Cloud Computing and the Enterprise Needs for Data Freedom

Dalia Kriksciuniene
Vilnius University
Muitines str. 8, Kaunas, Lithuania
dalia.kriksciuniene@khf.vu.lt

Donatas Mazeika
Vilnius University
Muitines str. 8, Kaunas, Lithuania
donatas.mazeika@khf.stud.vu.lt

Abstract— **The article aims to investigate the problem, if the willingness and success of transferring enterprise operations to the cloud-computing infrastructure are related to the level of freedom of managing enterprise data. This problem emanates from the raising awareness of the enterprises, that submission of data for processing by the cloud computing solutions invokes risk of further transformations of data formats by vendor, and failure of keeping its compatibility to own data. In this case the enterprise could be unable to make archives, or to switch to another cloud service provider. The pilot survey explores attitudes of enterprise members and specialists of information technologies (IT) to the designed shadow service and is potential to increase customer trust for cloud computing services. The concept of shadow service architecture is suggested, which models the customized backup plan adjusted for managing data of the company as the compatibility matrix.**

Keywords-cloud computing; vendor lock-in; trust; compatibility matrix; shadow service architecture

I. INTRODUCTION

The main idea of cloud computing is to transfer on-site solutions, based on ownership and management of hardware and software products, to the IT services that can be accessed by Internet. Cloud computing solution providers have started massive marketing for enterprises encouraging them to shift from traditional IT to software on-demand. Implementation of new software solution is quite long and expensive process, while cloud computing offers service provision with lower initial costs, quick launch, and flexibility of software and hardware resources.

Cloud computing issues are widely analyzed and discussed in scholarly literature. In [1], authors define cloud computing as a large-scale distributed computing paradigm that is driven by economies of scale. Cloud computing is claimed to be a new step in internet computing that provides large perspectives, but at the same time it raises issues in the architecture, design, and exploiting of existing networks and data centers [4]. In [3] Sakr, indicates that cloud computing is the transformation of IT from a product to a service. Many authors agree that this new type of computing brings new opportunities to enterprise, such as reduced costs, improved scalability, increased customization [3], experimenting with new services, ability to improve usage by adding more capacity at peak demand, and removing unexploited capacity [4]. Together with the new promising advantages, there have emerged issues of security, data privacy, technical risks, lack of system integration, which should be solved for building trust of users.

The aim of this work is to analyze enterprise needs for ability to control its own data in cloud solutions and to investigate if services from cloud computing vendors should be modified in order to increase trust of the end-users and to avoid threat of vendor lock-in.

The following section presents cloud-computing definitions, operating principles, risks and classification of products that are used in business practice. In the third section, the Software as a service (SaaS) type of cloud solutions is explored by conducting case study, which addresses needs of the end-users and analyses how the enterprise can manage its own data on cloud server. In the fourth section the concept of shadow service is substantiated. It aims to increase trust of user by enhancing flexibility of data monitoring. The findings of the research are summarized in the conclusion section of the article.

II. BENEFITS AND RISKS OF CLOUD COMPUTING

Cloud computing refers to the various software and hardware solutions that are provided by independent vendors for access via Internet. These solutions are easily scalable and simple to start using. The payments for these services are usually made for subscription. Cloud computing offers significant computing capability and economy of scale, and recently cloud computing has become a new promising mode of business computing [5][15].

Cloud computing services can be classified to three main layers, reflecting different purposes of use: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [3][8]:

- Software-as-a-Service (SaaS) layer applications are hosted by cloud computing providers and are available to customers over Internet, such as CRM, ERP, project management systems, document management systems, office suite programs etc. These solutions are targeted for business and home users.
- Platform-as-a-Service (PaaS) layer is targeted at software developers' needs, it offers both development environment and tools as a service.
- Infrastructure-as-a-Service (IaaS) layer delivers platform virtualization environment as a service [5]. Target users are system administrators, who analyse the needs for resources and ensure computing power [4].

 ISBN: 978-1-61208-154-0

IaaS and PaaS are mainly used by developers and Independent Software Vendors (ISV) [6], while SaaS services are targeted for enterprise use [4]. As company data are the critical sources of competitive advantage, the enterprises are very serious about their safety and effective use, without any obstacles emanating from the chosen cloud service.

SaaS applications can serve enterprise end-users, such as CEO, managers, administrators, for managing various business processes. The leading SaaS cloud-computing vendors are Salesforce.com, Google, IBM. The most popular SaaS solutions are the Google Apps office tools products, Salesforce.com, SugarCRM (CRM- customer relationship management systems), LotusLive (web-based collaboration tools). SaaS solutions have wide selection of functions [3]. In most cases they can be compared to analogous traditional software. They are easy to start, as they do not require specific programming or administration knowledge and could be especially suitable for small and medium enterprises that lack financial and human resources for investing to IT infrastructure: installing and maintaining hardware infrastructure and software applications [15]. By using "pay as you go" subscription model the enterprise can avoid costs of starting capital, and the running costs can be further regulated by subscribing resources and services that company needs at the time. Other benefits include scalability, reliability, security, ease of deployment, and ease of management for customers [7].

Despite of many benefits the question about getting back company's data is left open. If company subscribes some cloud services, it could be hard or impossible to make backup of all data and ensure that data could be reused of moved to other cloud vendor. One of the final conlusions of the extensive survey of 125 participants from over 100 organisations by the KPMG consulting company states that cloud computing vendors are establishing their own, partly incompatible, standards to complicate integration with other vendor's solutions [10].

Salesforce.com can serve as a case study of managing company data in cloud. In 2010, it had over 87 200 customers and was recognized as a world leading CRM provider by Gartner [12]. All the services provided by Salesforce.com are based on cloud computing. Its most known product is CRM system for sales force automation. The possibilities for the enterprise to manage its own data by using Salesforce.com CRM system can be evaluated as limited:

- It provides export possibility of all company data once a week only for subscribers of highest priced versions - Enterprise and Unlimited Editions.
- For Professional Edition this feature has to be ordered for additional fee.
- Group and Contact Manager editions do not have such feature [13].
- The provided backup is flat file format without any object relations [13].
- The additional possibility is to take enterprise data from Salesforce.com with the help of third party applications from AppExchange (application shop), yet without any guarantee that they are compatible to the current version of cloud solution for backup of company data.

The Saleforce.com case shows that one of the main disadvantages of cloud computing is the threat of locking company data's lock at one vendor and lack of integration with other systems.

Major threats of cloud computing can be summarized to the following risk categories [2]:

- Policy and organizational risks: lock-in, loss of governance, compliance challenges, loss of business reputation, and cloud service termination or failure;
- Technical risks: unavailability of service, resource exhaustion, intercepting data in transit, data transfer bottlenecks, and distributed denial of service;
- Legal risks: subpoena and e-discovery, changes of jurisdiction, data privacy, and licensing.

Policy and organizational risks can lead to difficulty of extracting data from the cloud service, and this is important reason why some companies refuse start using it [2]. It is recomended that cloud computing customers should have an alternative location for services, and the cloud provider would give proper data backup to ensure continuity even if the cloud computing provider went broke or acquired and swallowed up by a larger company [5].

Technical risks can cause short-term disorders such as services unavailability due to server or connection failure, resource exhaustion, or Denial of Service (DDoS) attacks. Major technical risk is loss of Internet connectivity [15]. Although the technical risks can be considered similar to the generic Internet-related issues, but their impact for performing business processes in cloud environment is crucial, as it ads two additional sources of risk from cloud vendor and his internet provider to the existing risks emanating at the customer side and his internet service.

Legal risks are likely to arise due to a customer data keeping in different countries. Due to different legislative systems there are risks that company's data was unintentionally disclosed, the centralization of storage and shared tenancy of physical hardware imparts more risk of unwanted data disclosure, especially in cases of cloud vendor hardware being confiscated by law enforcement agencies or through civil suits [2]. Cloud computing providers are free to offer confusing privacy policies that can be unilaterally modified at any time without notice to users [14]. It is important that users could be confident to their data security and availability upon request in order to be able to change cloud vendor for resolving any type of risk.

III. FEASIBILITY SURVEY

The pilot survey was conducted in order to reveal the point of view of customers and potential users as related to the vendor lock-in problem, to discuss issues which hinder trust, and provide insights to feasibility of new shadow service which could ensure availability of data backup from cloud upon the request from customer. This pilot survey was

 ISBN: 978-1-61208-154-0

targeted to IT-professionals, analysts, cloud providers, and users. The survey questions included:

- Expectations about cloud computing at the enterprise (Likert-type scale was applied [18]).
- Awareness of data management and level of control in cloud.
- The preference list of software solutions for most likely transfer to cloud according to trust in full control of enterprise data.

The respondents for survey were accessed virtually via business information systems related groups of LinkedIn social network site. The pilot survey consisted of 10 questions, 19 people responded to the survey. The majority of respondents (63%) indicated that they belong to IT-related industry, the other persons were from telecommunications, education, engineering, architecture, finance, banking, and insurance industries. Half of the respondents were IT professionals (53,3%), managers had a share of 26,6%, other professions were represented by one person each. The majority (75%) of respondents use SaaS cloud solutions (SalesForce.com, Google Apps, etc.). The IaaS cloud service layer was the second by popularity (18,8%). The remaining respondents did not use cloud services yet.

The question if correspondents knew about their data management abilities in cloud computing was answered differently: 47,6% respondents were fully aware of the enterprise data format applied for data download from cloud service vendor. 17,6% respondents did not know if they could safely get back their data from cloud, the remaining respondents had no practical experience in this sphere.

The perceived benefits and disadvantages of cloud computing were explored by Likert-type answer scale [18]. The highlighted benefits (strongly agree answer) were: increased mobility (56,30%), quick deployment (43,80%) and software flexibility (43,80%). Reduced costs were indicated as important benefit as well (agree answer – 43,80%). Majority of participants agreed that each type of concerns (lock-in in one vendor, cloud service termination or failure, possible unavailability of service, legal and data privacy risks) were equally important. Cloud service termination or failure was marked as one of the most important concerns (agree 41,18%; strongly agree 11,76%). The result shows that the enlisted risks are not yet resolved in satisfactory way and should be addressed in future.

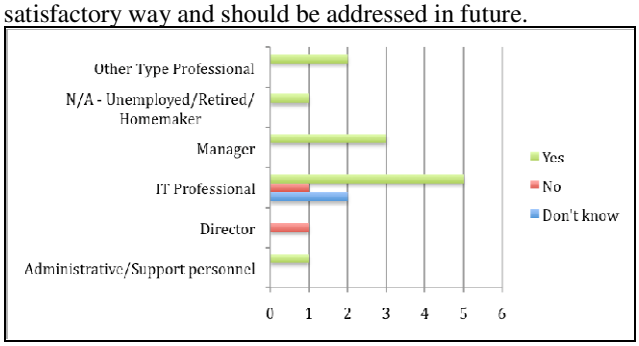

Figure 1. Question: "Does the full control of company data would increase the trust line of cloud computing?" (Distribution of answers by job roles)

The answers to question "Does the full control of company data would increase the trust line of cloud computing?" and their distribution by job roles are presented in Fig.1. The positive answers were given by 76 % of participants, who agreed that this aspect of cloud computing is very important. Only two respondents (IT professionals) answered that they don't know if the full control data management would reduce the risks. Three respondents (two IT professionals and director) expressed doubt of necessity of full data control.

The remaining two questions were to determine which of IT solutions of the enterprise the participants would transfer without hesitation to the infrastructure of cloud computing, and what types of solutions could be transferred only in case of ability to fully control company's data.

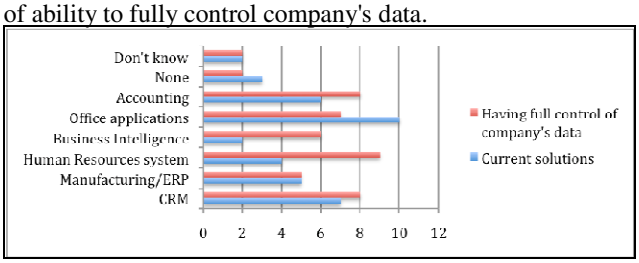

Figure 2. Question: "Which local IT solutions would you transfer to cloud?"

From Fig. 2, it is evident, that full control of company's data participants would encourage transfer of more sophisticated and customized enterprise-based IT solutions to cloud, such as accounting software, business intelligence, human resources system, CRM. Office applications were most easily considered for use in current cloud solutions.

The comments of the participants given to the question "What improvements are needed (if needed) to transform cloud computing to inevitable future model of IT?" revealed their confidence that cloud computing is the future model of IS. Security, safety, control were mentioned as absolute necessity. Less marketing, more transparent services, reviews and free market were recommended as well.

The survey confirmed that the users and potential customers are aware of benefits of cloud computing solutions for enterprise, but are strongly concerned about risks. The risks of lock-in in one vendor, cloud service termination or failure, possible unavailability of service, legal and data privacy issues were of equal importance to most participants. Software mobility, quick deployment and flexibility were the main drivers which could encourage companies for IT transfer to the cloud. The ability to control data could increase trust to cloud services and subsequently lead to transfer to cloud not only standard, but the important customized IT solutions as well. The idea for enhancement of cloud service is proposed in the following chapter as the shadow service model.

IV. SHADOW SERVICE ARCHITECTURE MODEL

In order to avoid vendor lock-in, cloud service termination or failure, the shadow service idea is suggested. New service is aimed to shadow the compatibility of interests between customer and cloud vendor and always

 ISBN: 978-1-61208-154-0

monitor enterprise data for "ready" status, which means that enterprise can access and backup its data from cloud anytime and in various formats.

In Fig. 4, the Architectural Framework for Cloud Computing is presented. It enhances basic scheme of cloud service, presented in [16]. The component of data compatibility management acts as the shadow service module, which could calibrate the interests of both customer and the cloud vendor. The shadow service has to track the amendments and changes of the cloud environment, which are performed by the vendor and to compare to the customer requirements for data formats. If the updates cloud vendor lead to further transformation of data, the rules have to be created and checked in order to inform the customer, if he can retrieve the data back from cloud in the required format for his in-site use, or possibly transfer it to the other vendor.

The shadow service module serves as a special layer, consisting of data compatibility tables and data compatibility rule engine. The data compatibility tables define the cloud service plan and compatibility matrix. The service plan is built before starting to use cloud services. The plan should include questionnaire, agreement, and other written typical documents for definition of

- Most critical data to business continuity;
- Data or documents that should be accessed without internet connection (synchronizing feature);
- Maximum retrieval time of most critical and complete data.
- Required backup formats (CSV, XML, JSON etc.).

The compatibility matrix creates the layer of interaction between the customer and cloud vendor. It serves for monitoring status of customer data within cloud and is based on preparing specific backup plans for cloud computing customers. The desired importable, processing and exportable data formats for company (or default format selection if it is not known by company) are declared here. The matrix defines the possibilities from the providers' side as well. It is used to trace changes of the customer data which could affect the compatibility of the backup data for using off-line or to transfer them to the other cloud vendor . It serves as a buffer when the connection or other technical risks arise. The shadow service idea aims for optimal data synchronization between cloud servers and company's local computers. As the full data synchronization is difficult to achieve, the priorities have to be specified for data "freshness" and readiness for backup. One of the solutions is suggested in [17] as Re:FRESHiT protocol. It specifically addresses the problems of replication management in data clouds. Re:FRESHiT organizes virtual trees based on the sites' freshness levels, and introduces a routing mechanism for reading data, while at the same time allowing users to specify their own freshness requirements [17].

Compatibility matrix-based layer serves as a set of rules for processing data and managing files according to the request of the customer and capabilities of the cloud service vendor. This feature would generally allow company to get the agreed (or default) format of data before or after using cloud services. The basic structure of compatibility matrix is shown in Fig. 3. Data compatibility management structure consists of three objects, used for describing data formats (importable data, processing data, and exportable data), and aggregated object for data compatibility management. The descriptions of objects of the compatibility matrix are filled-in when customer starts using service (either by asking customer to fill-in the questionnaire, negotiating service plan, or by offering default options).

The preferences of the customer for the importable and exportable data's formats are checked by cloud service vendor. If it is impossible for cloud vendor to import, process or export according to the requested specification, the corresponding rules within compatibility matrix are defined and further used for all company's data backup and for -local data synchronization.

Application of the shadow service and monitoring the compatibility matrix for data management could resolve problems with temporary connection breaks, because all critical data could be backuped according to the defined priorities and available for access from a local computer.

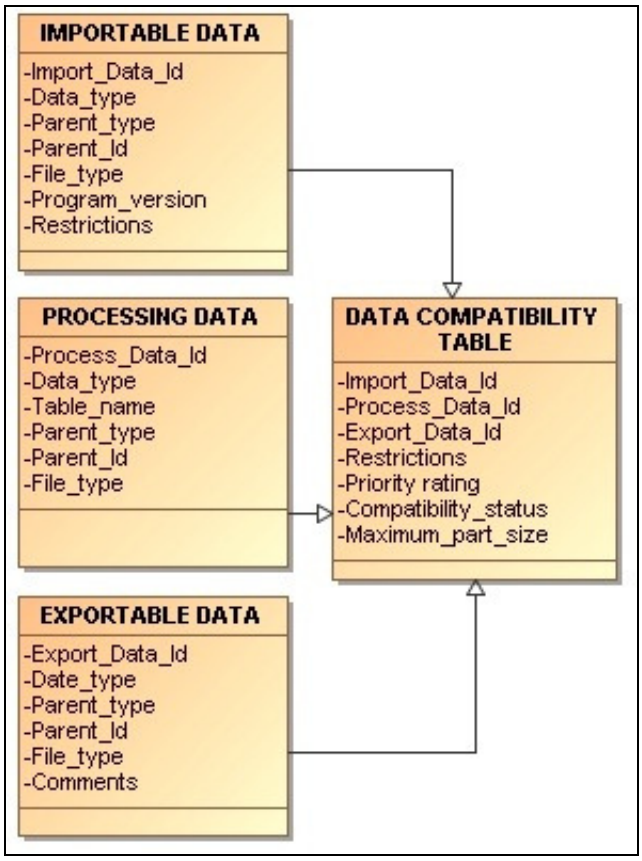

Figure 3. Compatibility matrix structure

The main idea of the compatibility matrix is to maintain importable, processing and exportable data formats, types, relations, backups, priorities by the agreed requirements. Inevitably, the cloud service provider can develop and change its environment by changing data structure or formats, but all these changes should be updated in compatibility matrix and not to violate the existing rules.

 ISBN: 978-1-61208-154-0

Data compatibility management module consists of rule engine and compatibility matrix (Fig. 3). Compatibility matrix is filled-in jointly by the customer the vendor. Some default or standard options can be applied, as well as the customized articles of mutual agreement can be introduced here. The vendor has to check the customer information and requirements about data management, and generate rules for maintaining them. In case some rules become impossible to fulfill due to changes in the cloud environment, the compatibility matrix can show the changes and require new steps from customer and the cloud vendor for altering conditions of the service agreement.

Data compatibility matrix stores information about all types of data:

Importable data: defines, what type of data and files are imported to cloud service:

- The importable data type (Date_type) could be file type or information of data array, which will further be organized to data tables.
- Parent_type and Parent_id: defines how importable data is related with another customer data, for example companies and contacts are related via company key attribute.
- File_type defines the import data source (XLS, CSV etc) or the allowed data format for upload.
- Restrictions define maximum file sizes, and requirements for files information.
- Program_version indicated the software, which was applied for creating original file.

Processing data: combines metadata information about the status of processing customer files in cloud service.

The Process_data_id and Data_type withing cloud environment are defined, changes in organizing data tables and the relations with other data (*Parent_type* and *Parent_id*), file types are set and checked by the vendor for defining rules of data compatibility.

Exportable data: area is calibrated among customer and vendor. The backup formats, relations and downloadable file types are defined.

Data compatibility table aggregated the information about importable, processing and exportable data interdependencies. Any existing restrictions, priority rates for synchronisation, statuses, maximum size and other information are filled considering mutual agreement.

The rule engine in created for managing and implementation of the compatibility matrix. If the agreement contains any particular requirement which is impossible to fulfil, it alters its status. In case some customer requirements for importable/exportable data were initialy confirmed, but later became incompatible due to changes in cloud environment, the status should change to „impossible“ or „partial“. In this case it could inform the customer about the increased risk and lead to further steps from his side or envoke obligations of the vendor to meet the requirements by customizing service.

The shadow service layer has to constantly reveal status of compatibility of interests between customer and vendor, technical status of data and information of any vulnerabilities which could hinder interests fo preserving and managing enterprise data. Transparency of techical compatibility increases customer trust and reduces risk of lock-in as well.

This shadow service enables companies to forecast further IT related strategies in order to fully control strategic data, and to reduce dependence on specific cloud service, because they know the format and integration level of enterprise data available from cloud vendor.

The suggested model brings additional costs for cloud computing vendors due to support of enhanced functionality aimed to constant track of data conversion for fulfilling requests of the customers. Their compensation would be increase in trust from clients, and building potential for solving most urgent disadvantages of cloud computing.

V. CONCLUSIONS AND FUTURE WORKS

Cloud computing vendors provide their customers various IT solutions that are easy to start using, flexible and mobile. Nevertheless the issues of concern dealing with security, data privacy, technical risks related to getting company data back from cloud in the appropriate format should be resolved for successful usage of cloud solutions.

The feasibility survey was conducted for exploring attitudes of the users and potential customers. It showed that main obstacles which hinder usage of service are related to possible cloud service termination or failure and vendor lock-in.

New shadow service architecture was suggested, which is based on creating the technical solution for monitoring agreement between cloud service and the customer vendor for defining priorities, rules and formats for managing data. The data compatibility management model, consisting of compatibility matrix and rule engine is proposed as the framework for enabling shadow service. It allows achieving transparent technical compatibility of interests between customer and services by cloud vendor, and always monitors enterprise data for “ready” status.

The compatibility matrix defines data freshness priorities, formats for data backup. It enables tracking changes performed with the data. The rule engine component enables to inform the customer, if he can retrieve the data back from cloud in the required format and ensures possibility to use the backup data with the local system of the customer and prevent from vendor lock-in situation.

The requirement for enhancement of cloud services by monitoring the shadow service layer via compatibility matrix could increase the workload and costs. But it can increase trust of customers and encourage them to transfer to cloud the most specific and strategic solutions of the company such as accounting software, customer relationship management (CRM), and human resource management systems.

REFERENCES

[1] I. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud computing and grid computing 360-degree compared,“ Grid Computing Environments Workshop (GCE '08), 2008.

[2] T. Betcher, “Cloud Computing: Key IT-Related Risks and Mitigation Strategies for Consideration by IT Security Practitioners,” 2010.

 ISBN: 978-1-61208-154-0

[3] M.F. Sakr, “Cloud Computing/Virtualization,” The SANS (SysAdmin, Audit, Network, Security) Institute, 2010.

[4] G. Pallis, “Cloud Computing The New Frontier of Internet Computing, “ ISSN: 1089-7801, 2010.

[5] J. Yang, “Cloud Computing Research and Security Issues,” Computational Intelligence and Software Engineering (CiSE), 2010 International Conference

[6] L. Qian, Z. Luo, Y. Du, and L. Guo, “Cloud Computing: An Overview,” 2009, pp. 626-631, DOI: 10.1007/978-3-642-10665-1_63.

[7] H. Erdogmus, “Cloud Computing: Does Nirvana Hide behind the Nebula?, “ 2009, DOI: 10.1109/MS.2009.31

[8] M. Ahrens, “Cloud Computing and the Impact on Enterprise IT,” 2010, DOI: 10.1007/978-3-642-15877-3_16

[9] Salesforce, “Benefits of SaaS, “, 2010. <http://www.salesforce.com/saas/benefits-of-saas/> [last access 22/07/2011]

[10] KPMG, “KPMG is a global network of professional firms providing Audit, Tax and Advisory services,“ 2010, <http://www.kpmg.com/Global/en/WhoWeAre/Pages/default.aspx> [last access 22/07/2011]

[11] J. Schofield, “Freedom to move data is vital when it's in the clouds,“ 2009, http://www.guardian.co.uk/technology/ 2009/jun/17/cloud-computing- jack-schofield> [last access 22/07/2011]

[12] Salesforce, “Salesforce.com Positioned as a Leader in the Magic Quadrant for Sales Force Automation,“ 2010, https://www.salesforce.com/company/news-press/press-releases/2010/08/100825.jsp [last access 22/07/2011]

[13] Salesforce, “Salesforce Pricing & Editions - Sales Cloud,” 2010, http://www.salesforce.com/crm/editions-pricing.jsp, [last access 22/07/2011]

[14] R. Sprague, “Cloud Privacy: Normative Standards for Information Privacy Management within Cloud Computing. University of Wyoming College of Business,” 2009.

[15] K.R. Choo,”Cloud computing: Challenges and Future Directions” in Trends&Issues in Crime and Criminal Justica Vol.400, October 2010. Australian Institute of Criminology, 2010.

[16] R.Clarke, ”User Requirements for Cloud Computing Architecture,” Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing CCGRID '10, IEEE Computer Society, 2010, doi>10.1109/CCGRID.2010.20

[17] T. Cristiana, H. Schuldt, Y. Breitbart, and H. Schek, „Flexible Data Access in a Cloud based on Freshness Requirements," Proceedings of the 2010 IEEE 3rd International Conference on Cloud Computing CLOUD '10, IEEE Computer Society, 2010, doi>10.1109/CLOUD.2010.75

[18] R. Likert, "A Technique for the Measurement of Attitudes" in Archives of Psychology Vol 14, 1932: p.p. 1–55.

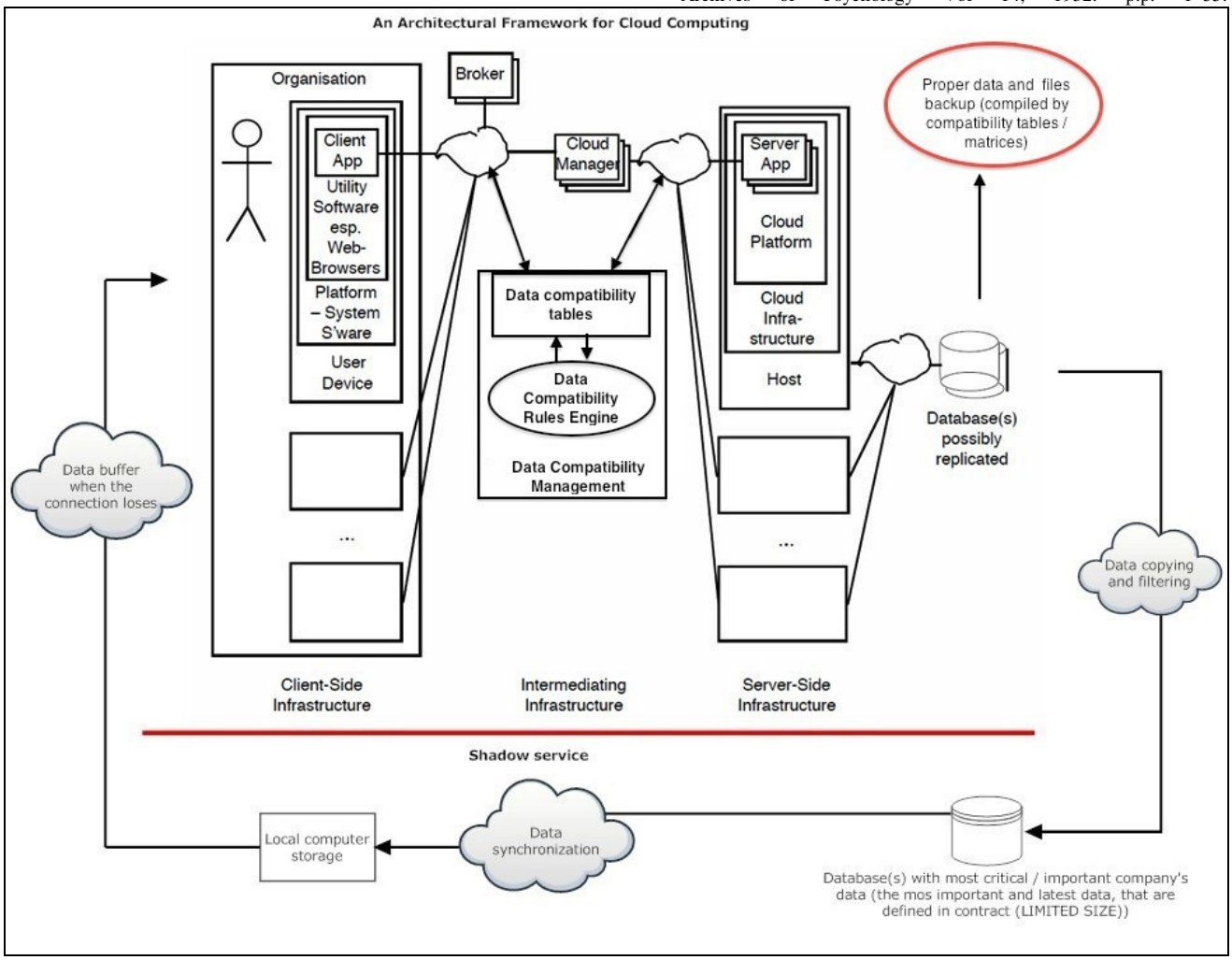

Figure 4. An Architectural Framework for Cloud computing service with compatibility layer for shadow service