global, 2013, pp. 1-45.

[27] L. Tawalbeh, Y. Haddad, O. Khamis, F. Aldosari, E. Benkhelifa, Efficient Software-Based Mobile Cloud Computing Framework, *2015 IEEE International Conference on Cloud Engineering (IC2E),* Tempe, AZ, 2015, pp. 317-322.

[28] R. Buyya, R. N. Calheiros, J. Son, A. V. Dastjerdi, Y. Yoon, Software-Defined Cloud Computing: Architectural Elements and Open Challenges, *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI),* Delhi, India, 2014, pp. 1-12.

[29] A. Darabseh, M. Al-Ayyoub, Y. Jararweh, E. Benkhelifa, M. Vouk, A. Rindos, A Novel Framework for Software Defined based Secure Storage Systems, *Simulation Modelling Practice and Theory*, Vol. 77, pp. 407-423, September, 2017.

[30] C. Monsanto, J. Reich, N. Foster, J. Rexford, D. Walker, Composing Software Defined Networks, *the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, Lombard, IL, 2013, pp. 1-13.

[31] D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, S. Uhlig, Software-Defined Networking: A Comprehensive Survey, *Proceedings of the IEEE,* Vol. 103, No. 1, pp. 14-76, January, 2015.

[32] E. Thereska, H. Ballani, G. O'Shea, T. Karagiannis, A. Rowstron, T. Talpey, R. Black, T. Zhu, IOFlow: A Software-defined Storage Architecture, *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles*, Farmington, PA, 2013, pp. 182-196.

[33] A. Parakh, S. Kak, Online Data Storage using Implicit Security, *Information Sciences*, Vol. 179, No. 19, pp. 3323-3331, September, 2009.

[34] Y. Jararweh, O. Al-Sharqawi, N. Abdulla, L.A. Tawalbeh, M. Alhammouri, High-Throughput Encryption for Cloud Computing Storage System, *International Journal of Cloud Applications and Computing (IJCAC)*, Vol. 4, No. 2, pp.1-14, 2014.

[35] J. Al-Badarneh, Y. Jararweh, M. Al-Ayyoub, M. Al-Smadi, R. Fontes, Software Defined Storage for Cooperative Mobile Edge Computing Systems, *Fourth International Conference on Software Defined Systems (SDS)*, Valencia, Spain, 2017, pp. 174-179.

[36] B. Lynn, A*uthenticated Identity-Based Encryption*, IACR Cryptology ePrint Archive: Report 2002/07, June, 2002.

[37] G. Wei, X. Yang, J. Shao, Efficient Certificateless Authenticated Asymmetric Group Key Agreement Protocol, *KSII Transactions on Internet and Information Systems*, Vol. 6, No. 12, pp. 3352-3365, Decemebr, 2012.

[38] S. More, S. Chaudhari, Third Party Public Auditing Scheme for Cloud Storage, *Procedia Computer Science*, Vol. 79, pp. 69-76, 2016.

[39] K. Yang, X. Jia, *Security for Cloud Storage Systems*, Springer-Verlag, 2014.

[40] M. Al-ruithe, E. Benkhelifa, K. Hameed, Key Dimensions for Cloud Data Governance, *The IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud 2016)*, Vienna, Austria, 2016, pp. 379-386.

[41] M. K. A. Hassan, Governance, Risk and Compliance "GRC" for Internet of Things "IOT", *International Journal of New Technology and Research*, Vol. 2, No. 3, pp. 148-152, March, 2016.

[42] M. Al-ruithe, S. Mthunzi, E. Benkhelifa, Data Governance for Security in IoT & Cloud Converged Environments, *IEEE/ACS 13th International Conference on Computer Systems and Applications (AICCSA 2016)*, Agadir, Morocco, 2016, pp. 1-8.

[43] T. Gillespie, Governance of and by Platforms, in: J. Burgess, A. E. Marwick, T. Poell (Eds.), *Sage Handbook of Social Media*, Sage, 2017, pp. 254-278.

[44] Government of India, *Framework & Guidelines for Use of Social Media for Government Organisations*, pp.1-38, April, 2012.

[45] L. Tawalbeh, Y. Jararweh, F. Ababneh, F. Dosari, Large Scale Cloudlets Deployment for Efficient Mobile Cloud Computing, *Journal of Networks*, Vol. 10, No. 1, pp. 70-76, January, 2015.

[46] M. Quwaider, Y. Jararweh, A Cloud Supported Model for Efficient Community Health Awareness, *Pervasive and Mobile Computing*, Vol. 28, pp. 35-50, June, 2016.

[47] M. Quwaider, Y. Jararweh, An Efficient Big Data Collection in Body Area Networks, *2014 5th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2014, 1-6.

[48] M. T. Beck, M. Werner, S. Feld, T. Schimper, Mobile Edge Computing: A Taxonomy, *AFIN 2014- The Sixth International Conference on Advances in Future Internet*, Lisbon, Portugal, 2014, pp. 48-54.

[49] A. Ahmed, E. Ahmed, A Survey on Mobile Edge Computing, *10th IEEE International Conference on Intelligent Systems and Control, (ISCO 2016)*, Coimbatore, India, 2016, pp. 1-8.

[50] E. Ahmed, A. Gani, M. K. Khan, R. Buyya, S. U. Khan, Seamless Application Execution in Mobile Cloud Computing: Motivation, Taxonomy, and Open Challenges, *Journal of Network and Computer Applications*, Vol. 52, pp. 154-172, June, 2015.

[51] M. Beck, S. Feld, A. Fichtner, C. Linnhoff-Popien, T. Schimper, Me-volte: Network Functions for Energy-efficient Video Transcoding at the Mobile Edge, *2015 18th International Conference on Intelligence in Next Generation Networks (ICIN),* Paris, France, 2015, pp. 38-44.

[52] O. Makinen, Streaming at the Edge: Local Service Concepts Utilizing Mobile Edge Computing, *2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies*, Cambridge, UK, 2015, pp. 1-6.

[53] A. Darabseh, M. Al-Ayyoub, Y. Jararweh, E. Benkhelifa, M. Vouk, A. Rindos, SDstorage: A Software Defined Storage Experimental Framework, *IEEE International Conference on Cloud Engineering (IC2E 2015)*, Tempe, AZ, 2015, pp. 341-346.

[54] Y. Jararweh, M. Al-Ayyoub, A. Darabseh, E. Benkhelifa, M. Vouk, A. Rindos, Software Defined Cloud: Survey, System and Evaluation, *Future Generation Computer Systems*, Vol. 58, pp. 56-74, May, 2016.

[55] Y. Jararweh, A. Doulat, A. Darabseh, M. Alsmirat, M. Al-Ayyoub, E. Benkhelifa, SDMEC: Software Defined System for Mobile Edge Computing, *2016 IEEE International Conference on Cloud Engineering Workshop (IC2EW)*, Berlin, Germany, 2016, pp. 88-93.

[56] Y. Jararweh, A. Doulat, O. AlQudah, E. Ahmed, M. Al-Ayyoub, E. Benkhelifa, The Future of Mobile Cloud Computing: Integrating Cloudlets and Mobile Edge Computing, *2016 23rd International Conference on Telecommunications (ICT)*, Thessaloniki, Greece, 2016, pp. 1-5.

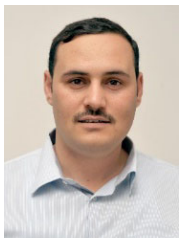
[57] Y. Jararweh, M. Alsmirat, M. Al-Ayyoub, E. Benkhelifa, A. Darabseh, B. Gupta, A. Doulat, Software-Defined System Support for Enabling Ubiquitous Mobile Edge Computing, *The Computer Journal*, Vol. 60, No 10, pp. 1443-1457, October, 2017.

[58] C.-M. Yu, C.-Y. Chen, H.-C. Chao, Proof of Ownership in Deduplicated Cloud Storage with Mobile Device Efficiency, *IEEE Network Magazine*, Vol. 29, No. 2, pp. 51-55, March, 2015.

Biographies

Majid Al-Ruithe is a Ph.D. student of Computer Science at Staffordshire University, U.K, with extensive experience in working with KSA industry and public sector on real world business problems. His research interests include cloud data governance, cloud computing, IoT and data science. He is the author of over 10 publications. He is a member of the Mobile Fusion Applied Research Centre and Cloud Computing and Applications Research Group at Staffordshire University.

Elhadj Benkhelifa is a Professor of Computer Science at the Staffordshire University. He is the Founding Head of the Cloud Computing and Applications Research Lab. He was the Faculty Director of the Mobile Fusion Applied Research Centre (2014-2016). He is a member of several research committees within the university and externally. He has co-founded and chaired a number of international conferences and workshops and edited a number of conference proceedings and journals special issues. Elhadj esearch interest spans cross the areas of cloud computing, IOT, Software Engineering and applied soft computing



Yaser Jararweh received his Ph.D. in Computer Engineering from the University of Arizona in 2010. He is currently an associate professor of computer sciences at Jordan University of Science and Technology. He has co-authored several technical papers in established journals and conferences in fields related to cloud computing, HPC, SDN, security and Big Data. He is co-chairing many IEEE events such as ICICS, FMEC, SDS, MCSMS, IoTNAT, CCSNA, OSNT,, SNAMS, BDSN, IoTSMS, ISCW, and many others.

Chirine Ghedira is a Full Professor of computer science at Jean Moulin Lyon 3 University, Lyon, France. She is the president of the committee of experts in Mathematics Computing at the the university of Lyon 3 and responsable for Masters course in management information Systems. She is the deputy head of the SOc research group, part ofth eLIRIS-CNRS (UMR5205) Lab. Chirine is also a core member of the doctoral school InfoMaths at the university. She is also a core member of the VAE/VAP commission of iaelyon. Chirine's research interest include service oriented architecture, Cloud Computing, Web services, Security & Privacy, Data Quality and quality of services, Integration and data management in distributed systems