Practical implementation of new algorithm for restricting data fusion in cloud computing with use of information Kalman filtering

Mohamadreza Mohamadzadeh

Department of Electronic Engineering, Science and Research branch,Islamic Azad University, Neyshabur,Iran

m_mohamadzadeh.talent@yahoo.com
mohamadreza_mohamadzadeh@hotmail.com

Received: June 28, 2019. Revised: July 15, 2021. Accepted: October 16, 2021. Published: November 26, 2021.

***Abstract*—These days' lots of technologies migrate from traditional systems into cloud and similar technologies; also we should note that cloud can be used for military and civilian purposes [3]. On the other hand, in such a large scale networks we should consider the reliability and powerfulness of such networks in facing with events such as high amount of users that may login to their profiles simultaneously, or for example if we have the ability to predict about what times that we would have the most crowd in network, or even users prefer to use which part of the Cloud Computing more than other parts – which software or hardware configuration. With knowing such information, we can avoid accidental crashing or hanging of the network that may be cause by logging of too much users. In this paper we propose Kalman Filter that can be used for estimating the amounts of users and software's that run on cloud computing or other similar platforms at a certain time. After introducing this filter, at the end of paper, we talk about some potentials of this filter in cloud computing platform. In this paper we demonstrate about how we can use Kalman filter in estimating and predicting of our target, by the means of several examples on Kalman filter. Also at the end of paper we propose information filter for estimation and prediction about cloud computing resources.**

Keywords- **Information Technology, Information filter Cloud computing and its resources, Security, Kalman Filter.**

I. INTRODUCTION

Cloud computing introduces changes that necessitate a re-assessment of legal risks. The cloud provider contract on offer must be examined in detail and favourable and constructive terms negotiated with the provider to ensure that they are appropriate to the work that your institution carries out. Legal responsibility for technology use and for data, particularly personal data, remains with the institution despite using remote services such as cloud computing. Whether data is 'at rest' within a cloud service or 'in transmission' to, from or within the cloud service, it is necessary to clarify the security aspect of such information [12].

Risk with technology usage is not new but because cloud technology changes how data is managed and processed particular risks need to be clarified and re-assessed. These include questions with regard to information ownership, responsibility for data protection compliance, what law applies when a dispute arises, as well as issues such as access and scrutiny in terms of law enforcement. Software as a Service (SaaS) provides complete applications hosted by a cloud provider and delivered over the internet. Entire administrative, operational and research capabilities can be provided online. Resources are shared but data and access capabilities are segregated within the application offering economies of scale. For IT directors considering outsourcing software as a service from a legal point of view, ensuring that security is robust is probably the starting point. Awareness rising for individual users is also a consideration as it is necessary to inform users by whom their data is being processed and for what purposes [12].

Infrastructure as a Service (IaaS) is the delivery of computer hardware (servers, networking technology, and storage and data centre space) as a service. The service is typically paid for on a usage basis. All cloud infrastructures depend on virtualisation. This includes the aggregation and partitioning of computing resources across multiple data centres enabling cloud providers to manage their capacities more efficiently. This inevitably means that servers and software are processing data for many different data controllers simultaneously. When IT directors are considering using such services, an assessment of the risks should include how resilient the service is, for example, in terms of availability and response to demand and an examination of the cloud provider's security measures [12].

With Platform as a Service (PaaS) the cloud provider delivers more than infrastructure. It delivers capabilities to manage all software development stages from planning and design to building and deployment to testing and maintenance. Once again, concerns with regard to access and data security are inherent in the relationship and therefore need to be addressed in the service contract at the outset [12].

As a manager in an FE or HE institution you will know that the Freedom of Information (FOI) legislation gives individuals a right of access to information held by the institution. The legislation covers all records and information held whether digital or print, current or archived. For senior decision makers considering introducing cloud services it is important therefore to address this issue with any cloud provider and ensure that specific data, if requested, can be retrieved within twenty working days. Even though, the information is stored in the cloud, an institution will still be deemed to be holding it for the purposes of FOI. This converts into a legal requirement to ensure that access is possible and that such incidents as outages and failures at the cloud provider's end do not prevent the institution fulfilling its legal obligations to respond with information as requested [12].

What jurisdiction new information is created in will have a bearing on how ownership of it is determined and if an institution is conducting research on virtual computers it is advisable that the issue of ownership is clarified and detailed in the service contract before the work takes place. Thus in addition to discrete information outputs, it is necessary for the institution in the service contract to address and agree ownership of information which can be derived or deduced from how the cloud data is used [12].

There are many occasions when information is required to be kept confidential by administrative staff or researchers at an institution. This will include handling personal health data, some types of employment related data and management related data that may be sensitive commercially. Before entering into a cloud service agreement you as a senior manager will want the proposed systems tested to ensure that confidential data can be processed without being compromised [12].

The Data Protection Act 1998 governs how the personal data of individuals is processed. Institutions, as data controllers, are required to ensure that all processing of personal data for which they are responsible is fair and lawful, even where the data processing is carried out by a cloud provider. The legal obligations fall on the senior office holders at the institution to make sure that any cloud provider that is processing its personal data has appropriate security practices and procedures in place [12].

Cloud technology can be central to the policy-making process in many important areas, including digital inclusion, education, e-government, aging, and employment. Microsoft believes the industry can achieve more through cooperation than through isolation. To this end, we encourage governments to support and foster multi-stakeholder partnerships that leverage accessible technology and the cloud to address key policy challenges and deliver positive benefits for individuals and society at large [12].

By coordinating policy development with industry leaders, assistive technology vendors, nonprofits, and consumers, governments can promote increased investment in both the Cloud and accessible technologies so that all Americans, including the millions with disabilities, benefit.

Under s.49 RIPA, properly authorised persons (such as members of the law enforcement, security and intelligence agencies) may serve notice on an institution requiring the disclosure of protected (e.g. encrypted) information which they lawfully hold, or are likely to, in an intelligible form. S.49 limits the information to which the right to serve such a notice applies but an example could be material seized by police under a judicial warrant or intercepted under a warrant authorised in accordance with Chapter I, Part I of RIPA [12]. An FE or HE institution receiving the appropriate notice must comply with it by disclosing the information specified in the notice in an intelligible form or by disclosing any key to the information which is in their possession. It is now necessary to have appropriate procedures in place to carry this out should data be hosted in the cloud [12].

II. HOW CAN WE ENHANCE INNOVATION IN THE CLOUD COMPUTING

To enhance innovation in the cloud, governments should facilitate movement of data across borders while maintaining legal protection for consumers [12].

Like the Internet, cloud services are global in nature. Being able to move data among large data centers in multiple geographic areas allows cloud computing providers to pool IT resources and consolidate overheads and purchasing power; this, in turn, results in significant cost and efficiency benefits for consumers, as well as the environmental benefits that flow from using fewer data centers. From an operational standpoint, cloud computing providers move data between data centers in order to offer key services to customers, including 24 hour technical support and round-the-clock product development. Data transfer likewise is essential to data back-up and resiliency. As noted in a recent report by the Lloyd’s insurance market, “The digital world is still susceptible to physical disasters such as flooding, earthquakes and hurricanes,” and thus “geographic concentration” of data may increase risk of loss. The cloud provides a perfect vehicle for ensuring that critical information does not disappear forever as a result of natural or man-made disasters [12].

Rules governing the transfer of data and information across borders, however, do not accommodate the current realities of broadband-enabled computing. While not their intention, these rules limit the innovation and economic development otherwise made possible by the cloud, and often do not produce any corresponding benefit to consumer privacy. As the European Commission has recognized, “there is a general need to improve the current mechanisms for international transfers of data” in light of the vastly increased delivery of services over the Internet since the Data Protection Directive was adopted 15 years ago. The Directive as it now stands broadly restricts the transfer of personal data from within Europe to any country whose domestic laws do not provide a level of protection that the EU considers “adequate.” In practice, only those countries – less than 10 to date – that provide the same precise methods of protection as the EU have been deemed adequate. Other governments go even further and impose near-complete bans on certain types of data transfer, such as in Nova Scotia and British Columbia. There, most personal data held by public bodies cannot be moved to any jurisdiction outside of Canada [12].

Regardless of the nature of an unduly strict cross-border data restriction – whether it is the result of an express prohibition on data export, a limitation based upon an adequacy requirement, or inconsistent laws across jurisdictions – the unintended consequence is to depress investment, reduce trade, and deprive consumers and enterprises of the benefits of cloud computing and other innovations. Cloud service providers subject to inflexible cross-border data restrictions are forced to implement cumbersome and expensive processes in order to legitimize the data transfers, even when more pragmatic procedures could provide the same or even a better level of protection to users. Alternatively, the provider may be forced to store the data locally in the jurisdiction that imposes the export restriction, thereby preventing the provider from being able to offer customers the cost and efficiency benefits that stem from being able to move data to multiple geographic areas, and eliminating the potential energy efficiencies and environmental benefits of consolidating resources in fewer data centers. In a nutshell, the desire for local data centers is in conflict with the efficiencies associated with the scale economics of cloud computing [12].

Cloud Computing is a model for enabling ubiquitous, 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. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider [12].

The IT industry has recently been included as a significant factor in global greenhouse gas (GHG) emissions, at approximately 2-3% of emissions, roughly equivalent to the international airline industry [IUSE, 2009]. An ever increasing innovation cycle with shorter product lifecycles, exponentially increasing demand for data and processing power, and more energy intensive processing have fueled this emissions increase. We suggest a strategy that can substantially dematerialize the industry by increasing computing utilization rates through centralization of computing power, and distribution of computing as a service. This concept, cloud computing, has been emerging for the past decade, but only to a certain degree, and without any evidence of dematerialization. As an academic demonstration, we seek to show the potential that cloud computing has to fully service the computing industry, significantly reducing material and energy consumption while enhancing performance and productivity [12].

The concept of cloud computing can be implemented to varying degrees. In its most simple and existing form, it offers online data storage on a remote server that can be accessed via the internet. The next level of the cloud is where not only data but applications are accessed via the cloud, known as software as a service (SAAS) or platform as a service (PAAS) deployment [Right Scale]. This eliminates the need to install and manage the hardware-software interface internally and can ease the burden of providing sufficient computing power and IT management for a firm. Companies such as Google, Amazon, and Microsoft have begun investing in this type of infrastructure, and Google Docs is a simplified application of this concept. The ultimate level of cloud computing, known as infrastructure as a service (IAAS) is where software, operating systems, and server hardware and infrastructure are all managed as a service within the cloud. Computing resources can be distributed amongst one or many remote servers and computers, and delivered via the internet to the end user [12].

The potential energy, material, time and cost savings for this ultimate form of cloud computing are vast. The utilization rates of a common desktop in a business environment are between 10-20% [Zhou, 2009]. By centralizing the computing power in the cloud, the need for end-user processor power is minimized, and utilization rates can be vastly increased to as high as 80-90% [Zhou, 2009]. Not only is this far more energy efficient, but reduces the material needs for procuring local machines. Instead of high performance desktops, thin- client machines can be used that are vastly simpler, lighter smaller, and less energy intensive. Programs like Windows Remote Desktop and GoToMyPC.com are the closest thing to true cloud computing that exists today, however these services are not intended to necessarily replace your desktop, but to complement it. Coordination between machine makers, software designers, and internet infrastructure and bandwidth providers will have to continue to develop, so that a seamless IASS service can eliminate the need for desktops [12].

The economies of scale of a national scale cloud computing infrastructure are a complex system to simulate. For the purposes of our analysis, we chose to compare the life cycle of a collection of 100 traditional desktops vs. a network of 100 thin-client machines (w/server allocation) to simulate a micro-scale version of a cloud. This 100 client cloud system is comparable to the computing needs of a small business. Small businesses often do not have the capital to invest in internal computing infrastructure and IT, and cloud computing would perfectly suit their needs. Companies such as Right Scale are targeting these kinds of customers for their cloud services. Full cloud computing for larger firms is less realistic, as they have their own IT capabilities, higher utilization rates, and security concerns that would discourage cloud use, at least in its infancy stages before security and lock-in issues are resolved [Right Scale]. The scope of our analysis is small, and it should be recognized that economies of scale in a true cloud infrastructure would further enhance the potential economic and environmental savings [12].

Our analysis looks at the life cycle of a thin-client + server, desktop, from material procurement, precomponent and component production, final product assembly, use, and disposal. Our quantitative analysis focuses on manufacturing and use phase environmental and economic costs. Quantitatively we only consider global warming potential (GWP) as an environmental indicator while waste, water, and toxicity indicators are described qualitatively. Our economic scaling factor could be used for a quantitative comparison for the remaining environmental indicators. This is defensible as the major difference between the two supply chains is the use of less material, which reduces costs over almost every link of the supply chain [12].

As these shifts pick up steam, Microsoft believes it is important to expand the vision of an accessible computing ecosystem to encompass technologies, tools, products, and services that enable all people to benefit from the cloud and to more easily control the devices around them to live a more independent life. As a starting point, people will always need a smart client device with built-in accessibility capabilities to serve as a gateway to the cloud and to perform tasks offline or when no Internet access is available. When browsing the Internet, accessing social networking sites, or tapping into cloud services, users will use applications or a web browser such as Internet Explorer to access and retrieve information. These devices also enable individuals who may require AT applications or specialty computing hardware that enables text to be read, video to be captioned, or that facilitates hands-free entry of information [12].

Over time, we envision a world of accessible, cloud-connected devices powered by Microsoft technologies that an individual can rely on to understand and interpret their needs, preferences, and immediate surroundings to create an adaptive experience. In combination with reliable broadband connectivity and the virtually unlimited computing power of the cloud, it will be possible for these preferences to follow users wherever they go and to supplement the capabilities of any particular device in ways that increase communication and collaboration and help reduce social isolation at work and at home [12].

III. AN INTRODUCTION INTO KALMAN FILTER

The **Kalman filter**, also known as **linear quadratic estimation** (LQE), is an algorithm which uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those that would be based on a single measurement alone. More formally, the Kalman filter operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state. The algorithm works in a two-step process: in the prediction step, the Kalman filter produces estimates of the current state variables, along with their uncertainties. Once the outcome of the next measurement (necessarily corrupted with some amount of error, including random noise) is observed, these estimates are updated using a weighted average, with more weight being given to estimates with higher certainty. Because of the algorithm's recursive nature, it can run in real time using only the present input measurements and the previously calculated state; no additional past information is required [14].

The Kalman filter uses a system's dynamics model (e.g., performances of Cloud user's in this model), known as control inputs to that system - Cloud, and multiple sequential measurements (such as from controlling and monitoring the movements of users between different pages and databases) to form an estimate of the system's varying quantities (its state) that is better than the estimate obtained by using any one measurement alone. As such, it is a common controlling and monitoring fusion and data fusion algorithm. Basic equations in Kalman filter are as follows [14]:

$$\tilde{X}_{t|t-1} = F_t \breve{X}_{t-1\,|t-1} + B_t U_t$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t$$

$$\breve{X}_{t|t} = \breve{X}_{t|t-1} + K_t \left(Y_t - H_t \breve{X}_{t|t-1}\right)$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1}$$

$$K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}$$

All measurements and calculations based on models are estimates to some degree. Noisy monitoring data – data that are obtained from motions of users, approximations in the equations that describe how a system changes, and external factors that are not accounted for introduce some uncertainty about the inferred values for a Cloud state. The Kalman filter averages a prediction of a Cloud state with a new measurement using a weighted average. The purpose of the weights is that values with better (i.e., smaller) estimated uncertainty is "trusted" more. The weights are calculated from the covariance, a measure of the estimated uncertainty of the prediction of the Cloud state. The result of the weighted average is a new state estimate that lies in between the predicted and measured state, and has a better estimated uncertainty than either alone. This process is repeated every time step – every preferred time to can avoid surveillance actions and/or avoid permeating hackers into Cloud databases, with the new estimate and its covariance informing the prediction used in the following iteration [14].

To this point we obtained general information about Kalman filtering and cloud computing, from now on we want to purpose: "Information Filter" for using in predicting and updating information about presence of hacker in such networks [14].

IV. Information filter:

In the information filter – Kalman Information filter, or inverse covariance filter, the estimated covariance and estimated state are replaced by the information matrix and information vector respectively. These are defined as [14]:

$$\mathbf{Y}_{k|k} = \mathbf{P}_{k|k}^{-1}$$

$$\hat{\mathbf{y}}_{k|k} = \mathbf{P}_{k|k}^{-1}\hat{\mathbf{x}}_{k|k}$$

Similarly the predicted covariance and state have equivalent information forms, defined as [14]:

$$\mathbf{Y}_{k|k-1} = \mathbf{P}_{k|k-1}^{-1}$$

$$\hat{\mathbf{y}}_{k|k-1} = \mathbf{P}_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1}$$

As have the measurement covariance and measurement vector, which are defined as:

$$\mathbf{I}_k = \mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k$$

$$\mathbf{i}_k = \mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{z}_k$$

The information update now becomes a trivial sum [14].

$$\mathbf{Y}_{k|k} = \mathbf{Y}_{k|k-1} + \mathbf{I}_k$$

$$\hat{\mathbf{y}}_{k|k} = \hat{\mathbf{y}}_{k|k-1} + \mathbf{i}_k$$

The main advantage of the information filter is that *N* measurements can be filtered at each time step simply by summing their information matrices and vectors [14].

$$\mathbf{Y}_{k|k} = \mathbf{Y}_{k|k-1} + \sum_{j=1}^{N}\mathbf{I}_{k,j}$$

$$\hat{\mathbf{y}}_{k|k} = \hat{\mathbf{y}}_{k|k-1} + \sum_{j=1}^{N}\mathbf{i}_{k,j}$$

To predict the information filter the information matrix and vector can be converted back to their state space equivalents, or alternatively the information space prediction can be used [14].

$$\mathbf{M}_k = [\mathbf{F}_k^{-1}]^{\mathrm{T}}\mathbf{Y}_{k-1|k-1}\mathbf{F}_k^{-1}$$

$$\mathbf{C}_k = \mathbf{M}_k[\mathbf{M}_k + \mathbf{Q}_k^{-1}]^{-1}$$

$$\mathbf{L}_k = I - \mathbf{C}_k$$

$$\mathbf{Y}_{k|k-1} = \mathbf{L}_k\mathbf{M}_k\mathbf{L}_k^{\mathrm{T}} + \mathbf{C}_k\mathbf{Q}_k^{-1}\mathbf{C}_k^{\mathrm{T}}$$

$$\hat{\mathbf{y}}_{k|k-1} = \mathbf{L}_k[\mathbf{F}_k^{-1}]^{\mathrm{T}}\hat{\mathbf{y}}_{k-1|k-1}$$

Note that if *F* and *Q* are time invariant these values can be cached. Note also that *F* and *Q* need to be invertible [14].

V. Conclusion:

In this article, we tell about basic definitions and concepts of cloud computing. Also we tell about uses of this technology in different societies and industries and also about how they use it and about their future plans. An important factor in using cloud computing and migrating into this network is its security and durability. In this paper the authors propose Kalman filtering for increasing security and durability of such networks for the first time. If we implement this algorithm on such networks – for example on the edge of such networks, we can estimate and predict the amount of users that use the resources – software and hardware resources – at anytime. Also we can estimate and predict the amount of users that logging onto a certain account, and by the means of that we can avoid surveillance entering of bad users – we estimate the location of user by the previous location and its background data. Also we can use it for estimating the amount of user that use a certain application on such networks, and by knowing that amount we can improve power of our network to be able to support our users. Furthermore by using this algorithm we can increase the security of this technology, by estimating and predicting the point of presence of bad users. In this paper we demonstrate about how we can use Kalman filter in estimating and predicting of our target, by the means of several examples on Kalman filter. Also at the end of our paper we purpose and discuss about information filter that can be used for estimation and prediction in our network.

References:

[1] Darbandi "Applying Kalman Filtering in solving SSM estimation problem by the means of EM algorithm with considering a p ractical example"; published by the Journal of Computing – Springer, 2012; USA.

[2] http://www.cs.unc.edu

[3] mohamadreza mohamadzadeh " An overview of the effects of processing on C loud Computing dramatic present and provide new security solutions "; published by the life science journal - **marsland press,** 2013; USA

[4] http://info.acm.org/pubs/toc/CRnotice.html

[5] Microsoft's Accessible Technology Vision and Strategy; September 2011.

[6] *www.wikipedia.org*.

[7] C.G. Atkeson and J.M. Hollerbach. 1985. "Kinematic features of unrestrained vertical arm movements," Journal of Neuroscience, 5:2318-2330.

Ali Azarbayejani and Alex Pentland. June 1995. "Recursive Estimation of Motion, Structure, and Focal Length," *IEEE Trans. Pattern* Analysis and Machine Intelligence, June 1995, 17(6).

[8] Ronald Azuma and Mark Ward. 1991. "Space-Resection by Collinearity: Mathematics Behind the Optical Ceiling Head-Tracker," UNC Chapel Hill Department of Computer Science technical report TR 91-048 (November 1991).

[9] Ronald Azuma and Gary Bishop. 1994. "Improving Static and Dynamic Registration in an Optical See-Through HMD," SIGGRAPH Conference Proceedings, Annual Conference Series, pp. 197-204, ACM SIGGRAPH, Addison Wesley, July 1994. ISBN0-201-60795-6.

[10] Ronald Azuma. 1995. " Predictive Tracking for Augmented Reality," Ph.D. dissertation, University of North Carolina at Chapel Hill, TR95-007.

[11] Ted J. Broida and Rama Chellappa. 1986. "Estimation of object motion parameters from noisy images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, January 1986, 8(1), pp. 90-99.

[12] R. G. Brown and P. Y. C. Hwang. 1992. *Introduction to Random* Signals and Applied Kalman Filtering, 2nd Edition, John Wiley & Sons, Inc.