Alma Mater Studiorum – Università di Bologna
In collaborazione con LAST-JD consortium:
Università degli studi di Torino
Universitat Autonoma de Barcelona
Mykolas Romeris University
Tilburg University
e in cotutel con
THE Luxembourg University

DOTTORATO DI RICERCA IN
Erasmus Mundus Joint International Doctoral Degree in Law, Science and Technology

Ciclo 30 – A.Y. 2014/2015

SSD: INF/01 – INFORMATICA
SC: 01/B1 – INFORMATICA

Design and Implementation of Legal Protection for Trade Secrets in Cloud Brokerage Architectures relying on Blockchains

Presentata da: Muhammad Umer Wasim

Coordinatore Dottorato	**Supervisore**
Prof. Giovanni Sartor	Prof. Pascal Bouvry Prof. Monica Palmirani
	Co-Supervisore
	Assoc. Prof. Tadas Limba

Esame finale anno 2018

Alma Mater Studiorum – Università di Bologna
In partnership with LAST-JD consortium:
Università degli studi di Torino
Universitat Autonoma de Barcelona
Mykolas Romeris University
Tilburg University
and in cotutorship with the
THE Luxembourg University

PhD Programme in
Erasmus Mundus Joint International Doctoral Degree
in Law, Science and Technology

Ciclo 30 – A.Y. 2014/2015

SSD: INF/01 – INFORMATICA
SC: 01/B1 – INFORMATICA

Design and Implementation of Legal Protection for Trade Secrets in Cloud Brokerage Architectures relying on Blockchains

Submitted by: **Muhammad Umer Wasim**

PhD Programme Coordinator	**Supervisor**
Prof. Giovanni Sartor	Prof. Pascal Bouvry Prof. Monica Palmirani
	Co-Supervisor
	Assoc. Prof. Tadas Limba

Year 2018

PhD-FSTC-2018-30
The Faculty of Sciences, Technology and Communication

University of Bologna
Law School

DISSERTATION

Defence held on 26/04/2018 in Bologna
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN INFORMATIQUE

AND

DOTTORE DI RICERCA
in Law, Science and Technology

by

MUHAMMAD UMER WASIM

Born on 22nd September, 1979 in Murree (Pakistan)

Design and Implementation of Legal Protection for Trade Secrets in Cloud Brokerage Architectures relying on Blockchains

Dissertation defence committee

Dr Jesus Carretero, Chairman
Professor, University Carlos III of Madrid

Dr Anh Tuan Trinh, Vice Chairman
Professor, Budapest University of Technology and Economics

Dr Pascal Bouvry, Dissertation Supervisor
Professor, University of Luxembourg

Dr Seredynski Franciszek
Professor, Cardinal Stefan Wyszynski University

Dr Monica Palmirani
Professor, University of Bologna

Dr Tadas Limba
Assoc. Professor, Mykolas Romeris University

Acknowledgement

The PhD research presented in this document marks the conclusion of Erasmus+ Joint International Doctoral (Ph.D.) Degree in Law, Science and Technology and Doctoral (Ph.D.) Degree in Informatics (Informatique). The University of Bologna, Italy, and the University of Luxembourg, Luxembourg, within the LAST-JD consortium will award a Ph.D. degree in “Law, Science and Technology” and Ph.D. degree in Informatics (Informatique), respectively. The development and completion of both the degrees would not have been possible without those who have helped me during the research.

Firstly, I would like to express my deepest gratitude to my PhD supervisors: Prof. Pascal Bouvry from Univerisity of Luxembourg and Associate Prof. Tadas Limba from Mykolas Romeris University. I would like to extend my appreciation for their trust, optimism, advice, support, and availability. Overall, this team of supervisors, with their different backgrounds and areas of research, provided me with a very interesting environment for my development as a PhD researcher and as a person. I am also very thankful to Mr. Abdallah A. Z. A. Ibrahim, a PhD researcher at Uniervsity of Luxembourg for providing his support towards my research work which helped me to test my models in an emulated cloud environment.

A special thanks to: Erasmus Mundus Law, Science and Technology PhD (LAST-JD) program for allowing me to pursue this research; Prof. Monica Palmirani, Director of LAST-JD program, Prof. Guido Boella, Vice-Director of LAST-JD program, and Ms. Dina Ferrari, Coordinator of LAST-JD program, for their guidance and continuous support in every steps during the research; and faculties and staff of University of Bologna and University of Turin, Italy, Mykolas Romeris University, Lithuania, and University of Luxembourg, Luxembourg, for their administrative support and academic discussions.

Last and most importantly, I would like to thank my beloved wife and my family. I would not have been able to go through this without their support.

Abstract

This multidisciplinary Ph.D. research focuses on legal protection for trade secrets in the cloud, a topic that is relatively unexplored in the literature. The primary objective was to provide legal protection for trade secrets in the cloud brokerage architecture. However, as per overwhelming evolution of blockchains in the cloud, secondary objective was also included in the research. The latter was to provide legal protection for trade secrets over a blockchain. The following abstract summarizes the research in context of the aforementioned objectives in respective paragraphs.

Data Protection legislation has evolved around the globe to maximize legal protection of trade secrets. However, it is becoming increasingly difficult to prove trade secret violations in cloud context. Embedding legal protection as a preemptive measure could effectively reduce such burden of proof in a court of law, which can be implemented by an online broker in the cloud. The primary aim of this research was to propose a model for an online broker that embeds legal protection as preemptive measure to reduce burden of proof during litigation. This is a novel area of inter-disciplinary research whose body of knowledge is not yet well established. The underlying concept in the proposed model was built upon the notion of factor analysis from the discipline of unsupervised machine learning. For evaluation, two-stage procedure was implemented that showed application of legal protection as preemptive measure and subsequently, reduced burden of proof in a court of law. A real time quality of service based dataset for cloud storage providers (Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, SugarSync, and Zip Cloud) was used for the technical evaluation. The simulation results showed better results of proposed model as compared to its counterparts in the field, which in court of law can be used as a part of evidence to reduce burden of proof. For legal validation of such conclusion, questionnaires were sent to law and ICT experts. There were total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). The sample (5 out of 6 respondents) agreed that results of our model could be used in the court

(or judiciary) as a part of evidence to reduce burden of proof. Theoretically, this part of research (focused on primary aim) is a pioneer effort on providing legal protection to trade secrets in the cloud. Practically, it will benefit an enterprise to negotiate contract with service providers to minimize trade secret misappropriation in the cloud.

However, for enterprise that is using decentralized architecture in the cloud e.g. blockchains, contracts could emerge towards smart contracts (an autonomous software program running over blockchains). In this context, a well negotiated contract will not be a solution to minimize trade secret misappropriation. In fact, for this case it is particularly relevant to instantiate role of judiciary over a blockchain. The secondary aim of this research was to develop a model that can be implemented over the blockchain to automatically issue preliminary injunction (or temporary restraining order by court of law) for the breach of contract that can potentially lead to trade secret misappropriation. This part of the research extended the previously proposed model by using stochastic modeling from the discipline of data science. High performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu) and docker (a software container platform) were used to emulate contractual environment of three service providers: Redis, MongoDB, and Memcached Servers. The results showed that court injunction(s) was issued only for Redis and MongoDB Servers. Technically, this difference could be attributed to the fact that Memcached is simply used for caching and therefore, it is less prone to breach of contract. Whereas, Redis and MongoDB as databases and message brokers are performing more complex operations and are more likely to cause a breach. For legal validation of the results, questionnaires were sent to law and ICT experts. There were total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). The sample (4 out of 6 respondents) disagreed “**ONLY**” using the results of the model by the court of law (or judiciary) to issues a preliminary injunction (or temporary restraining order) for the breach of contract. Theoretically, this part of the research is a pioneer attempt for providing legal protection over the blockchain. Practically, it will benefit blockchain driven enterprises to control and stop breach of contract that can potentially lead to trade secret misappropriation.

In addition to above mentioned applied benefits, following list briefly presents research contributions of this multidisciplinary Ph.D. research in the domain of Law.

- It is first in-line to focus on legal protection for trade secrets in the cloud. A well-established similar concept is “information security”, which provides technical protection for trade secrets in the cloud e.g. encryption, hashing etc.
- In the domain of case law, despite of the jurisdiction constraint i.e. precedents (or court rulings) are binding on all courts within the same jurisdiction, this research is first in-line to use case law together with newly proposed Delphi Sampling method to provide legal protection for trade secrets in borderless online cloud environment.
- It is first in-line to implement notion of “confidentiality by design”, which focuses on a legal person or an enterprise. A well-established similar concept is “privacy by design” that focuses on a physical person or human being.
- By defying the myth that “smart contracts cannot be breached” and in the context of contract law, this research is first in-line to automate role of the court (evidential hearing).

In addition to the above mentioned research contribution in the domain of Law, following list briefly presents research contribution in the domain of ICT.

- In the context of multi-criteria decision analysis, this research is first in-line to identify and analyze noise in the data and solves related issue of structural uncertainty (or misspecification of criteria).
- In the context of machine learning, this research is first in-line to propose “self-regulated multi-criteria decision analysis” that operates without decision maker’s interference and hence, it can be used in the context where automation of decision making process is required.
- In the context of multidisciplinary research, this study is first in-line to

propose a method of Delphi Sampling that seeks inter-disciplinary validation for research results.

Abstract

(Italian Translation)

Questa tesi multidisciplinare di dottorato si focalizza sulla protezione legale dei segreti commerciali sul Cloud, argomento ancora relativamente poco esplorato in letteratura. Il principale obiettivo è stato quello di fornire protezione legale per i segreti commerciali nell'architettura di brokeraggio Cloud. Tuttavia, a causa della considerevole evoluzione della blockchain sul Cloud, un obiettivo secondario è stato incluso nella ricerca. Questo consiste nell'offrire tutela giuridica per i segreti commerciali attraverso la blockchain. Il presente abstract riassume la ricerca nel contesto degli obiettivi sopra menzionati in rispettivi paragrafi.

La legislazione a livello mondiale sulla protezione dei dati si è evoluta verso la massimizzazione della protezione dei segreti commerciali. Ciononostante, sta diventando sempre più difficile provare le violazioni del segreto commerciale nel contesto del Cloud. Includere la tutela legale come misura preventiva potrebbe ridurre efficacemente l'onere della prova nei tribunali, se implementata da un broker online sul Cloud. Lo scopo primario di questa ricerca è quello di proporre un modello per un broker online che includa la protezione legale come misura preventiva per ridurre l'onere della prova durante il processo. Questa è una nuova area di ricerca interdisciplinare il cui insieme di conoscenze non è stato ancora ben definito. Il concetto sottostante al modello proposto è costruito sulla nozione di analisi fattoriale proveniente dall'area dell'apprendimento automatico non supervisionato. Per la valutazione tecnica, è stato applicato un metodo a due fasi che mostrava l'applicazione della protezione legale come misura preventiva e, conseguentemente, un ridotto onere della prova in un'aula di tribunale. Per la valutazione, è stata usata un insieme di dati sulla qualità del servizio dei fornitori di archiviazione Cloud (Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, SugarSync e Zip Cloud). La simulazione effettuata con il modello proposto ha mostrato risultati migliori rispetto ai suoi equivalenti nel campo, che in tribunale possono essere usate come prove per ridurre l'onere della prova. Per una convali-

da legale di tale conclusione, sono stati mandati dei questionari a degli esperti in diritto e informatica. Un totale di 6 persone hanno risposto al questionario (due provenienti da discipline informatiche, due da discipline giuridiche e due da informatica giuridica). Il campione (5 su 6 persone) si è dichiarato d'accordo sul fatto che, se i risultati del nostro modello possono essere verificati, possono essere usati in tribunale come parte delle prove per ridurre l'onere della prova. A livello teorico questa ricerca interdisciplinare è un tentativo pionieristico di fornire protezione legale per i segreti commerciali su Cloud. Allo stesso tempo, a livello pratico, darà beneficio alle imprese nel negoziare contratti con i provider dei servizi per ridurre l'appropriazione indebita sul Cloud.

Ciononostante, per un'impresa che usa l'architettura decentralizzata sul Cloud, come la blockchain, i contratti potrebbero svilupparsi in smart contract (un software autonomo che funziona sulla blockchain). In questo contesto, i contratti ben negoziati non forniranno una soluzione per minimizzare l'appropriazione indebita di segreti commerciali. Infatti, per questo caso è particolarmente importante rappresentare il ruolo della magistratura nella blockchain. Il secondo scopo della ricerca consiste nello sviluppare un modello che possa essere applicato sulla blockchain al fine di emettere un'ordinanza preliminare (o un ordine restrittivo preliminare di un tribunale) sulla violazione di un contratto che potrebbe portare all'appropriazione indebita di segreti commerciali. Questa parte della ricerca estende un modello proposto in precedenza usando la modellazione stocastica proveniente dalla disciplina della scienza dei dati (data science). Il cluster di calcolo ad alte prestazioni (High Performance Computing o HPC) dell'università di Lussemburgo (HPC @ Uni.lu) e il docker (una piattaforma contenitore software) sono stati usati per emulare un ambiente contrattuale di tre provider di servizi: i server di Redis, MongoDB e Memcached. I risultati dimostrano che le ordinanze del tribunale sono state emesse solo per i server di Redis e MongoDB. A livello tecnico, questa differenza può essere attribuita al fatto che Memcached è semplicemente usato per la memorizzazione temporanea (caching) e di conseguenza ha una tendenza minore alla violazione di un contratto. Invece, Redis e MongoDB, in quanto banche dati e message broker, compiono operazioni più complicate e hanno più possibilità di causare una violazione. Per una convalida legale di tale conclusione, sono stati mandati dei questionari a degli esperti in diritto e informatica. Un totale di 6 persone hanno risposto al questionario (due provenienti da discipline informatiche, due da discipline giuridiche e due da informatica giuridica). Il campione (4 su 6 persone) non è d'accordo con l'uso "ESCLUSIVO" ei risultati del nostro modello da parte dei tribunali per emettere un'ingiunzione preliminare (o un ordine restrittivo temporaneo) per la violazione di un contratto. A livello teorico, questa parte della ricerca è un tentativo pionieristico di fornire protezione legale sulla blockchain. D'altra parte, a livello pratico, aiuterà quelle imprese basate sulla blockchain a controllare e fermare la violazione di un contratto che potrebbe potenzialmente portare all'appropriazione indebita di segreti commerciali.

Oltre ai già citati benefici applicati, la seguente lista illustra brevemente i contributi per le discipline giuridiche di questa ricerca dottorale multidisciplinare:

- È la prima ricerca a concentrarsi sulla tutela giuridica per i segreti commerciali sul Cloud. Un simile concetto consolidato è quello di sicurezza dell'informazione, che fornisce protezione tecnica per segreti commerciali nel Cloud, come il criptaggio, l'hashing, eccetera.
- Presenta un approccio per costruire argomentazioni legali usando l'analisi della giurisprudenza e ridefinirla come concetto tecnico dal dominio delle tecnologie dell'informazione e della comunicazione (ICT).
- Nel campo della giurisprudenza, nonostante dei limiti giuridici, cioè i precedenti (o le decisioni del tribunale), siano vincolanti per tutti i tribunali sotto la stessa giurisdizione, questa è la prima ricerca a combinare la giurisprudenza con l'innovativo metodo Delphi Sampling per dare protezione legale ai segreti commerciali in un ambiente Cloud online senza frontier.
- È la prima ricerca ad applicare la nozione di confidentiality by design (confidenzialità fin dalla progettazione) che si concentra su una persona giuridica o un'impresa. Un simile concetto consolidato è quello di tutela della vita privata fin dalla progettazione (privacy by design), che

si concentra su una persona fisica o essere umano.

- Sfidando il mito che “gli smart contracts sono inviolabili” e nel contesto del diritto contrattuale, questa ricerca è la prima ad automatizzare il ruolo del tribunale (udienza probatoria).

Oltre ai contributi scientifici sopracitati nel campo del diritto, la seguente lista presenta i contributi nel dominio informatico:

- Nel contesto dell’analisi decisionale basata su criteri multipli, questa ricerca è la prima a identificare e analizzare il rumore (noise) nei dati e a risolvere i relativi problemi di incertezza strutturale (o l’errata specifica dei criteri)
- Nel contesto dell’apprendimento automatico, questa ricerca è la prima a proporre un’ “analisi decisionale basata su criteri multipli autoregolamentata” che opera senza l’intervento di un decisore e può quindi essere usata nei contesti dove è richiesta l’automazione del processo decisionale
- Nel contesto della scienza dei dati, questa ricerca è la prima a proporre un metodo per Delphi Sampling che ricorre alla validazione interdisciplinare dei risultati della ricerca.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction

This chapter mainly presents an overview of the PhD research. Sections 1.1 presents research focus and questions; section 1.2 presents research methodology and challenges; section 1.3 presents sources of law used during the research; section 1.4 presents research constraints both in terms of law and ICT, sections 1.5 and 1.6 present research contributions in the field of law and ICT respectively; and finally, sections 1.7 and 1.8 present thesis structure and list of published and under review research papers respectively.

1.1 Research Focus and Questions

Law differentiates between real human beings and enterprises by using the terms natural person and legal person respectively. in the context of data protection. This research focuses on data protection for a legal person with Research and Development (R&D) as one of the core activities of its business model. Such enterprise invests in R&D for acquiring, developing and applying know-how to defend its competitiveness in the market [1-3]. It has different means for commercial disclosure and exclusivity of applications developed from such know-how. Use of intellectual property rights (IPR) such as patents, copyrights, and trademarks are among them [4]. However, there is another type of know-how known as trade secrets [5, 6].

Fundamentally, a trade secret is information that provides an enterprise with a competitive advantage over other enterprises not having that information [7]. Unlike patent and copyrights, which provide enterprise with certain benefits after disclosure, for trade secrets, the enterprise must derive value from their secrecy. While the secret formula for Coca-Cola is the classic example of a trade secret, it is not the type of trade secret generally stored in the cloud. Instead, secret information in the form of customers list/profile, computer source code, and product designs and schematics are examples of trade secrets commonly stored in the cloud today [8]. One of the major risks in the cloud that can impair secrecy of these trade secrets is big data analytics.

Big data analytics is a data mining and analysis technique used in the cloud to explore data, usually large amount and business related - also known as "Big Data", to discover useful information. A growing use of Industrial Internet of Things (IIoT) by R&D based enterprises embrace the fact that corpus of Big Data can contain trade secret(s). Therefore performing big data analytics on such corpus may lead to trade secret misappropriation in the cloud [9, 10]. However, this particularly does not hold true when big data analytics is performed on public data [8]. One of the recent cases in a court of law that highlighted this aspect is *PeopleBrowsr, Inc. v. Twitter, Inc*[1]. During the case proceedings, the court noted that Twitter's big data analytics market consisted of companies that used analytics to derive insights from the flow of information generated on Twitter. PeopleBrowsr, one of such companies, receiving every tweet posted on Twitter through the Twitter "Firehose" and paid over $1 million per year for this access. It analyzed tweets and provided three major services: (a) Inference Measurement, which provides a unique visual stream that allow clients to identify others with like interest, as well as those who are influential in those communities; (b) Action Analytics for Government and Enterprises, which tracks all activities related to a brands or particular market in order to identify trends, competition, technology development etc.; and (c) Financial Data Service, which spot trends in Twitter data in order to more quickly detect when market changes are occurring.

On the contrary, in trademark litigation of *Tiffany (NJ), Inc. v. eBay, Inc.*[2] court observed that the results similar to PeopleBrowsr services together with advance data mining techniques can be used to generate *persona scores* and subsequently *customers list/profile* i.e. a trade secret. And in *Allied Portables LLC v. Youmans*[3], it was concluded that information illegally accessed i.e. customers list/profile, constituted a trade secret and is subjected to misappropriation claim. Thus, despite

[1] PeopleBrowsr, Inc. v. Twitter, Inc., U.S. Dist. LEXIS 31786; 2013 WL 843032 (2013)

[2] Tiffany (NJ), Inc. v. eBay, Inc., U.S. Dist. Ct. 576 F.Supp.2d 463 (2008). An expert for Tiffany testified that "using data mining techniques commonly used by corporations, eBay could have designed programs that identified listings of Tiffany items likely to be counterfeit, and that identified sellers thereof, using an algorithm to produce a *suspiciousness score*".

[3] Allied Portables LLC v. Youmans, No. 2:15-CV-294-FTM-38CM, (M.D. Fla. Nov. 6 (2015)

the fact that big data analytics is legitimate for open data as mentioned in *PeopleBrowsr, Inc. v. Twitter, Inc.,* the discussion on *Tiffany (NJ), Inc. v. eBay, Inc.* and *Allied Portables LLC v. Youmans* shows that it could be imputed for misappropriation when the data is not public. However, for such litigation claim to stand, the plaintiff must establish that the misappropriation has resulted in injury or damage [8]. In cloud context, however, proving such injury or damage could be complex phenomenon. One of the lawsuits that highlighted such aspect is *JetBlue Airways Corp. Privacy Litigation*[4]. In this case the court stated that *"it is apparent based on the briefing and oral argument held in this case that the sparseness of the damages allegations is a direct result of plaintiffs' inability to plead or prove any actual contract [or other] damages".*

On the contrary, rather than waiting for the litigation to unfold, embedding legal protection as a preemptive measure [11] could effectively reduce burden of proof in a court of law [8]. This was indicated in *EPIC v. the Department of Homeland Security (DHS)*[5]. In 2005, the Transportation Security Administration (TSA), a component of the DHS, began testing whole body imaging technology to screen air travelers. These scans produce detailed, three-dimensional images of individuals. In 2010, EPIC legally challenged the TSA's unilateral decision to make whole body imaging technology the primary screening technique in U.S. airports. EPIC argued that this technology violate the U.S. Video Voyeurism Prevention Act of 2004, which specifically prohibits the intentional capture of an image of a private area of an individual without their consent under circumstances in which the individual has a reasonable expectation of privacy. Whereas in defense, TSA proclaimed that its whole body imaging technology incorporates a *privacy algorithm* that eliminates much of the detail shown in the images of the individual while still being effective from a security standpoint. Such implementation of an algorithm by TSA to preserve privacy of a natural person is an excellent example of legal

[4] In re JetBlue Airways Corp. Privacy Litig., 379 F.Supp.2d 299 (U.S. Dist. Ct., Eastern Dist. of NY, August 1, 2005).

[5] EPIC v. the Department of Homeland Security, Case No. 09-02084(RMU) (D.D.C.filed Nov. 9, 2009)

protection (i.e. privacy) embedded as a preemptive measure. Furthermore, during the litigation it reduced burden of proof for DHS based upon the evidence that shows accuracy of an algorithm for preserving privacy.

Respectively in the cloud, participating in the same degree is an online broker. It is a software agent used to embed preemptive measure in the cloud [11]. However, the discussion in section 3.2 shows that online broker is still at initial level when it comes to provisioning legal protection. The primary aim of this research is to propose a model for an online broker that embeds legal protection as preemptive measure to reduce burden of proof during litigation. More specifically, the primary research question addressed in this research is: *how an online broker can embed legal protection as preemptive measure to reduce burden of proof in a court of law?*

For R&D based enterprise that employee online broker, the answer of above research question will benefit in negotiating a contract with service providers to minimize trade secret misappropriation in the cloud. However, if the enterprise starts using decentralized architecture in the cloud e.g. blockchains, the contract could emerge towards a smart contract [12], an autonomous software program running over blockchains [13]. In this context, well negotiated contract is not the solution to minimize trade secret misappropriation. In fact, in such case it is particularly relevant to instantiate role of judiciary over a blockchain [12].

Blockchain is an emerging technology for decentralized and transactional data sharing across a large network of untrusted participants [14]. The first generation of the blockchain was a public ledger for monetary transactions with very limited capability to support programmable transactions. The typical example is cryptocurrency or Bitcoin [15]. The second generation of the blockchain became a generally programmable infrastructure with a public ledger that records computational results. In this generation, smart contracts were introduced as autonomous programs that are deployed by the components connected to the blockchain to reach agreements and solve problems with minimal trust [13]. Autonomous Decentralized Peer-To-Peer Telemetry (ADEPT), a project of IBM is an excellent implementation of smart contracts to enable programmable transaction in cyber-physical system or internet of things [16].

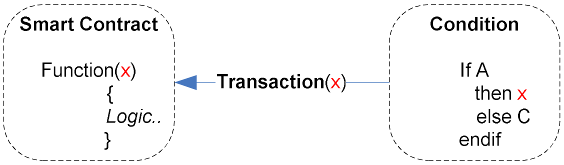


Fig. 1.1 Smart Contract

A smart contract is a piece of code that resides on a blockchain and is identified by a unique address. It includes a set of executable functions and state variables. The function is executed when a transaction is invoked by a certain condition (or by an electronic event or data). These transactions include input parameters that are required by the functions in the contract, see Figure 1.1. Upon the execution of a function, the state variables in the contract change depending on the logic implemented in the function. This execution is self-enforceable i.e. once a smart contract is concluded, its further execution is neither dependent on intend of contractual parties or third party, nor does it require any additional approvals or actions from their side [17]. Thus, any malicious intent of the party i.e. breach of contract, and role of third party addressing the malicious intent i.e. judiciary, becomes irrelevant during the execution of a smart contract [18].

However, in addition to dealing with breaches, contract law also encompasses deviations in pre-defined outcomes [19]. Even though breach of contract and role of judiciary become irrelevant during the execution of a smart contract, what if an output of a smart contract is considered as a breach by court of law? For example, a court may acknowledge deviation in output of a contract as a breach, if average uptime of a web service is 90% instead of agreed 95%. The secondary research question addressed in this research is: *what happens when the outcome of a smart contract deviates from the outcome that the law demands?* The answer to this research question will eventually benefit blockchain driven R&D based enterprises to control and stop breach of contract that could potentially lead to trade secret misappropriation.

1.2 Research Methodology and Challenges

Figure 1.2 presents flow chart of this PhD research that shows how the primary and secondary research questions identified in previous section are addressed.

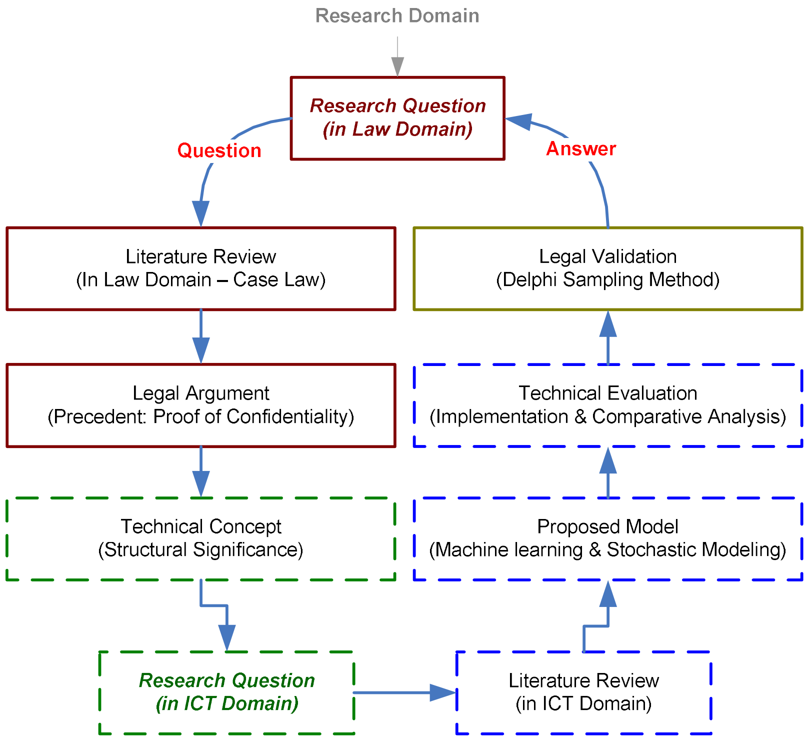


Fig. 1.2 Research Methodology

In the figure, the dotted rectangles show the research activities related to the field of ICT whereas the rest are related to the field of Law. As there is no law that specifically talks about protection of trade secrets in the cloud, see section 2.3, therefore the ***first challenge*** in this research was to build legal argument for protection of trade secrets in the cloud. The legal argument (precedent: proof of confidentiality) was identified during literature review of legal text (case law analysis) that addressed the related ***research question (in law domain)*** as shown in the flow chat. This challenge is addressed in section 3.1.

The ***second challenge*** was a “twofold transformation” i.e. to find the technical concept that correspond to the legal argument and then build a related research question in ICT domain. The prior i.e. transformation into technical concept, was a time consuming task because there are numerous sub-domains in the field of ICT. For example, in section 3.1 and table 3.1, a part of legal argument “…*proof of confidentiality: a proof for reasonable efforts made by the owner to protect trade secret in the cloud”* was transformed into technical concept of “*structural significance*” that belongs to the domain of multi-criteria decision analysis (MCDA), which belongs to the domain of operation research, which further belongs to the domain of decision science in the field of ICT. This challenge is addressed in section 3.1.

The ***third challenge*** was a review of ICT literature to check if answer to the ***research question (in ICT domain)*** already exists or not. As it did not exist, this PhD research proposed a solution and performed its technical evaluation in a cloud environment. The two datasets used during the evaluation were “*feedback from customers*” and “*feedback from servers*” on Quality of Service (QoS) of cloud storage providers. The first dataset i.e., feedback from customers, was compiled using leading review websites such as Cloud Hosting Reviews, Best Cloud Computing Providers, and Cloud Storage Reviews and Ratings. The second dataset i.e., feedback from servers, was generated from cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). This challenge is addressed in sections 3.2, 3.3, 3.4, and 4.2.

The ***fourth challenge*** was to propose a method that can be used to legally validate results of the PhD research (including results of activities in the field of ICT). In this regards, the research proposed method of “Delphi Sampling”, which seeks inter-disciplinary (ICT and law) validation for the results. This proposed method is based on “Delphi forecasting technique [20]” from the field of policy analysis. In this method, several rounds of questionnaires are sent out to inter-disciplinary experts (or sample), and the anonymous responses on the results are accumulated and shared with the group after every round. The experts are allowed to modify their response in succeeding rounds. Since multiple rounds of questions are asked and the panel is told what the group thinks as a whole, the Delphi Sampling seeks to reach the inter-disciplinary validation for the results through consensus. Based on the universal fact of Dominant Minority i.e. opinion of all (experts in the world) is dominated by the opinion of few (most experienced and well reputed experts) [21], the results of Delphi Sampling is an approximation technique for universal validation of multi-disciplinary research results. This challenge is addressed in sections 3.5, 4.3, and 6.4.

1.3 Sources of Law for the Research

Several regulations are potentially related to cloud computing including sector specific regulations e.g. health sector and financial sector regulations [22]. In addition, the emerging trends are: a) use of case law for cloud computing [23]; b) use of opinions e.g. at EU level, opinion of Article 29 Working Party[6]; and c) regulations in the form of contracts and standardization documents created by the private sector [24]. This research uses case law as a source to build a legal argument for protection of trade secrets in the cloud, see sections 1.5.3, and 2.3 for more details.

1.4 Law and ICT based Research Constraints

Following list presents law and ICT related research constraints that were encountered during the execution of research methodology presented in figure 1.2.

1. Many regulations are potentially applicable to cloud computing [22-24]. Given the extensiveness and density of these laws, complete analysis was not possible in this research.
2. The scope of literature review in this research can be enhanced by including publications presented in languages other than English. For

[6] The "Article 29 Working Party" is the short name of the Data Protection Working Party established by Article 29 of Directive 95/46/EC. It provides the European Commission with independent advice on data protection matters and helps in the development of harmonised policies for data protection in the EU Member States.

example, for systematic review in section 3.2, the research published in English language between January 2010 and March 2017 was explored by using the following databases: ACM Digital Library, Google Scholar, IEEE Xplore, ScienceDirect, and SpringerLink.

3. Datasets (and sources) used in this research if integrated with additional methods e.g. implementing proposed model in Amazon cloud and monitoring data streams for information security, could have increased the scope and depth of analyses and results.
4. Communication of normative and empirical research results between the disciplines [25] of Law and ICT is one of the barriers in achieving genuine interdisciplinary validation [26]. For example, there is 100% chance that the empirical results that are valid in ICT domain receives rejection based on the normative claim made by a lawyer.

1.5 Research Contributions in the Field of Law

In addition to following applied benefits of this PhD research, following subsections briefly presents novel research contributions in the field of Law.

- This research will benefit R&D based enterprises in negotiating a contract with service providers to minimize trade secret misappropriation in the cloud.
- This research will benefit blockchain driven R&D based enterprises to control and stop breach of contract that could potentially lead to trade secret misappropriation.

1.5.1 Legal Protection of Trade Secrets in the Cloud

Contrary to the belief that the cloud is a virtual environment, basically it is number of computer installed geographically at many locations (e.g. countries) [27]. Since, the enterprise using the cloud is not aware of these geographical locations, the whereabouts of the uploaded data (or trade secrets) and its management is a matter of great worry [28]. In the domain of ICT, such concern is (or can be) minimized by implementing number of information security measures e.g. cryptography (using encryption and hashing) and access management (using access keys and firewalls) [29]. However, even after adopting these measures one thing is for sure i.e. once the trade secret is uploaded in the cloud, owner loses its control. In fact, given the unknown geographical locations of the computers, the responsibility of the owner extends to the level where he must ensure that the service provider has necessary information security measures in place to protect trade secrets in the cloud [21]. If the provider does not guarantee such measures, the risk e.g. big data analytics (see section 1.1), could lead to misappropriation of a trade secret. In law, the duty of an owner to produce the evidence for misappropriation is known as "*burden of proof*" [30]. In cloud context, such burden could be extremely complex, see discussion on *JetBlue Airways Corp. Privacy Litigation* in section 1.1.

This research uses ICT (unsupervised machine learning) to the help owner of a trade secret to reduce burden of proof in the court. In doing so, it is first in-line to focus on "legal protection" for trade secrets in the cloud as compared to the well-established similar concept of "information security", which provides technical protection for trade secrets in the cloud e.g. encryption, hashing etc.

1.5.2 Implementing Notion of Confidentiality by Design

The idea of incorporating law into ICT design is not completely new. Privacy by Design (PbD) is one of such established concepts [31]. Privacy is a legal concept that is related to a physical person (human being). PbD includes the idea that ICT design should minimize the amount of personal data processing that could lead to identification of a physical person [31].

The underlying notion in this PhD research is also about incorporating law into ICT architecture. However, unlike PbD that focuses on privacy of a physical person, this research focuses on confidentiality of a legal person (an enterprise) and proposes a new concept of Confidentiality by Design (CbD). CbD includes the idea that ICT architecture should scale down burden of proof in the court of law, which could help in proving trade secret misappropriation, see chapter 3. Unlike PbD, CbD is a novel area of inter-disciplinary research whose body of knowledge is not yet well established. This PhD research is first in-line to implement notion of CbD in an online cloud environment.

1.5.3 Case Law Analysis for Trade Secrets in the Cloud

Common law is one of the two main legal systems in the present world, the other one is civil law [32]. Case law is the part of common law that consists of judgments given by courts for cases brought before them. These judgments are called precedents and they are binding on the courts within the same jurisdiction for similar cases [32]. Whereas, Civil law is a predefined and highly structured code of rules in which a judge decides cases without any reference to precedent(s) [33].

Legal systems (common or civil law) are only applicable to a particular geographic region (e.g. country) [33]. Whereas, because of universal footprint of the cloud i.e. computer installed geographically at many locations (e.g. countries), implementing legal protection in the cloud could be a challenge [34]. This research is first in-line to use case law together with newly proposed method of Delphi Sampling (see section 1.2) to provide legal protection for trade secrets in the cloud. In this regards, in the domain of case law, precedents set by previous court rulings on trade secret misappropriation (in United States of America - USA) were identified, see table 3.1. Afterwards, using Delphi Sampling, it was established that identified precedents are applicable in any jurisdiction (or most of them) around the globe and hence, they are also applicable to the cloud, see section 3.5.

1.5.4 Automating Role of Judiciary over Blockchains

Before trade secret misappropriation trial starts, enterprises (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing [28, 35]. In such hearing, court determines whether there is enough evidence to start a trial. Initially, it assesses significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it examines if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial [28, 35, 36].

By defying the myth that "smart contracts are unbreachable [18]" and in the context of contract law [19], chapter 6 presents automation of above mentioned role of the court (evidential hearing). In this regards, it uses unsupervised machine learning and stochastic modeling together with blockchain (smart contract). This PhD research is first in-line to automate role of the judiciary over blockchains.

1.6 Research Contributions in the Field of ICT

This section briefly presents novel contributions of the PhD research in the field of ICT.

1.6.1 Self-Regulated Multi-criteria Decision Analysis - MCDA

Multi-criteria Decision Analysis (MCDA), one of the prevalent branches of operations research, aims to design mathematical and computational tools for selecting the best alternative among several choices [37]. It prescribes a methodology that deals with the most important components in the process of decision making and aims at supplying reliable information to take an unbiased decision. These components include an objective that is a pre-established goal achievable under given constraints. These constraints are criteria that are used to rank potential alternatives. Such ranking is generated with respect to criteria and their significance provided by a decision maker (DM) [37]. An unbiased selection and valuation of criteria by DMs strongly relates to their profound knowledge of the subject matter. Hence, the approach is termed ineffective when the DM has insufficient subject knowledge. In the context of online cloud environment, this PhD research is first in-line to propose self-regulated MCDA that operates without DM interference and well suited for the context where automation of decision making is required, see chapter 4.

1.6.2 Identifying & Analyzing Noisy Data in MCDA (Machine Learning)

Real-world data, which is the input for data processing and analytics, are affected by many factors; among them, the presence of noise is a main factor. It is an unavoidable problem, which influence data processing and analytics. Noisy data in MCDA generally means that the decision making take account of insignificant correlations (or criteria), which could result in selection of sub-optimal or least optimal alternative [38]. Using unsupervised machine learning (or factor analysis); this PhD research is first in-line to identify and analyze noisy data in MCDA, see sections 3.4.1 and 4.2.1.

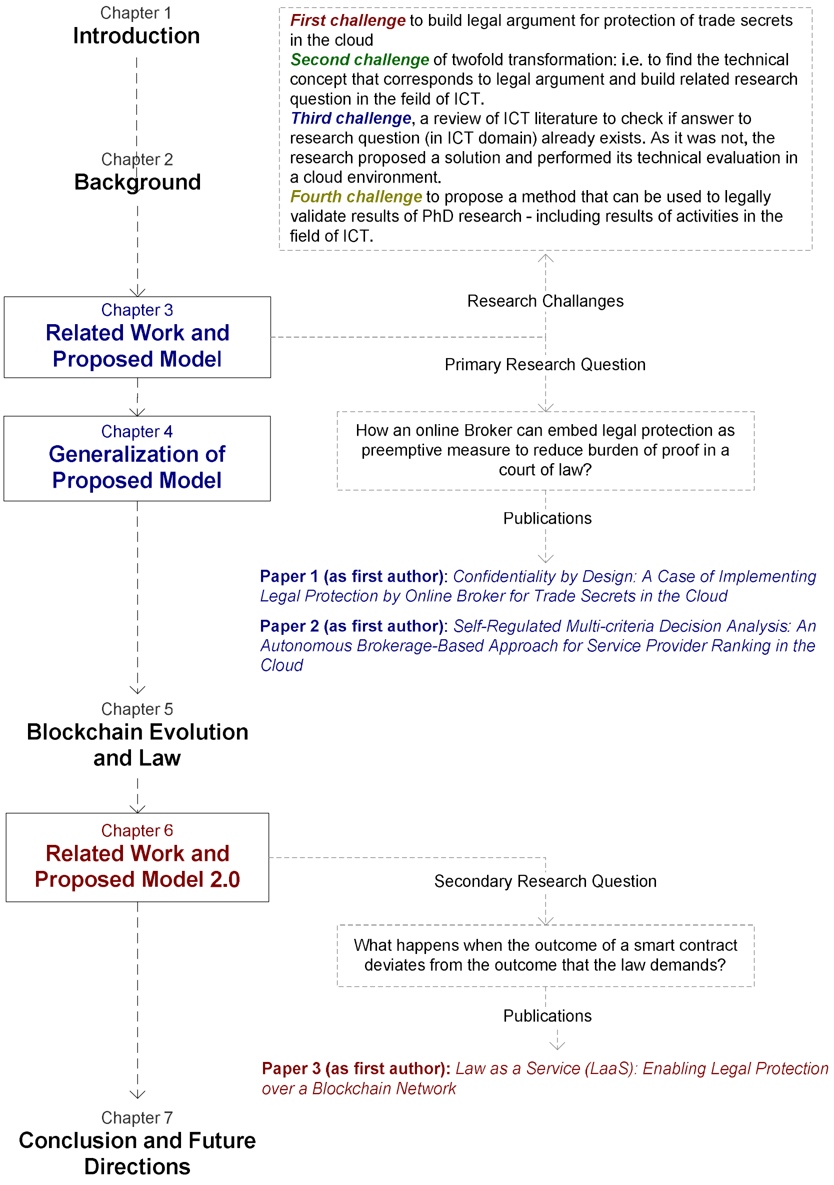
1.6.3 Delphi Sampling Method (Multidisciplinary Research)

Communication of normative and empirical research results between the disciplines of law and ICT is one of the barriers in achieving genuine interdisciplinary validation. The proposed method of Delphi Sampling is an approximation technique for universal validation of multidisciplinary research results. Sections 3.5, 4.3, and 6.4 present use of Delphi Sampling to seek inter-disciplinary (ICT and law) validation of the results in this PhD research.

1.7 Thesis Structure

Figure 1.3 presents pictorial presentation of the thesis structure and related research publications. Chapter 2 (**Background**) and chapter 5 (**Blockchain Evolution and Law**) presents information on essential concepts necessary for the understanding of the PhD research. Chapter 3 (**Related Work and Proposed Model**) successfully addresses:

- The primary research questions identified in section 1.1 i.e. *how an online broker can embed legal protection as preemptive measure to reduce burden of proof in a court of law?*
- The following four challenges of the PhD research presented in section 1.2:
    1. The ***first challenge*** to build legal argument for protection of trade secrets in the cloud


Chapter 1
Introduction
Chapter 2
Background
Chapter 3
Related Work and Proposed Model
Chapter 4
Generalization of Proposed Model
Chapter 5
Blockchain Evolution and Law
Chapter 6
Related Work and Proposed Model 2.0
Chapter 7
Conclusion and Future Directions
Research Challanges
Primary Research Question
Publications
Secondary Research Question
Publications

First challenge to build legal argument for protection of trade secrets in the cloud
Second challenge of twofold transformation: i.e. to find the technical concept that corresponds to legal argument and build related research question in the feild of ICT.
Third challenge, a review of ICT literature to check if answer to research question (in ICT domain) already exists. As it was not, the research proposed a solution and performed its technical evaluation in a cloud environment.
Fourth challenge to propose a method that can be used to legally validate results of PhD research - including results of activities in the field of ICT.

How an online Broker can embed legal protection as preemptive measure to reduce burden of proof in a court of law?

Paper 1 (as first author): *Confidentiality by Design: A Case of Implementing Legal Protection by Online Broker for Trade Secrets in the Cloud*

Paper 2 (as first author): *Self-Regulated Multi-criteria Decision Analysis: An Autonomous Brokerage-Based Approach for Service Provider Ranking in the Cloud*

What happens when the outcome of a smart contract deviates from the outcome that the law demands?

Paper 3 (as first author): *Law as a Service (LaaS): Enabling Legal Protection over a Blockchain Network*

Fig. 1.3 Thesis Structure and Publications

2. The ***second challenge*** of twofold transformation: i.e. to find the technical concept that corresponds to legal argument and build related research question in ICT domain.
3. The ***third challenge***, a review of ICT literature to check if answer to ***research question (in ICT domain)*** already exists. As it was not, the research proposed a solution and performed its technical evaluation in a cloud environment.
4. The ***fourth challenge*** to propose a method that can be used to legally validate results of PhD research - including results of activities in the field of ICT.

Chapter 4 (**Generalization of Proposed Model**) presents generalization of model proposed in chapter 3. This is one of the major requirements of the second PhD degree "PhD in Informatics (Informatique)" at University of Luxembourg, Luxembourg. Furthermore, the dataset used in chapter 3 for evaluation of the proposed model is secondary data (data that was collected by someone other than the user). This chapter takes the evaluation one step further and test the proposed model in cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). Chapter 6 (**Related Work and Proposed Model 2.0**) successfully addresses the secondary research questions identified in section 1.1 i.e. *what happens when the outcome of a smart contract deviates from the outcome that the law demands*? Finally, chapter 7 (**Conclusion and Future Research**) concludes the research by presenting main research findings, limitations of models in chapter 3, 4, and 5, and suggests related research directions.

1.8 Research Publications

Research publication of chapter 1, 2, and 3 is paper 1. Paper 1, *Confidentiality by Design: A Case of Implementing Legal Protection by Online Broker for Trade Secrets in the Cloud*, is submitted to the IEEE Journal - IEEE Transactions on Services Computing, and it is currently under review.

Research publication of chapter 4 is paper 2. Paper 2, *Self-Regulated Multi-*

criteria Decision Analysis: An Autonomous Brokerage-Based Approach for Service Provider Ranking in the Cloud, is a generalization of proposed model in paper 1. It also tests the model of paper 1 in cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). The paper is accepted in 9th IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2017), December 11-14, Hong Kong China.

Research publication of chapter 5 and 6 is paper 3*, Law as a Service (LaaS): Enabling Legal Protection over a Blockchain Network*. The paper is accepted in 14th International Conference on Smart Cities: Improving Quality of Life using ICT & IoT (HONET-ICT 17), October 09-11, Irbid Jordan. (2017). Abstract of the paper 1, 2, and 3 are presented below.

1. *Confidentiality by Design: A Case of Implementing Legal Protection by Online Broker for Trade Secrets in the Cloud*

   Authors: Muhammad Umer Wasim, Pascal Bouvry, Tadas Limba

   Submitted to: IEEE Transactions on Services Computing (Journal)

   Status: Under Review

   Abstract— Data Protection legislation has evolved around the globe to maximize legal protection of trade secrets. However, it is becoming increasingly difficult to prove trade secret violations in cloud context. Embedding legal protection as a preemptive measure could effectively reduce such burden of proof in a court of law, which can be implemented by an online broker in the cloud. This research proposes a model for an online broker that embeds legal protection as preemptive measure to reduce burden of proof during litigation. This is a novel area of inter-disciplinary research whose body of knowledge is not yet well established. For evaluation of proposed model, two-stage procedure was implemented that shows implementation of legal protection as preemptive measure and subsequently, reduced burden of proof in a court of law. A real time Quality of Service based dataset for cloud storage providers (Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, Sug-

arSync, and Zip Cloud) was used for the evaluation. Theoretically this multi-disciplinary research is a pioneer discussion on providing legal protection to trade secrets in the cloud. Whereas, the beneficiary of the research would be R&D based enterprises that see trade secret misappropriation as limiting factor for acquisition of cloud services.

Index Terms—legal protection, trade secret, cloud computing, big data analytics, burden of proof, online broker, multi-criteria decision analysis (MCDA), analytic hierarchy process (AHP), technique for order of preference by similarity to ideal solution (TOPSIS), unsupervised machine learning, factor analysis, principal factor analysis, quality of service (QoS).

2. *Self-Regulated Multi-criteria Decision Analysis: An Autonomous Brokerage-Based Approach for Service Provider Ranking in the Cloud*

   Authors: Muhammad Umer Wasim, Abdallah A. Z. A. Ibrahim, Pascal Bouvry, Tadas Limba

   Submitted to: 9th IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2017), December 11-14, Hong Kong China. (2017)

   Weblink: http://2017.cloudcom.org/

   Status: Accepted

   Acceptance Ratio: 29.4%

   Abstract—The use of multi-criteria decision analysis (MCDA) by online broker to rank different service providers in the Cloud is based upon criteria provided by a customer. However, such ranking is prone to bias if the customer has insufficient domain knowledge. He/she may exclude relevant or include irrelevant criterion termed as 'misspecification of criterion'. This causes structural uncertainty within the MCDA leading to selection of suboptimal service provider by online broker. To cater such issue, we propose a self-regulated MCDA, which uses notion of factor analysis from the field of unsupervised machine learning. Two QoS based datasets were used for evaluation of proposed model. The prior dataset *i.e., feedback from customers, was compiled using leading review websites such as Cloud Hosting Reviews, Best Cloud Computing Providers, and Cloud Storage Reviews and Ratings. The later dataset i.e., feedback from servers, was generated from cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). The results show better performance of proposed model as compared to its counterparts in the field. The beneficiary of the research would be enterprises that view insufficient domain knowledge as a limiting factor for acquisition of cloud services.*

Keywords—multi-criteria decision analysis (MCDA), online broker, misspecification of criteria, structural uncertainty, unsupervised machine learning, factor analysis, quality of service (QoS).

3. Law as a Service (LaaS): Enabling Legal Protection over a Blockchain Network

   Authors: Muhammad Umer Wasim, Abdallah A. Z. A. Ibrahim, Pascal Bouvry, Tadas Limba

   Submitted to: 14th International Conference on Smart Cities: Improving Quality of Life using ICT & IoT (HONET-ICT 17), October 09-11, Irbid Jordan. (2017)

   Weblink: http://honet-ict.org/

   Status: Accepted

   Abstract— In the current world of online contracts i.e. service level agreements (SLAs), contract breaches are usually compensated by gift vouchers, however in an emerging world of online contracts i.e. smart contracts, the breaches could potentially lead to court injunctions over blockchains. This research proposes Probability based Factor Model (PFM) that can be implemented over the blockchain to automatically issue court injunction for the breach, which has a potential to create substantial damage and has high probability to occur in the future. The underlying concept in PFM is built upon the notion of factor analysis and stochastic modeling from the discipline of Data Science. High perfor-

mance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu) and docker (a software container platform) were used to emulate contractual environment of three service providers: Redis, MongoDB, and Memcached Servers. The results showed that court injunction(s) was issued only for Redis and MongoDB Servers. Technically, this difference could be attributed to the fact that Memcached is simply used for caching and therefore, it is less prone to breach of contract. Whereas, Redis and MongoDB as databases and message brokers are performing more complex operations and are more likely to cause a breach. The beneficiary of the research would be an enterprise that views breach of contract as a limiting factor for implementation of smart contract in cyber-physical system or internet of things.

Keywords— blockchain, smart contract, contract law, breach of contract, court injunction, unsupervised machine learning, factor analysis, stochastic modeling, structural equation modeling.

Chapter 2 : Background

This chapter describes the background information on essential concepts necessary to understand PhD research addressing the primary research question identified in section 1.1. Section 2.1 provides a brief overview of cloud computing, its service models, deployment models, and stakeholders. Afterwards, section 2.2 discusses data protection in the cloud in terms of personal data (privacy) and business data (trade secrets). Section 2.3 presents current efforts and related issues for legal protection of trade secrets in the cloud. Finally, section 2.4 summaries the discussion and findings of the chapter.

2.1 Cloud Computing

Cloud is a shared infrastructure allowing customers to access computing resources remotely [27]. Consumers of cloud services connect to these resources over the internet for their computing requirements. In addition to the basic requirements like sending and receiving emails, consumers store everything from valuable commercial data to photographs/videos on the cloud [21]. This data is stored on the computers located at different geographic locations (e.g. countries). From a technical viewpoint, the location of the data is often considered irrelevant, however, it has legal implications [21].

Cloud computing allows the consumers to outsource their computing requirements in a proficient and cost effective manner [27]. Popular cloud service like Dropbox is common examples of the cloud based storage service. It has many advantages like: global access to documents, inexpensive data backup, and access to new and innovative business solutions (e.g. Dropbox for Businesses) [27]. However, in addition to these advantages, new challenges have also evolved. For example, data storage at different geographic locations has created challenges for regulators, particularly in the areas of intellectual property, data protection, and compliance in many industry sectors such as finance or healthcare [21]. Some of the challenges are specific to type of service and deployment models used in the cloud. The following subsection presents service and deployment models of cloud computing.

2.1.1 Cloud Computing Service and Deployment Models

The extent to which a consumer can have control over their data depends on the cloud model under use. In general, following are the three cloud computing service models available [27]:

- *Infrastructure as a Service (IaaS)* provides the consumer with computing resources such as processing power (and/or storage) e.g. Google Compute Engine. Under this model, the consumer has most of the control over the data in the cloud.
- *Platform as a Service (PaaS)* provides the consumer with the platform (software environment) for developing (and commonly deploying) custom applications e.g. Google App Script/Engine. Under this model, the consumer has less control over the data as compared to IaaS.
- *Software as a Service (SaaS)* provides the consumer with access to the software e.g. Gmail. Under this model, the consumer has the least amount of control over the data.

In addition to different service models discussed above, not all clouds are created with equal accessibility. In general, following are the three cloud computing deployment models available [27]:

- On the most secure model in terms of accessibility is a *private cloud*. This model is often dedicated to a single enterprise, or shared by members of the same corporate group. The owner of private cloud owns the data center(s) and other physical facilities. The outsourcing in this model does not generally take place, providing for a greater level of data security. It has some of the advantages of cloud computing, like global access, but do not capitalize the cost saving obtained through shared networks. This model is appropriate for enterprises or corporations with sensitive computing needs (or sensitive data processing needs) including those in the financial and health sectors.
- *Community cloud* is similar to a private cloud in a way that it has con-

trolled access to the computing resources. Instead of only available for an enterprise (like in private cloud), this model serves many enterprises with similar security requirements e.g. banking cloud or healthcare cloud. The benefit of this model is a sharing of ICT resources allowing for lower cost, whereas, the down side is reduced security (as per increase in number of enterprises).

- On the less secure side is a *public cloud* e.g. Amazon and Google cloud. It provides access to many consumers. It has lowest cost and most commonly used model. However, low cost, flexibility, and accessibility come at the cost of security, as it may expose the data of their consumers to the greatest risks of misappropriation. Moreover, the data may be monitored for secondary clients or reused by third party applications. As a result of their large size and implementation, it is difficult to determine location of the data at any given time.
- *Hybrid cloud* combines public and private cloud models to provide a higher level of security e.g. sensitive data is kept or transferred over private cloud while less sensitive data is kept or transferred over public cloud. By using this model, the cloud consumer takes advantage of economy of scale and advanced security.

The major stakeholders that plan, deliver, and consume above mentioned cloud computing service and deployment models are discussed in following section.

2.1.2 Major Stakeholders in Cloud Computing

In general, following are the four major stakeholders involved in planning, delivering, and consumption of cloud computing service and deployment models discussed in preceding section [27]:

- *Consumer*: The final end user of a cloud computing service. The other terms used for an end-user of a cloud service are: cloud client and cloud subscriber.
- *Service Provider*: Cloud service provider is the enterprise making the cloud service available to the consumer. Depending on the services

models (SaaS, PaaS, and IaaS), the role of the service provider varies. For example in the SaaS, service provider will provide all features of the cloud service (e.g. Gmail) to the consumer. In the PaaS, service provider is in control of the underlying platform and put the consumer in control of applications running on the platform. In the IaaS, service provider even shares access to platform with consumer.

- *Auditor*: The auditor is an external agent that evaluates the cloud service. The typical function of an auditor is to verify compliance in reference to regulation, standard, or contract. An auditor is seen as playing an increasingly important role in cloud security, privacy protection, and overall trust in the cloud. Many public bodies require third party audits for evaluation of cloud services (or service providers).
- *Broker*: Cloud broker is an intermediary agent between consumers and service providers. It play critical role in finding a desired cloud service(s) for consumers and helps in establishing a contractual relationship between consumers and service providers.

Trust between above mentioned stakeholders is critical for planning, delivering, and consumption of cloud computing service and deployment models. The following section discusses notion of data protection in the cloud that could aid or impair such trust.

2.2 Data Protection and Cloud Computing

Data security is the leading concern that could aid or impair the trust between stakeholders in the cloud [28]. Threats to data security can emanate from the *consumers* themselves as shared infrastructure of cloud computing opens the possibility for interference or espionage [39]; from the *insider* (service provider); from *third party insiders* (sub-contractor) [40]; or from the *outsiders* e.g. spammers are using phishing campaigns and hackers are using cryptographic key cracking [39]. These threats mainly emerge from lack of control on the resources, increased exposure of internal infrastructure, and insufficient adaptation of security measures. This implies that both service providers as well as consumers have to be aware of the existence of such threats and take appropriate measures to address them. Taking such measures is not just based on business reasons but is also due to mandatory legal requirements [41], which are different as per type of the data in the cloud. The following subsections present the two most common types of data in the cloud and the related legal requirements for their protection.

2.2.1 Protecting Personal Data (Privacy) in the Cloud

As discussed in section 1.5.2, privacy in the cloud begins with understanding the concept of 'personal data' and it's 'processing'. In the EU context i.e. by using European Data Protection Directive (or Regulation), the two concepts are explained as follows [41].

1. The term *processing* includes a range of actions related to data including the collection, recording, organization, storage, alteration, retrieval, consultation, use, transmission, dissemination, combination, blocking, and destruction. The directive is mainly focused on the *processing of personal data wholly or partly by automatic means*. The use of the term wholly or partly suggests that an automated operation that contains some manual use of personal data falls within the jurisdiction of the directive. Moreover, the directive is also valid to non-automated processing which forms part of a filing system or are intended to form part of a filing system (structured data). Fundamentally, the directive applies whenever personal data is processed using automated or non-automated means (except some exceptions). Given many operations included within the concept of data processing e.g. collection, recording, organization; processing of the data in the cloud may also involve one or more of these operations and hence, it is subjected to personal data protection regulation. For example, if IaaS provided storage is used for personal data, then it will be subjected to personal data protection regulation e.g. European Data Protection Directive (or Regulation).
2. The *personal data* is any information relating to an identified or identifiable natural person. Identification requires features that describe a person

in such a way that he/she can be distinguished from others. Such identification of the individual could happen directly from the information being processed or could be by combining the information being processed with other information. To conceal identify of a person during data processing, following are the most common techniques in use [31]:

- ***Anonymisation*** is a process by which data is concealed to make it difficult to identify data subjects. This can be done by deleting identifying details.
- ***Pseudonymisation*** involves replacing names or other direct identifiers with codes or numbers.
- ***Encryption*** is the process of changing a plain text in to cipher-text. A ciphertext is unreadable by a human or computer without the cipher (or decryption key).

A combination of these techniques, for example anonymisation, pseudonymisation, and encryption can enhance the protection of the personal data in the cloud.

2.2.2 Protecting Business Data (Trade Secrets) in the Cloud

Discussions regarding trade secrets protection in the cloud begin with understanding the concept of *contracts*. In the EU, there is no single definition of a contract. Existing definitions are found in various regulations related to commerce (or electronic commerce) [42]. In the cloud, service providers enter into contracts with consumers in a number of ways. For some consumers, the contract follows the old contracting scheme (paper and pen), while others agree to terms electronically (electronic contract). Also, the term electronic contract does not have a standard definition [43]. In general, electronic contract is an agreement where a service is formally defined and relevant factors for data protection, among others, are decided between service providers and consumers in an online environment. Most common of these factors for data protection include followings [43-45]:

- ***Availability***: Availability enables authorized consumers to access data and to receive it in the desired time.

- ***Accuracy***: Data is accurate if it is free from errors and it has the format that the consumers want. If data has been altered intentionally or unintentionally, it is no longer accurate.
- ***Authenticity***: Authenticity of data is the state of being original. Data is unauthentic if it is not in the state in which it was created, placed, stored, or transferred.
- ***Confidentiality***: Data is confidential if it is protected from unauthorized access and if unauthorized access is made to the data, confidentiality is breached.
- ***Integrity***: Data has integrity when it is complete and remains uncorrupted. Many malwares are designed with the aim to corrupt the data.
- ***Utility***: The utility of data is its format. If data is accessible, but is not in a format that is meaningful to the consumer, it is not useful or has no utility.
- ***Possession***: The possession of data is its control. Data is said to be in the possession, if one has obtained it (regardless of its format). While a breach of confidentiality always results in a breach of possession, a breach of possession does not always result in a breach of confidentiality. For example, a company has secured its data using encryption. An x-employee decides to take a copy of the data and sell it to the competitor. The stealing of the data from protected environment is a breach of possession. But, because the data is encrypted and cannot be used without decryption; therefore, there is no breach of confidentiality.
- ***Security Measures and Standards***: Given the fact that cloud is a shared infrastructure, security measures and industry standards (e.g. ISO 2700 standards) play a central role in protecting data in the cloud.
- ***Acceptable Use Policies***: Acceptable use polices are applied on consumers to refrain them from unauthorized use of the service.
- ***Intellectual Property Rights (IPR)***: In general, service providers do not claim any ownership rights on the data stored by the consumer in

the cloud. However, data that is created during the life of the service may be claimed as the exclusive property of the service provider e.g. algorithms developed while optimizing the consumer data in the cloud.

- ***Data Breach Notification and Liability***: The requirement to notify data breach comes from the terms of the contract. Most standard electronic contracts offer little (or nothing) in the way of liabilities for data misappropriation.
- ***Unilateral Amendment of Contract***: The contract must allow consumers with the ability to object unilateral changes in the contract that relates to the data protection in the cloud.
- ***Subcontracting***: If multiple providers e.g. service provider, infrastructure provider, software provider etc., are involved in handling data in the cloud, there must be a liability clause for each provider in the contract.
- ***Location of Data***: Consumers can use the contract to define the location of data in motion, at rest, and geographic locations for backup.
- ***Portability*****:** Consumers can use the contract to minimize lock-in effect. For example, use of proprietary data format for storage by service provider makes the consumer's data unusable with another provider. Options for migration to other service providers must be addressed in the contract.
- ***Jurisdictions***: As service providers commonly operate across multiple jurisdictions. Under the general principles of freedom of contract, consumers and service providers have choice in determining the forum and the jurisdiction(s) that will be applied to their dispute(s) related to data misappropriation.
- ***Termination***: The contract must address the liabilities related to data misappropriation even after its termination (in normal or abnormal conditions).

Trade secrets in the cloud could be stored in different jurisdiction at the same time [27]. It is often neither practical nor viable to limit the storage to one jurisdiction, although as discussed above, contracts can be used to limit the storage to certain jurisdictions. Moreover, it was also mentioned that under the general principles of freedom of contract, consumers and service providers have choice in determining the jurisdiction(s) that will be applied to dispute(s) related to trade secret misappropriation. Although this may reduce some of the confusion and provide greater certainty for trade secrets protection in the cloud, the jurisdictional problems do not completely go away [43-45]. For example, in a typical cloud setup, where a trade secret is stored in many jurisdictions, it might be difficult to point to the location where the misappropriation has occurred. This is because the damage that gives rise to liability can also be distributed in the same manner as the setup of the cloud across different jurisdictions. In the following section we discuss current efforts and related issues for protection of trade secrets at cross-jurisdiction level.

2.3 Rule of Law and Protecting Trade Secrets in the Cloud

At cross-jurisdiction level, World Trade Organization's *Trade-Related Aspects of Intellectual Property Rights (TRIPS)* agreement provides certain basic remedies which signatory countries should make available to the owner of a trade secret in case of misappropriation [46, 47]. However, among the signatory countries, this benchmark does not successfully serve the purpose of prompting uniformity because it has not been implemented, or has been implemented with different specifications [48].

Likewise at EU level, Table 2.1 summarizes such disparity in legislative panorama of twenty seven members states of the European Union [49] for trade secret protection. It can be observed that, most of the member states have not applied the Intellectual Property (IP) law for trade secrets protection as per definition of TRIPS agreement since they do not consider rights in trade secrets to be Intellectual Property Rights (IPR). However, absence of a specific law e.g. IP law, does not seem to necessarily entail an inadequate level of protection for trade secrets. Sensitive information which meets certain minimum requirements is protected in all relevant regulations [50], see table 2.1. Nevertheless, absence of uniformity in different jurisdictions may lead to insubstantial retribution for misappropriation at the cross- jurisdiction level. To deal with such discrepancy, on 28 November 2013, the European Commission (EC) published a draft directive to harmonize trade secret protection across the EU. This directive aims at: a) making it easier for national courts to deal with the misappropriation, b) remove infringing products from the market, and c) make it easier for victims to receive compensation for violation of their trade secrets.

Table 2.1 Trade Secret Protections in EU-27

	Criminal Law	Labor Law	Civil Law	IP Law	Contract Law	Tort Law	Others
Greece	×	×	×	×		×	×
Belgium	×	×	×		×	×	×
France	×	×	×	×	×	×	
Romania	×	×	×	×			×
Austria	×	×		×			×
Hungary		×	×	×	×		×
Italy	×	×	×	×			×
Latvia	×	×				×	×
Netherlands	×	×	×			×	×
Spain	×	×		×			×
Bulgaria		×	×				×
Cyprus	×				×		×
Finland	×	×					×
Germany	×	×					×
Lithuania	×	×	×				×
Poland	×	×	×				×
Slovenia	×	×			×		×
Sweden	×						×
Denmark	×	×					×
Estonia	×	×					×
Ireland					×	×	
Luxembourg	×				×	×	×
Portugal	×	×		×			
Slovakia	×	×					×
Czech Rep.	×						×
Malta			×		×		
UK					×	×	

The successful application of the proposed directive by EC relies on the assumption that the location and responsibility of data is known and understood i.e. jurisdiction for a trade secret is transfixed (EU region). However, because of universal footprint of the cloud (cross-jurisdiction setup around the globe), proposed directive and similar regulations may fail to protect a trade secret in the Cloud.

2.4 Summary

This chapter discusses essential concepts necessary to understand the section of PhD research that is addressing the primary research question identified in section 1.1. In this regards, this chapter majorly discusses topic of data protection in the cloud. It was observed that the trust between the stakeholders in the cloud (consumer, service provider, auditor, and broker) is critical for planning, delivering, and consumption of cloud computing service and deployment models. One of the major issues that could aid or impair such trust is data protection. For an enterprise, data protection is protection of its business data or trade secrets in the cloud. Despite of the fact that contract can provide greater certainty for trade secrets protection in the cloud, the jurisdictional problems do not completely go away and may result in failure of legal protection of trade secrets in the cloud.

Chapter 3 : Related Work and Proposed Model

This chapter addresses the challenges that were presented in section 1.2. By doing so, it successfully answers the primary research question: *how an online broker can embed legal protection as preemptive measure to reduce burden of proof in a court of law?* The answer to this research question will benefit R&D based enterprises to negotiate a contract with service providers to minimize trade secret misappropriation in the cloud. Section 3.1 addresses the first challenge i.e. to build legal argument for protection of trade secrets in the cloud. It also addresses the second challenge of twofold transformation i.e. to find the technical concept that corresponds to legal argument and build related research question in ICT domain. Sections 3.2, 3.3, and 3.4 address the third challenge. Section 3.2 presents review of ICT literature to check if the answer to research question (in ICT domain) already exists or not. As it was not, section 3.3 proposes a solution and section 3.4 presents its technical evaluation in a cloud environment. Section 3.5 addresses the fourth challenge i.e. to legally validate the results of this chapter. Finally, section 3.6 summaries the discussion and findings in the chapter.

3.1 Related Work (Law – Case Law Analysis)

Considering the gap identified in section 2.3 i.e. *because of universal footprint of the cloud (cross-jurisdiction setup), regulations around the globe may fail to protect a trade secret in the cloud*, and to investigate plausible implementation of law for a trade secret protection in the cloud, in the domain of "case law", precedents set by previous court rulings in United States of America (USA) were identified, see table 3.1.

Table 3.1 Precedents set by Court Rulings for Trade Secret Protection in USA

	Precedent	Court Cases
Presence	Customer can store different types of data in the cloud. However, based on opinions in cases 1, 2, and 3, not all of them would come within the ambit of trade secret protection until data is not generally known to industry or public and the Customer has taken all possible measures to keep it secure.	COURT CASE 1: Religious Technology Ctr. v. Netcom On-Line Communication Servs[a]: One of the leading opinions in this case was, *“even if one person knows about the trade secret that could derive economic benefit from it, then the data could lose its trade secret status”.* But what if the data stored by Customer in the cloud has open source elements in it e.g. source code derived from open source software? In Essex Group v Southwire Corp.[b], the court stated that “*the trade secret can exist in a combination of characteristics and components, each of which is in public domain, but the unified process design and operation of which in unique combination, affords a competitive advantage and protective trade secret*”. COURT CASE 2: J.T. Healey & Son, Inc. v. James A. Murphy & Son, Inc.[c]: One of the leading opinions in this case was, “*if the person entitled to a trade secret wishes to have its exclusive use in his own business, he must not fail to take all proper and reasonable steps to keep it secret. . .*”. COURT CASE 3: Merrill Lynch, Pierce, Fenner & Smith, Inc. v Dumm[d]: One of the leading opinions in this case was, “*the trade secret owner has to take reasonable efforts to maintain secrecy*”.

Confidentiality	*Structural Significance*: Service provider can provide different criteria for security e.g. encryption, firewalls, access control etc. If a Customer fails to endorse significance of these criteria as per intend or a goal e.g. trade secret protection, then based on opinions in cases 4 and 5, he has not exercised a reasonable effort to maintain secrecy of the trade secret.	COURT CASE 4: Carboline Co v. Lebeck[e]: One of the leading opinions in this case was, ”*the trade secret owner had not taken reasonable measures as per intend to maintain secrecy where, among other things, it took no measures to protect information in the hands of suppliers or customers*”. COURT CASE 5: Heartland Home Fin., Inc v. Allied Mortgage Capital Corp.[f]: One of the leading opinions in this case was, *“the use of an encrypted email to transmit the alleged trade secret and the password protection were insufficient as per intend (given the lack of other security criteria)”.*

	Contract Compliance: If a customer uses cloud services that discloses trade secret to a service provider then based on opinions in cases 6 and 7, data will not lose its trade secret status if a contract between the two complies with non-disclosure regulations.	COURT CASE 6: Lac Minerals Ltd. v. International Corona Resources Ltd.[g]: One of the leading opinions in this case was, *"A duty of confidence arises when a person acquires knowledge of confidential information, including trade secrets, under circumstances in which the person has notice or agreed that the information is confidential as per law"*. COURT CASE 7: Saltman Engineering Coy Ltd. v. Campbell Engineering Coy. Ltd[h]: One of the leading opinions in this case was, *"if information is given by one trader to another in circumstances which make that information confidential as per law, then the second trader is disentitled to make use of the confidential information for purposes of trade by way of competition with the first trader"*.

Misappropriation	Based on the opinions in cases 8 and 9, it can be established that performing big data analytics is unlawful when: a) Big Data is obtained illegally or b) contract is breached during its lifetime or even after termination.	COURT CASE 8: Kewanee Oil Co. v. Bricon Corp.[i]: One of the leading opinions in this case was "*trade secret law imposes a liability only when the data is obtained by improper means or under breach of an agreement. It does not impose a liability for mere copying of the data; others are free to inspect the publicly available data to reverse engineer to procure secret information from it*". COURT CASE 9: Cadbury Schweppes v. FBI Foods Ltd.[j]: One of the leading opinions in this case was, *"a licensor revealed to the licensee, under license, confidential information about a recipe for a tomato cocktail with clam broth. After receiving notice to terminate the license, the licensee used the confidential information to develop a competing product. The court held the licensee was under an obligation to protect the trade secret even after termination of the license"*.

[a] Religious Technology Ctr. v. Netcom On-Line Communication Servs, 10 Cal. Rptr. 3d (2004)

[b] Essex Group v Southwire Corp, 269 Ga.553,501 S.E.2d 501(1998)

[c] J.T. Healey & Son, Inc. v. James A. Murphy & Son, Inc., 357 Mass. 728, 737-39 (1970)

[d] Merrill Lynch , Pierce, Fenner & Smith, Inc. v Dumm, 191 F.Supp.2d 1346,1351 (M.D. Fla.2002)

[e] Carboline Co v. Lebeck, 990 F.Supp.762,767,-68 (E.D. Mo. 1997)

[f] Heartland Home Fin., Inc v. Allied Mortgage Capital Corp,No 1:05 CV 2659,2007 U. S Dist. LEXIS 8882

[g] Lac Minerals Ltd. v. International Corona Resources Ltd, [1989] 2 S.C.R. 574

[h] Saltman Engineering Coy Ld. v. Campbell Engineering Coy. Ltd, (1948)

[i] Kewanee Oil Co. V. Bricon Corp. 416 U.S .470 (1974)

[j] Cadbury Schweppes v. FBI Foods Limited , [1999] 1 S.C.R. 142

For misappropriation claim of trade secret in the cloud, table 3.1 shows that the plaintiff[7] must establish three things in a court of law. They are: a) *presence:* it's a proof of data in the cloud to be a trade secret, b) *confidentiality:* it's a proof for reasonable efforts made by the owner to protect trade secret in the cloud, and c) *misappropriation:* it's a proof for misappropriation of a trade secret by using big data analytics. Furthermore, to ensure reasonable efforts are in place for confidentiality, owner must also assess structural significance of criteria and inspect contract (or electronic contract) for compliance with non-disclosure regulations. Whereas, structural significance of criteria is similar to the concept of coefficient of determination in statistics [51]. Statistically, it's a "shared and common variance" among the criteria that represents a goal [52]. Its low value indicates presence of irrelevant criterion or absence of relevant criterion in relation to a goal.

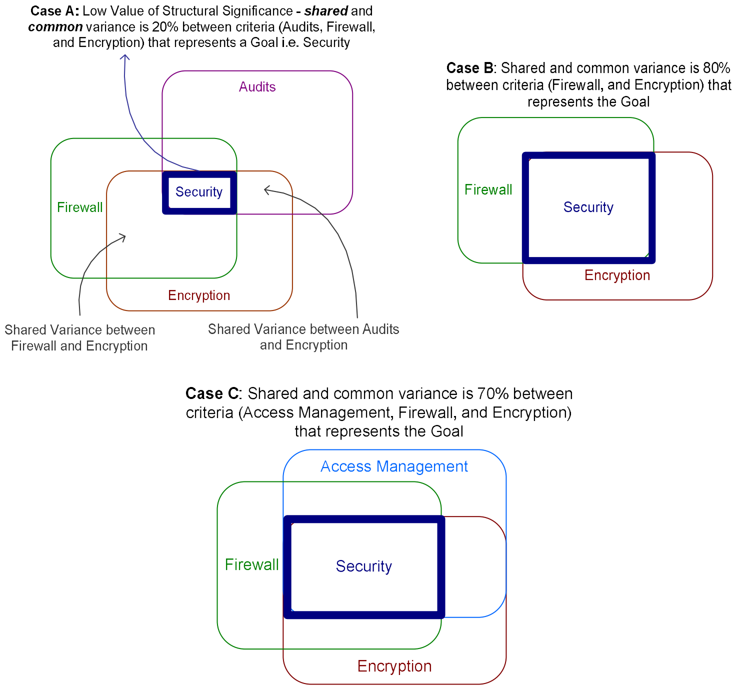

Case A: Low Value of Structural Significance - shared and common variance is 20% between criteria (Audits, Firewall, and Encryption) that represents a Goal i.e. Security
Audits
Firewall
Security
Encryption
Shared Variance between Firewall and Encryption
Shared Variance between Audits and Encryption
Case B: Shared and common variance is 80% between criteria (Firewall, and Encryption) that represents the Goal
Firewall
Security
Encryption
Case C: Shared and common variance is 70% between criteria (Access Management, Firewall, and Encryption) that represents the Goal
Access Management
Firewall
Security
Encryption

Fig. 3.1 Structural Significance of Criteria

[7] a person who brings a case against another in a court of law.

For example, figure 3.1 present three hypothetical cases with different values of structure significance of criteria. In case A, a pictorial presentation shows 20% of the shared and common variance between criteria (Audits, Firewall, and Encryption). As per contribution, Audits is a least relevant criterion in relation to a goal i.e. Security. In case B, after omitting Audits as an irrelevant criterion, 80% of shared and common variance is depicted between Firewall and Encryption in relation to the goal. In case C, a new criterion of Access management is added to the Case B and variance is depicted to be 70%. Among these three cases, Case B shows the highest structural significance of criteria i.e. 80%, in relation to the goal. However, in case C structural significance is also high i.e. 70%, which, in addition to Firewall and Encryption, justifies presence of Access management as a relevant criterion in relation to the goal.

The immediate lesson from preceding paragraph is that a misappropriation claim with the proofs for presence, confidentiality, and misappropriation is a sure recipe for litigation. However, as per conclusion of *JetBlue Airways Corp. Privacy Litigation* in chapter 1 - page 3, it is plausible that a fully fleshed-out proof for confidentiality that include evidence for structural significance and contract compliance, may complicate the burden of proof during the litigation. Thereupon, as per outcome of discussion on *EPIC v. the Department of Homeland Security* in chapter 1 – page 3 and 4, it is implied to use online broker to reduce such burden by embedding legal protection as preemptive measure. However, for online broker to do so, it must be capable to (1) inspect contract (or electronic contract) for compliance with non-disclosure regulations and (2) assess structural significance of criteria. For an affirmative response to both these requirements, the broker can then be assumed to be successfully providing legal protection for trade secrets in the cloud and subsequently, reducing burden of proof in a court of law.

3.2 Related Work (ICT – Systematic Review)

A review of relevant literature was performed to examine the status of online brokers for (1) inspecting contract compliance with non-disclosure regulations and (2) assessing structural significance of criteria. It was learned that services of online brokers are still at their initial level when it comes for provisioning legal protection. For example, the model for regulation aware online broker required for inspection of a contract for compliance with non-disclosure regulations has been recently developed in [11]. Moreover, it was also observed that, unlike contract compliance, structural significance is not directly and distinctly expressed in the reviewed literature. Therefore an additional attempt was made to analyze underlying contents by performing systematic review. Systematic review uses transparent procedure to find and analyze results of relevant research. This procedure is explicitly defined in advance in order to ensure that it can be replicated afterwards.

For systematic review, the research published between January 2010 and March 2017 was explored by using the following databases: ACM Digital Library, Google Scholar, IEEE Xplore, ScienceDirect, and SpringerLink. The primacy search term was “cloud service provisioning models”. Figure 3.2 present chronological distribution of identified models [37, 53-73]. Right hand side models uses data mining, whereas, left hand side models apply multi-criteria decision analysis (MCDA). It is evident from the figure that MCDA is the prevalent technique and hence, only MCDA based models were selected to identify an approach that is used by online broker to assess structural significance of criteria.

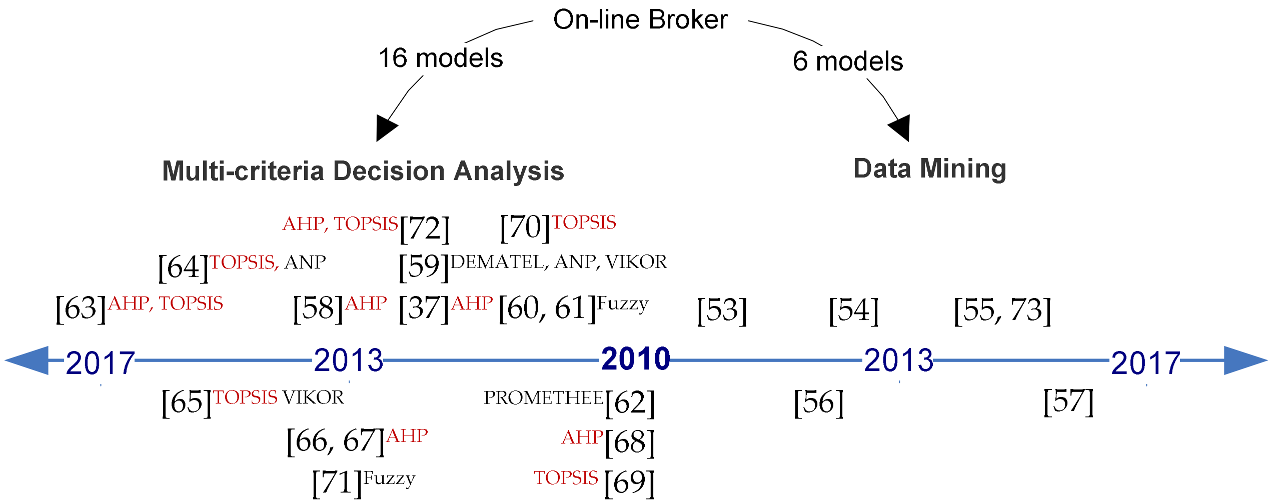


Fig. 3.2 Chronological Distribution of Models for Online Broker

MCDA is a methodology that deals with *objective, criteria,* and *alternatives* to reach a pre-established *goal* [37]. The goal or an overarching principle for an online broker could be the ranking of service providers. Whereas, the objective i.e. specific and measurable step, set to reach the goal could be data security. Once the objective is fixed, it is then necessary to establish criteria that are used to evaluate alternatives leading to the objective. For example, to evaluate service providers for data security in the cloud, online brokers can check type of security group in use. Security group is a virtual firewall that controls data flow in the cloud; therefore, service provider with its upmost implementation will be a leading alternative in the ranking.

During the review of MCDA based models, it was observed that the well-established *goal* for MCDA based online brokers is either ranking of service providers or optimization of cloud resources. In particular, optimization is realized through an *objective* of agility i.e. to sense opportunities or threats and allocate *alternatives* in an efficient and timely manner. The most common *criteria* used to observe such change is quality of service (QoS) e.g. response time, execution time, utilization etc. In [37] authors propose a broker for distributed resources management in the cloud using Analytic Hierarchy Process (AHP). They argue that, unresolved QoS issues cause service provider to suffer from unacceptable levels of performance. In this regards, AHP is used to recognize changes by performing pairwise comparison of system attributes structured in a hierarchal relationship. For a broker, such system is composed of resources and tasks. Incoming tasks are stored in a matrix configuration and sorted as per their priority that is measured by QoS criteria such as price or deadline etc. Likewise, resource matrix contains information on QoS of all resources. Overall, a broker contains two matrices, one for tasks and other for resources. The solution is to match the two in order for service provider to fetch the maximum return as per performance. In [58] authors propose a task-oriented-scheduling mechanism using AHP. They argue that, resource allocation is a complicated task in the cloud as there are many alternatives with varying capacities. In proposed mechanism, tasks are pairwise compared according to network bandwidth, complete time, task cost, and reliability of a task. Afterwards, weight for each task is calculated using AHP and resources are allocated respectively. In [72], a proposed model uses AHP and fuzzy based Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) to decide which cloud is the most suitable for offloading of tasks in fuzzy environments. Authors argue that, to extend the battery life and reduce execution time on mobile devices, computation tasks can be offloaded to the cloud. However, offloading the same task to different clouds may result in dissimilar amounts of computing (per unit time) due to difference in QoS. In this context, proposed model uses AHP to calculate task priority, then uses fuzzy based TOPSIS to identify an alternative (cloud) that is simultaneously closest to the ideal solution i.e. cloud with desired QoS, and the farthest from the anti-ideal solution and finally, perform offloading of tasks to this cloud as per assigned priorities. In [59] authors propose a model using Decision Making Trial and Evaluation Laboratory (DEMATEL), DEMATEL based Analytic Network Process (ANP), and VIKOR. They argue that, understanding Customer intentions and behaviors with regards to cloud services will help service providers to identify factors that affect their use and subsequently performance. ANP closely relates to AHP. While, AHP structures a decision problem into a hierarchy, ANP do it as a network. For proposed model, DEMATEL is used to construct a fuzzy scope influential network relationship map (FSINRM), which is then utilized to illustrate the influential relationships among criteria related to cloud services. Subsequently, DEMATEL based ANP and VIKOR methods are used to determine weights of criteria and gaps from the desired level of service delivery. The average gap between the actual and desired level indicate deficiencies in cloud services that must be addressed to improve performance. In [60] authors proposes dynamic service placement and replication (DSPR) framework to manage cloud services in a distributed environment. They argue that, services running on cloud still require service provider to plan distributed architecture carefully to leverage on the scalability offered by the cloud. In this regards, DSPR introduces a fuzzy inference engine to perform resource evaluation and allocation. DSPR uses team formation algorithm to continuously shift services to servers with better performance and at the same time, dynamic service replication algorithm autonomously form server pools to guarantee scalability. In [61] authors propose a model for resource allocation using a self-tuning fuzzy controller (STFCs). They argue that, design of an accurate and stable controller is challenging when response time is considered as a measured output. In this regards, DynaQoS is proposed as a two-layer QoS provisioning framework. The first layer is composed of a set of STFCs that measure response time, whereas, the second layer combines the requests from multiple STFCs to generate a single output for a resource management module to perform resource allocation. In [62] authors propose a new approach for dynamic autonomous resource management in cloud. They argue that, the optimal allocation of cloud resources such as virtual machines eventually relates to high profits for a service provider. In this regards, proposed approach perform dynamic resource management where main management task is further decomposed into independent subtasks. Each subtask is then performed by autonomous node agents (NA). NA uses PROMETHEE that perform QoS based pairwise comparison among alternatives i.e. resources, to identify and eliminate the alternative that is dominated by the other.

On the contrary, when the *goal* of an online broker is to generate ranking of service providers, the corresponding *objective* is benchmarking i.e. to assign relative weights to *alternatives*. The most common *criteria* used for assigning such weights are QoS e.g. security, reliability, availability etc. In [63] authors proposes a hybrid decision-making model based on affinity diagram, fuzzy AHP (FAHP) and fuzzy TOPSIS (FTOPSIS) to evaluate cloud solutions to host Big Data projects. In the first stage of this model, identification of evaluation criteria is performed by a decision-making committee using Affinity Diagram. Due to the varied importance of the selected criteria, a FAHP process is used in the second stage to assign weights for each criterion. FTOPSIS in the third stage employ these weighted criteria as inputs to evaluate and measure the performance of each alternative (cloud solutions). In the last step, a sensitivity analysis is performed to evaluate the impact of criteria weights on the final rankings of alternatives. In [64] authors discusses evaluation of Trade-offs based Methodology for Adoption of cloud based Services (TrAdeCIS) using TOPSIS and ANP. They argue that the decision to use such services is based upon criteria which can be mutually interdependent and conflicting and hence, a trade-offs-based methodology is needed to make such decisions. TrAdeCIS is the first methodology that supports an automated and quantified trade-offs based decision making for selection of a best cloud based service. In [65] authors compares behavior and quality of TOPSIS and VIKOR based multi-objective decision methods with the Pareto optimality solutions. In [66] authors propose a Service Measurement Index Cloud framework (SMICloud). It provides a holistic view of criteria to benchmark service providers. It is divided into seven categories that include accountability, agility, assurance, financial, performance, security and privacy, and usability. Each of these categories is further subdivided into three or more mid-level criteria. For example, mid-level criteria assigned to agility include, beside others, capacity and elasticity. Then within each mid-level criterion, a set of low-level criteria are defined for data collection. For example, low-level criteria assigned to capacity include, beside others, CPU and memory. For each criterion in these levels, relative weights are assigned using AHP to generate relative ranking. In [67] authors propose consumer centered cloud service selection model. They argue that, QoS criteria in the cloud are solely related to service provider. However, as cloud service spread all over the internet, part of them (e.g. availability and reliability) are largely influenced by a network which eventually impact Customers. For this reason, selection of a cloud service must be subjected to Customers interest. In this regards, AHP is used for ranking of service providers based on Customer preferences. In [68] authors propose fuzzy based AHP model for cloud service selection. They argue that, it is often difficult for a Customer to exactly quantify his or her opinion as a number. However, if expressed as an interval then it will be better description of an opinion. In this regard, proposed model combined interval valued fuzzy sets (IVFs) with AHP to generate ranking. In [69] authors propose fuzzy based TOPSIS model for cloud service selection. They argue that, QoS based cloud service selection can be treated as a multi-criteria group decision making problem when selection is performed by a group of experts with different experiences and skills. In this regard, proposed model uses triangular fuzzy numbers to represent opinions of experts. Afterwards, these fuzzy numbers are transformed into crisp numbers by using graded mean integration representation method. The canonical representation of addition and multiplication operations on triangular fuzzy numbers is then used to obtain the positive ideal solution (PIS) and the negative ideal solution (NIS). Due to the use of crisp number rather than triangular fuzzy number for canonical representation, the complicated calculations involving triangular fuzzy numbers is avoided. Afterwards, Minkowski distance function is applied to measure the distance of each alternative (cloud service) from the PIS and the NIS. The shortest distance from the PIS and the farthest distance from the NIS is selected as a best alternative. In [70] authors propose a model which uses Fuzzy TOPSIS for web service selection. Based on the fact that web service selection is highly influenced by Customer preferences, a simulated environment represented by 8∗8 LED matrices on a circuit board was used to demonstrate the selection. In [71] authors propose a cloud service selection model that uses subjective assessment of Customers and objective performance assessment conducted by a trusted third party. The model is composed of four services: (i) Cloud Selection Service – it chooses cloud services which meets all the objective requirements of a Customer; (ii) Benchmark Testing Service – this service is provided by a trusted third party which designs a variety of testing scenarios to conduct objective performance analysis; (iii) User Feedback Management Service – it is used to collect and manage the feedback from the Customers who are already consuming selected cloud services. For every performance aspect of a cloud service, a customer gives his/her subjective assessment (e.g., “good”, “fair” and “poor”); and (iv) Assessment Aggregation Service – it is responsible for accumulating assessments (subjective and objective) and perform benchmarking using fuzzy simple additive weighting system to generate ranking.

Table 3.2 Underlying Techniques and MCDA based Models

Underlying Technique	Objective	MCDA based Models
AHP	Pair-wise comparison of elements structured in a hierarchal relationship.	[37] [58] [63] [66] [67] [68] [72]
TOPSIS	Criteria based selection of an alternative that is closest to the ideal solution.	[63] [64] [65] [69] [70] [72]
Fuzzy	Evaluation of weights in terms of linguistic values represented by fuzzy numbers belonging to criteria.	[60] [61] [71]

Table 3.2 lists top three mostly used underlying techniques employed by MCDA based models discussed in the preceding paragraphs. They are: AHP, TOPSIS, and Fuzzy. However, among the three as in due course, AHP and TOPSIS are the most prevalent techniques as shown by left hand side models in figure 3.2. For AHP the prime objective is to decompose the decision problem into a hierarchical structure of objective, criteria and alternatives. Afterwards, evaluate them in a series of pair-wise comparisons that uses priorities provided by the decision maker [67]. TOPSIS on the other hand, compares a set of alternatives by using weights for each criterion provided by the decision maker. Afterwards, it calculate the geometric distance between each alternative and the expected ideal alternative [69].

It is evident that AHP and TOPSIS use distinct approaches to evaluate alternatives. However, at the very outset, they equally reply upon subjective judgments of the decision maker to ensure that all relevant criteria are included in the process. Apparently, this leads to conclusion that *MCDA based online brokers that use AHP or TOPSIS assume structural significance for criteria owning to subjective judgments of the decision maker.* In general, this conclusion reaffirms the observation identified in beginning of this section that an online broker is still at initial level when it comes to provisioning legal protection. Whereas explicitly, it acknowledges a need to develop a model that can assess structural significance of criteria for MCDA based online brokers that are using AHP and TOPSIS.

3.3 Proposed Model

As discussed in section 3.1, structural significance is a shared and common variance among criteria that represent a goal. To measure such variance, this part of PhD research uses notion of "factor loading" that belongs to broader concept of factor analysis from the domain of Unsupervised Machine Learning [52, 74]. However, despite of factor analysis being a technique for inferential statistic i.e. it is used to make generalizations; its results in this research do not extend beyond the given instance. Therefore, the prerequisites for generalization e.g. selecting a sample size, become void in this research.

Factor loading is a measure of a correlation between a criterion and a goal [52]. Such association can be linear or nonlinear in nature. As a stepwise progression, this research deal with the former as follows, whereas, the latter will be addressed in the future research.

$$\begin{aligned} \mathrm{x}_1 &= \lambda_1 \mathrm{f} + \mathrm{e}_1 \\ \mathrm{x}_2 &= \lambda_2 \mathrm{f} + \mathrm{e}_2 \\ \mathrm{x}_3 &= \lambda_3 \mathrm{f} + \mathrm{e}_3 \\ &\vdots \\ \mathrm{x}_\mathrm{n} &= \lambda_\mathrm{n} \mathrm{f} + \mathrm{e}_\mathrm{n} \end{aligned} \tag{1}$$

where,

n is total no of criteria

x_i is a criterion, where $0 < i \leq \mathrm{n}$

f is a goal

λ_i is a factor loading of x_i on f

e_i is a uniqueness of x_i not related to f

As correlation coefficient in above system of equations (1), factor loading (λ_i) measures the strength and the direction of a linear relationship between a goal (f) and a criterion (x_i). Its squared value $(\lambda_i)^2$ is called as communality, which is a shared and common variance of the criterion for the goal [52]. Whereas, structure significance of criteria (SS_c) i.e. shared and common variance among criteria, is the sum total of all communalities $(\sum(\lambda_i)^2)$. On percentage scale, it is given as:

$$\mathrm{SS_c} = \sum(\lambda_\mathrm{n})^2/\mathrm{n} \qquad (2)$$

However, above equation may fail to provide optimal results until it satisfies $(\lambda_i)^2 > \omega$. Where, ω is a controlled variables (or constant) and its value is assigned by a substantive specialist in the field or a statistical technique [52]. The value of ω lies between 0 and 1 and is used for identification of relevant criterion. For example, $\omega = 0.65$ ensure that a criterion which contributes more than 65% to the goal is selected for further processing. In figure 1.1, such was the case for “Firewall and Encryption in case B” and “Firewall, Encryption, and Access management in case C”. Accordingly, equation 2 can be rewritten as:

$$\mathrm{SS_c} = \left(\frac{\sum(\lambda_\mathrm{k})^2}{\mathrm{k}} \text{ where } 0 < \mathrm{k} \leq \mathrm{n} \text{ and } (\lambda_\mathrm{k})^2 > \omega\right) \qquad (3)$$

Equation 3 presents a model to assess structural significance of criteria for MCDA based online brokers that are using AHP and TOPSIS. In this model, the value of λ_k is estimated by Structural Equation Modeling (SEM). SEM is a statistical approach used to examine association between a latent variable(s) and observed variables [52, 74]. Latent variable is a theoretical construct that is analyzed through variables that are observed during the test or survey. For example, goal (f) in system of equations (1) is a latent variable since it represents intent of a Customer e.g. trade secret protection, and it is analyzed through variables (or criteria) $x_1, x_2, \dots, x_n$ that are observed during the test or survey e.g. data encryption, password protection, access control etc.

In SEM, the most popular and frequently used methods used to estimate λ_k are *Principal Factor Analysis (PFA)* and *Maximum Likelihood (ML)* [52, 74]. Considering that ML estimation assumes normal distribution of observed variables and this research is dealing with observed variables (or criteria) without making any prior assumption, so PFA is used to estimate λ_k. In PFA, the system of equations (1) that express linear associations between a latent variable and observed variables is summarized in the matrix expression as:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} [f] + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$\mathrm{X} = \Lambda\mathrm{F} + \mu \qquad (5)$$

where,

X is a $[n \times 1]$ matrix of $(x_1, x_2, \dots, x_n)$

F is a $[1 \times 1]$ matrix (or identity matrix) of f

Λ is a $[n \times 1]$ matrix of $(\lambda_1, \lambda_2, \dots, \lambda_n)$

μ is a $[n \times 1]$ matrix of $(e_1, e_2, \dots, e_n)$

In SEM, following two assumptions for variance (var) and covariance (cov) are linked to the system of equations (1) and equation 5 [51].

1. $var(e_i) = \psi_i$, each e_i have different variance ψ_i since it shows the respective uniqueness of x_i.
2. $cov(F, \mu) = 0$ and $cov(e_i, e_k) = 0, i \neq k$ implies that the latent variable account for all the correlations among the x_i, that is, all that the $x's$ have in common. Thus the emphasis in PFA is on modeling the correlations or covariance among the $x's$. And therefore, equation 5 in PFA is expressed in a variance-covariance matrix notation as:

$$cov(\mathrm{X}) = cov(\Lambda\mathrm{F} + \mu)$$

As per assumption $cov(F, \mu) = 0$, $\Lambda\mathrm{F}$ and μ are uncorrelated; therefore, the covariance matrix of their sum is the sum of their convince matrices.

$$cov(\mathrm{X}) = cov(\Lambda\mathrm{F}) + cov(\mu) \qquad (6)$$

Moreover, as per assumption $var(e_i) = \psi_i$ and $cov(e_i, e_k) = 0, i \neq k$, $cov(\mu)$ in above equation becomes:

$$cov(\mu) = \begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_n \end{bmatrix}$$

and reducing to ψ,

$$\begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_n \end{bmatrix} = \psi$$

Accordingly, we can write equation 6 as:

$$cov(\mathrm{X}) = cov(\Lambda\mathrm{F}) + \psi$$

By using covariance property $\mathrm{cov(AX) = A\,cov(X)\,A^T}$, $\mathrm{cov}(\Lambda\mathrm{F})$ in the right hand side of above equation can be expanded to following form:

$$cov(\mathrm{X}) = \Lambda\, cov(\mathrm{F})\, \Lambda^{\mathrm{T}} + \psi$$

Since F being an identity matrix has $\mathrm{cov}(\mathrm{F}) = 1$, $\Lambda\,\mathrm{cov}(\mathrm{F})\,\Lambda^{\mathrm{T}}$ in above equation can be reduced to:

$$cov(\mathrm{X}) = \Lambda\Lambda^{\mathrm{T}} + \psi$$

If X is not commensurate i.e. observed variables (or criteria) are measured in different units and scales, then standardized X is used. After standardization, covariance becomes correlation (r) and subsequently, covariance matrix $cov(\mathrm{X})$ becomes a correlation matrix R [74].

$$\mathrm{R} = \Lambda\Lambda^{\mathrm{T}} + \psi$$

If R shows no significant evidence of correlations then using system of equations (1) become void i.e. linear association does not exist, and it is suggested to use non-linear factor analysis. Otherwise, we can expand above equation as:

$$\begin{bmatrix} 1 & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} [\lambda_1\, \lambda_2 \ldots \lambda_n] + \begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_n \end{bmatrix}$$

Bringing ψ to left hand side,

$$\begin{bmatrix} 1 & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & 1 \end{bmatrix} - \begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} [\lambda_1\, \lambda_2 \ldots \lambda_n]$$

Preforming subtraction on left hand side,

$$\begin{bmatrix} 1-\psi_1 & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & 1-\psi_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} [\lambda_1\, \lambda_2 \ldots \lambda_n]$$

Subtracting unique variance from the one i.e. $1 - \psi_i$, will yield shared and common variance of an observed variable (criterion) for the latent variable (goal). And as mentioned in the start of this section, such variance is represented by communality $(\lambda_i)^2$. Respectively, $(\lambda_i)^2$ can replace $1 - \psi_1$.

$$\begin{bmatrix} (\lambda_1)^2 & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & (\lambda_n)^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} [\lambda_1\, \lambda_2 \ldots \lambda_n] \qquad (7)$$

where,

$$\begin{bmatrix} (\lambda_1)^2 & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & (\lambda_n)^2 \end{bmatrix} = \mathrm{R} - \psi \qquad (8)$$

Accordingly, in a reduce form, equation 7 becomes:

$$\mathrm{R} - \psi = \Lambda\Lambda^{\mathrm{T}} \qquad (9)$$

$\mathrm{R} - \psi$ is a 'reduced correlation matrix' with $(\lambda_1)^2$ on the diagonal. If $\mathrm{R} - \psi$ is positive semi-definite matrix i.e. it satisfy $\mathrm{R} - \psi = (\mathrm{R} - \psi)^{\mathrm{T}}$, then this implies that left hand side in equation 9 is symmetric and has a following spectral decomposition [74].

$$\mathrm{R} - \psi = \mathrm{UDU}^{\mathrm{T}} \qquad (10)$$

Spectral decomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvectors to identify latent variable(s) and corresponding eigenvalues to show strength of identified latent variable(s). In equation 10, U is the matrix of eigenvectors of $\mathrm{R} - \psi$ and D is the diagonal matrix of corresponding eigenvalues $\Theta_1\ \Theta_2 \ldots \Theta_{\mathrm{n}}$.

$$\mathrm{D} = \begin{bmatrix} \Theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Theta_n \end{bmatrix}$$

The important property of a positive semi-definite matrix is that its eigenvalues are always positive or null [74]. Hence, $\Theta_{\mathrm{i}} \geq 0$ and consequently, D can be factored into:

$$\mathrm{D} = \mathrm{D}^{1/2}\mathrm{D}^{1/2}$$

where,

$$\mathrm{D}^{1/2} = \begin{bmatrix} \sqrt{\Theta}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\Theta}_n \end{bmatrix}$$

Accordingly right hand side in equation 10 becomes:

$$\mathrm{R} - \psi = \left(\mathrm{UD}^{\frac{1}{2}}\right)\left(\mathrm{D}^{\frac{1}{2}}\mathrm{U}^{\mathrm{T}}\right) \qquad (11)$$

Equation 11 is in the form of equation 9 and accordingly, following can be deduced for Λ.

$$\Lambda = \left(\mathrm{UD}^{\frac{1}{2}}\right)$$

In an expanded form, right hand side in above equation can be written as:

$$\Lambda = \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nn} \end{bmatrix} \times \begin{bmatrix} \sqrt{\Theta}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\Theta}_n \end{bmatrix}$$

It can be observed that Λ (or $\mathrm{UD}^{1/2}$) is $[n \times n]$ matrix, however, for this research which involve single latent variable F, Λ must be $[n \times 1]$ matrix - see equation 5. Hence from the right hand side of above equation we take the largest eigenvalue $(\Theta_i > \Theta_k, i \neq k)$ and corresponding eigenvector U_i for calculation of Λ [74].

$$\Lambda = U_{\mathrm{i}}\sqrt{\Theta}_{\mathrm{i}} \tag{12}$$

Expanding the right hand side in above equation,

$$\Lambda = \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{ni} \end{bmatrix} \times \left[\sqrt{\Theta_i}\right]$$

The eigenvector (or U_i) in equation 12 represents the latent variable F (or goal). Each value it contains is an estimated unit-scaled loading or weight (u_{ii}) that is associated with each observed variable or criterion (x_i). The eigenvalue Θ_{i} is a shared variance among all the observed variables or criteria that represents the latent variable. Expanding left hand side in above equation and taking square of both sides.

$$\begin{bmatrix} (\lambda_1)^2 \\ (\lambda_2)^2 \\ \vdots \\ (\lambda_n)^2 \end{bmatrix} = \begin{bmatrix} (u_{1i})^2 \times \Theta_i \\ (u_{2i})^2 \times \Theta_i \\ \vdots \\ (u_{ni})^2 \times \Theta_i \end{bmatrix}$$

Taking sum of the values in above equation.

$$\sum(\lambda_n)^2 = \sum[(u_{ni})^2 \times \Theta_i]$$

Replacing $\sum(\lambda_k)^2$ in equation 3 will give a following PFA based model to assess structural significance of criteria (SS_c^{PFA}) for MCDA based online brokers that are using AHP and TOPSIS.

$$\mathrm{SS}_{\mathrm{c}}^{\mathrm{PFA}} = \left(\frac{\sum[(\mathrm{u}_{\mathrm{ki}})^2 \times \Theta_{\mathrm{i}}]}{\mathrm{k}} \mid [(\mathrm{u}_{\mathrm{ki}})^2 \times \Theta_{\mathrm{i}}] > \omega \text{ where } 0 < \mathrm{k} \leq \mathrm{n} \right) \tag{13}$$

3.4 Technical Evaluation and Results

A two-stage procedure was implemented in order to evaluate proposed model (equation 13) for legal protection of trade secrets in the cloud. In stage one; structural significance of criteria was assessed by using the proposed model. In stage two, a comparative analysis was performed between two types of MCDA based online Brokers. One type included the assessment of structural significance of criteria while the other did not.

The dataset used in this part of the research is comprised of consumer feedbacks[8] on QoS of cloud storage providers. It was compiled from leading review websites, which acknowledges data (or trade secret) misappropriation in the cloud a major factor influencing the feedbacks. These feedbacks were provided for the following QoS based criteria (or observed variables): Availability (AV), Response Time (RT), Price (PR), Speed (SP), Storage Space (SS), Ease of Use (EU), Technical Support (TS), and Customer Services (CS). Each of these criteria was assessed on the following ordinal scale: excellent (5), very good (4), good (3), satisfactory (2), and sufficient (1). In total, the dataset contained 390 feedbacks for seven cloud storage providers that included: Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, SugarSync, and Zip Cloud. The latent variable (or goal) was “Customer trust”, which was analyzed from 390 feedbacks that were: (a) influenced by data (or trade secret) misappropriation in the cloud, (b) collected for QoS based criteria (AV, RT, PR, SP, SS, EU, TS, and CS) on ordinal scale, and (c) provided for cloud storage providers that include: Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, SugarSync, and Zip Cloud.

The data analysis, scripting, and visualizations tools used during this two-stage evaluation of proposed model includes: STATA – Data Analysis and Statistical Software, IBM Statistical Analysis Software Package (SPSS), and Microsoft Excel.

[8] TrustFeedback@http://cs.adelaide.edu.au/~cloudarmor/ds.html

3.4.1 Structural Significance of Criteria

Following six steps assess structural significance of criteria. ***Step 1:*** the correlation matrix (R) is generated for QoS based criteria using the dataset. As these criteria are assessed on ordinal scale, the generated matrix contains polychronic correlations that are used to measure associations between ordinal variables.

Step 1: Polychoric Correlation Matrix								
	AV	RT	PR	SP	SS	EU	TS	CS
AV	1	**0.763**	0.740	**0.767**	0.571	**0.716**	**0.828**	0.786
RT	0.763	1	0.736	0.724	0.605	0.714	0.703	0.746
PR	0.740	0.736	1	0.751	**0.722**	0.709	0.681	0.715
SP	0.767	0.724	**0.751**	1	0.660	0.714	0.712	0.718
SS	0.571	0.605	0.722	0.660	1	0.627	0.555	0.584
EU	0.716	0.714	0.709	0.714	0.627	1	0.650	0.681
TS	**0.828**	0.703	0.681	0.712	0.555	0.650	1	**0.814**
CS	0.786	0.746	0.715	0.718	0.584	0.681	0.814	1

Step 2: In order to generate reduced correlation matrix, initial estimates for $(\lambda_i)^2$ were required, see equation 8. In [52] author lists several approximation techniques, among which the most commonly used are the "average correlation of a variable with other variables" and the "highest correlation of a variable". In this research we have used highest correlation of a variable as an initial estimate for $(\lambda_i)^2$.

Step 2: Highest correlation as initial estimates of $(\lambda_i)^2$							
AV	RT	PR	SP	SS	EU	TS	CS
0.828	0.763	0.751	0.767	0.722	0.716	0.828	0.814

Step 3: Reduced correlation matrix $\mathrm{R} - \psi$ is generated with $(\lambda_i)^2$ on the diagonal of the matrix, see equation 9. $\mathrm{R} - \psi$ is positive semi-definite matrix i.e. it satisfy $\mathrm{R} - \psi = (\mathrm{R} - \psi)^{\mathrm{T}}$, and so it is symmetric and has a spectral decomposition as per equation 10.

Step 3: Reduced Correlation Matrix $(\mathrm{R} - \psi)$								
	AV	RT	PR	SP	SS	EU	TS	CS
AV	**0.828**	0.763	0.740	0.767	0.571	0.716	0.828	0.786
RT	0.763	**0.763**	0.736	0.724	0.605	0.714	0.703	0.746
PR	0.740	0.736	**0.751**	0.751	0.722	0.709	0.681	0.715
SP	0.767	0.724	0.751	**0.767**	0.660	0.714	0.712	0.718
SS	0.571	0.605	0.722	0.660	**0.722**	0.627	0.555	0.584
EU	0.716	0.714	0.709	0.714	0.627	**0.716**	0.650	0.681
TS	0.828	0.703	0.681	0.712	0.555	0.650	**0.828**	0.814
CS	0.786	0.746	0.715	0.718	0.584	0.681	0.814	**0.814**

***Step 4**:* Using equation 11, the greatest eigenvalue Θ_i and corresponding eigenvector U_i is obtained from $\mathrm{R} - \psi$.

Step 4: Greatest Eigenvalue (Θ_i) and Corresponding Eigenvector (U_i)								
	AV	RT	PR	SP	SS	EU	TS	CS
U_i	0.373	0.357	0.359	0.360	0.311	0.342	0.359	0.364
Θ_i	5.710							

***Step 5**:* Using equation 12, $\Lambda = (\lambda_1, \ldots, \lambda_n)$ is calculated and afterwards, $(\lambda_i)^2$ is calculated. Based upon the opinion of substantive specialist in the field, the value of ω is assigned to 0.65 to select a criterion that contribute more than 65% to the goal. The result in this step shows that Storage Space (SS) with the value $0.552 < 0.65$ must be omitted for further processing.

Step 5: Finding λ_i and $(\lambda_i)^2$								
	AV	RT	PR	SP	SS	EU	TS	CS
λ_i	0.890	0.852	0.858	0.860	0.743	0.817	0.857	0.869
$(\lambda_i)^2$	0.793	0.727	0.736	0.740	0.552	0.669	0.735	0.755

***Step 6**:* The calculations are performed again from step 1 to step 4 by excluding SS from the dataset and respectively, using equation 12, $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ is generated and afterwards, $(\lambda_i)^2$. The result in step 6 shows that none of criteria have value less than 0.65. Afterwards, using equation 13, structural significance of QoS based criteria is calculated to be 73%. Such high value of structure signifi-

cance justifies presence of “Availability, Response Time, Price, Speed, Ease of Use, Technical Support, and Customer Services” as relevant QoS based criteria for analysis of the latent variable (or goal) i.e. Customer trust.

Step 6: Finding λ_i and $(\lambda_i)^2$ without SS							
	AV	RT	PR	SP	EU	TS	CS
λ_i	0.905	0.854	0.841	0.853	0.807	0.856	0.872
$(\lambda_i)^2$	0.819	0.729	0.707	0.727	0.651	0.734	0.760
$\boldsymbol{SS_c^{PFA}}$ = **0.732 (73%)**							

3.4.2 Comparative Analysis

In this stage, a comparative analysis is performed between MCDA based online broker that is assessing structural significance of criteria and the one that is not. More specifically, it’s a comparison between traditional AHP (identified in section 3.2) and AHP based upon proposed model. Whereas, the prior performs series of pair-wise comparisons for eight QoS based criteria by using weights provided by the decision maker, and later uses seven QoS based criteria (excluding SS) and weights assigned to each criterion based on $(\lambda_i)^2$ in step 6 of preceding section. For example, AV with $(\lambda_i)^2 = 0.819$ has been given the highest weight, followed by CS, TS, RT, SP, PR, and EU. When in fact, for pair-wise comparisons of alternatives i.e. cloud storage providers, both uses priorities provided by the decision maker. Moreover, a similar setting was also applied for comparison between traditional TOPSIS (identified in section 3.2) and TOPSIS based upon proposed model.

The motivation for performing two pairs of comparative assessment i.e. traditional AHP v. AHP based upon proposed model and traditional TOPSIS v. TOPSIS based upon proposed model, lies in the context which represent a certain and uncertain online cloud environment. For simulating uncertainty, high degree of randomness was induced by using random probability distribution in the dataset for traditional TOPSIS v. TOPSIS based upon proposed model.

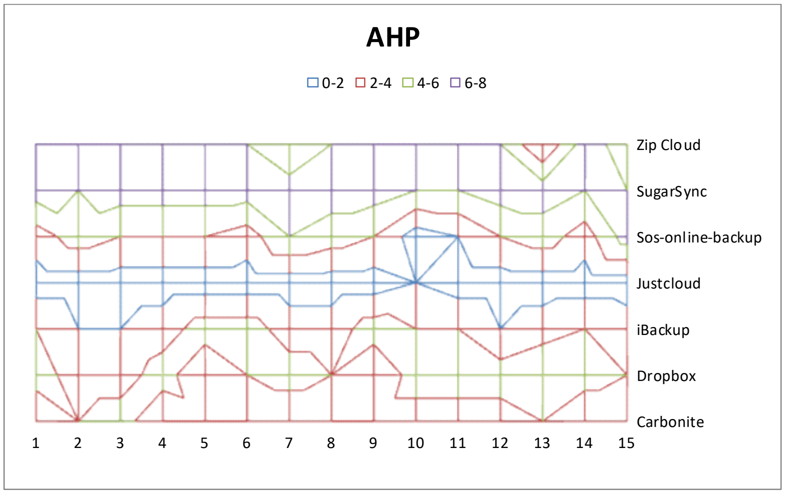
Figure 3.3 and 3.4 presents’ contour maps of comparative assessments for fifteen simulations. Each map represents four classes of Customer trust denoted by different colors. Blue color represents class of Customers with very high trust, red color represents class of Customers with high trust, green color represents class of Customers with some trust, and purple color represents class of Customers with low trust. For each simulation run in the map on top, the class membership is assigned to cloud storage providers on the basis of the ranking generated by traditional AHP and TOPSIS. Whereas, in the bottom map, the class membership is assigned on the basis of ranking generated by AHP and TOPSIS that are based upon proposed model.

In order to create direct correspondence between classes and generated ranking, the top two ranking positions are represented by the range of 0-2 in the maps and correspond to class of Customers with very high trust. Similarly, third and fourth positions are represented by range of 2-4 and correspond to class of Customers with high trust; fifth and sixth positions are represented by range of 4-6 and correspond to class of Customers with some trust; and lastly, seventh position is represented by the range of 6-8 and correspond to class of Customers with low trust.

In Figure 3.3, class memberships of cloud storage providers are much more explicit in the bottom map as compared to the map on top.

- Considering all simulations of both maps in Figure 3.3, it can be observed that Justcloud commonly falls in the class of Customers with very high trust. However, as per simulation 10 of the map on top, assigned membership for Justcloud is the class of Customers with high trust, whereas, for corresponding simulation in the bottom map, it is the class of Customers with very high trust.
- Considering all simulations of both maps in Figure 3.3, it can be observed that Zip Cloud commonly falls in the class of Customers with some trust and class of Customers with low trust. However, as per simulation 13 of the map on top, assigned membership to Zip Cloud is the class of Customers with high trust, whereas, for corresponding simulation in the bottom map, assigned class for Zip Cloud in is the class of Customers with some trust.
- Considering all simulations of both maps in Figure 3.3, it can be observed that SugarSync commonly falls in the class of Customers with low trust. However, as per simulations 10 and 11 of the map on top, assigned

membership for SugarSync is the class of Customers with some trust, whereas, in corresponding simulations in the bottom map, it is the class of Customers with low trust.



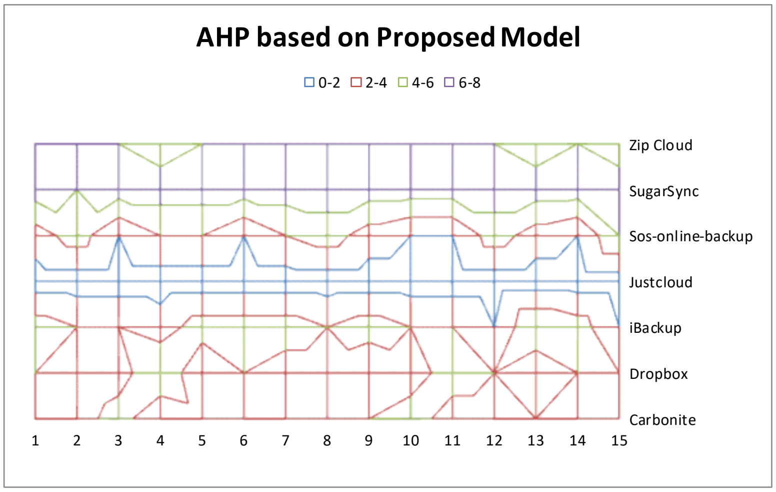


Fig. 3.3 Comparative Assessment of Proposed Model with AHP

For remaining cloud storage providers (Sos-online-backup, iBackup, Dropbox, Carbonite) the assigned memberships does not shown any significant improvement in the bottom map as compared to map on the top.

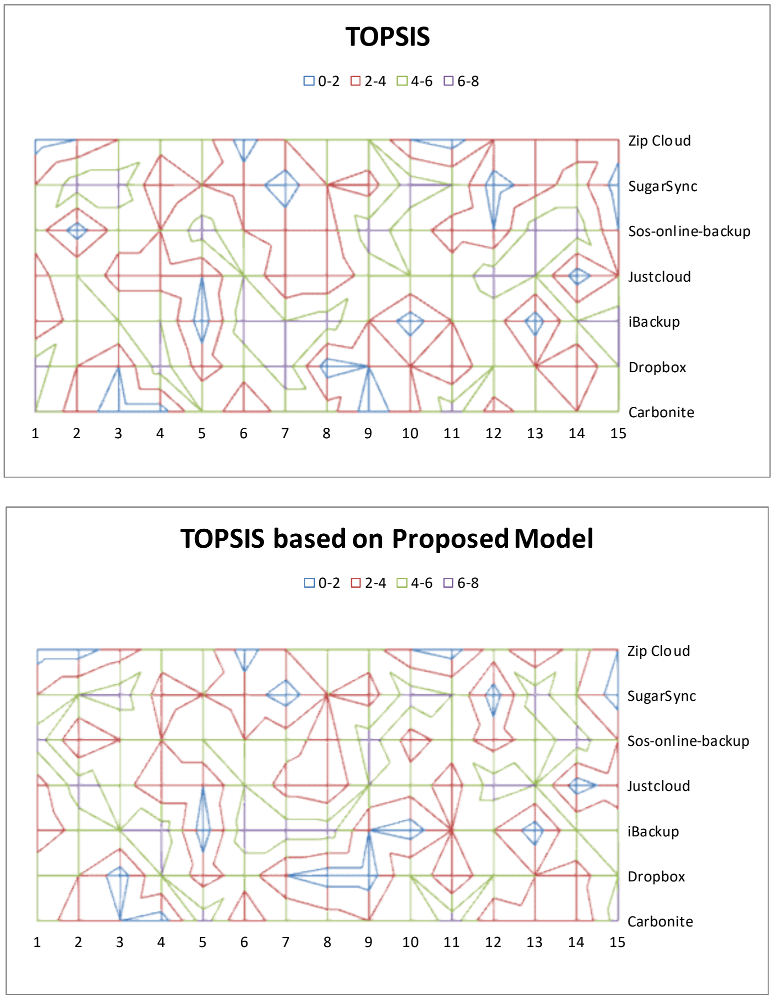


Fig. 3.4 Comparative Assessment of Proposed Model with TOPSIS

In figure 3.4, both the maps clearly show effects of induced uncertainly and respectively, the memberships of every cloud storage provider span all over the four classes in all fifteen simulations. However, in the bottom map, for Sos-online-backup the membership has reduced from four classes to three classes as compared to the map on top. This certainly highlights the limited capacity of proposed model to produce better results even in presence of uncertainty. But, this also shows limitation of proposed model and suggests a direction of future research for assessing structural significance of criteria in the presence of uncertainly.

Based on above observations, it can be stated that MCDA based online brokers that are using AHP and TOPSIS equipped with the proposed model are producing more accurate classification of service providers in term of Customer trust. In fact, a benefit of achieving such accuracy in results is significantly related to litigation, particularly for reducing burden of proof in a court of law. This was highlighted in section 1.1 on page 3 during the discussion on how *privacy algorithm* reduces burden of proof based on the evidence that focuses on algorithm accuracy for preserving privacy.

Overall, the above conclusion regarding accuracy of results to reduce burden of proof in the court of law and results of section 3.4.1 concerning structural significance to embed legal protection as preemptive measure, shows that this part of PhD research has successfully addressed the main research question (*how an online broker can embed legal protection as preemptive measure to reduce burden of proof in a court of law?*) and have implemented notion of *confidentiality by design* in the cloud.

3.5 Legal Validation and Results

As mentioned in section 1.4, communication of normative and empirical research results between the disciplines of law and ICT is one of the barriers in achieving genuine interdisciplinary validation. The proposed method of Delphi Sampling (see section 1.6.3) is an approximation technique for universal validation of multidisciplinary research results. For legal validation of research findings of table 3.1, using Delphi Sampling, following two questions were sent as part of a questionnaire.

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

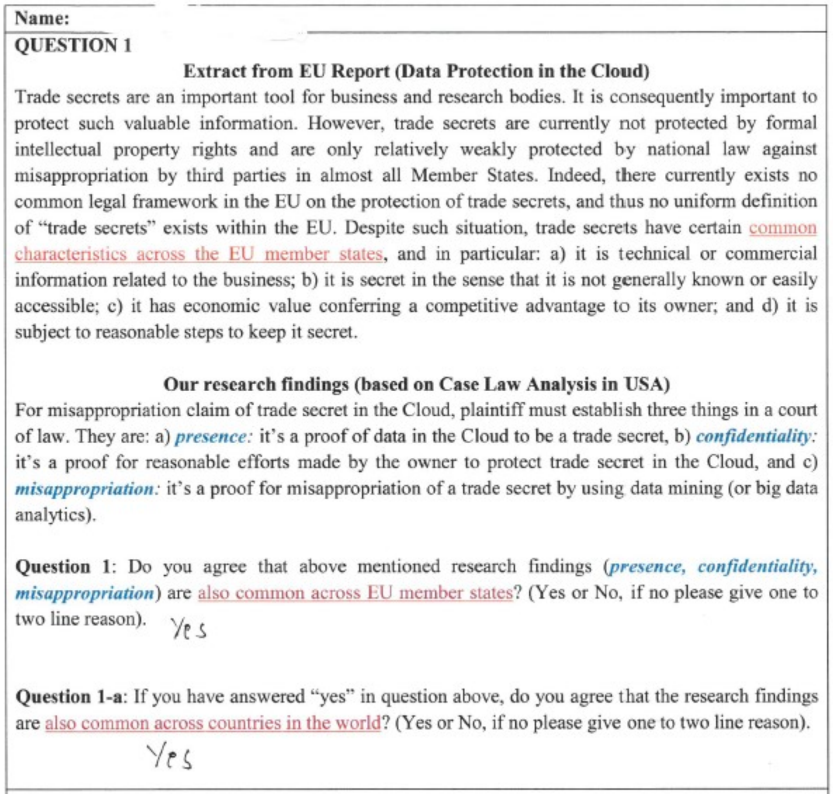
For misappropriation claim of trade secret in the cloud, plaintiff must establish three things in a court of law. They are: a) ***presence****:* it's a proof of data in the cloud to be a trade secret, b) ***confidentiality****:* it's a proof for reasonable efforts made by the owner to protect trade secret in the cloud, and c) ***misappropriation****:* it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason).

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

The screenshots of responses are presented below. There were a total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). There were two rounds. As per requirement of Delphi Sampling i.e. keeping anonymity in following rounds, names (of respondents) are hidden in the screen shorts. After two rounds the sample **(5 out of 6 respondents)** agreed that our research findings (*presence, confidentiality, misappropriation*) are common across EU member states and other countries in the world.

Respondent 1



Name:

QUESTION 1

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

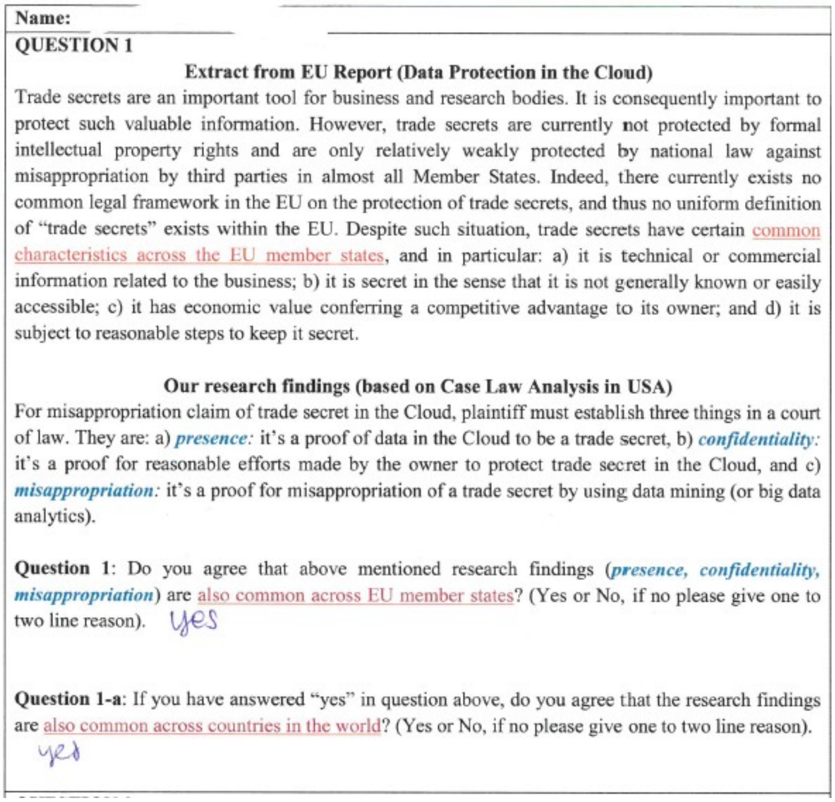
For misappropriation claim of trade secret in the Cloud, plaintiff must establish three things in a court of law. They are: a) ***presence:*** it's a proof of data in the Cloud to be a trade secret, b) ***confidentiality:*** it's a proof for reasonable efforts made by the owner to protect trade secret in the Cloud, and c) ***misappropriation:*** it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason). Yes

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

Yes

Respondent 2



Name:

QUESTION 1

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

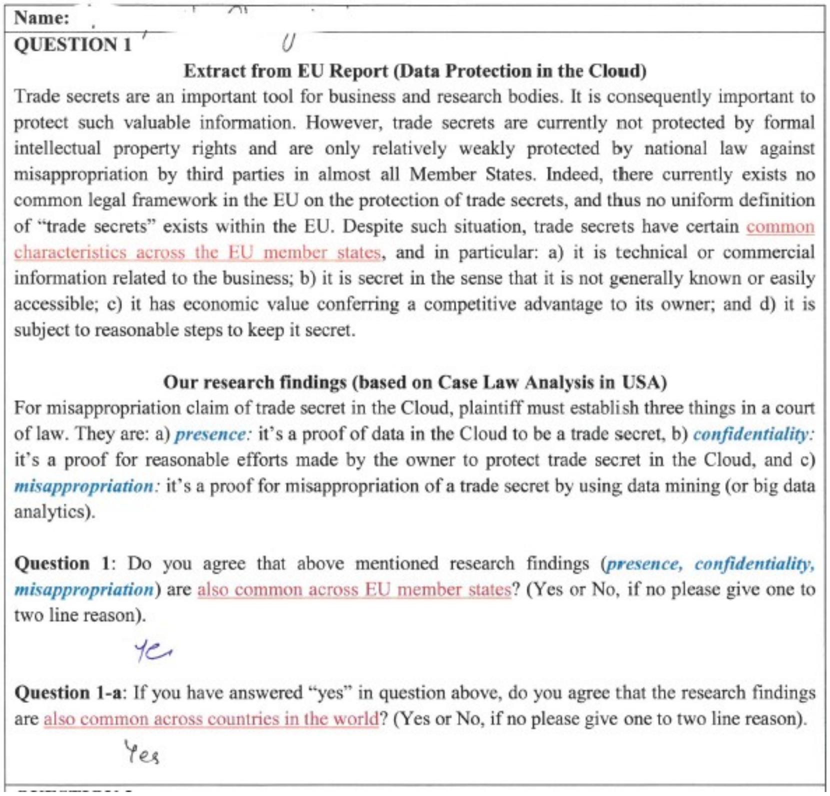
For misappropriation claim of trade secret in the Cloud, plaintiff must establish three things in a court of law. They are: a) ***presence:*** it's a proof of data in the Cloud to be a trade secret, b) ***confidentiality:*** it's a proof for reasonable efforts made by the owner to protect trade secret in the Cloud, and c) ***misappropriation:*** it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason). yes

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

yes

Respondent 3



Name:

QUESTION 1

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

For misappropriation claim of trade secret in the Cloud, plaintiff must establish three things in a court of law. They are: a) ***presence:*** it's a proof of data in the Cloud to be a trade secret, b) ***confidentiality:*** it's a proof for reasonable efforts made by the owner to protect trade secret in the Cloud, and c) ***misappropriation:*** it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

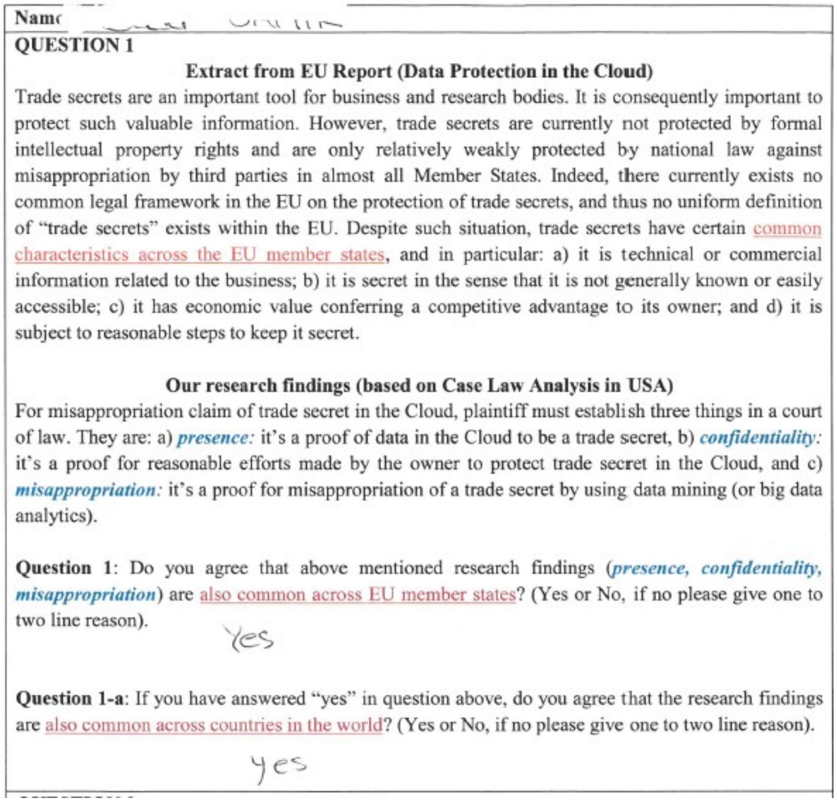
Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason).

Yes

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

Yes

Respondent 4



Name [illegible]

QUESTION 1

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

For misappropriation claim of trade secret in the Cloud, plaintiff must establish three things in a court of law. They are: a) ***presence:*** it's a proof of data in the Cloud to be a trade secret, b) ***confidentiality:*** it's a proof for reasonable efforts made by the owner to protect trade secret in the Cloud, and c) ***misappropriation:*** it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

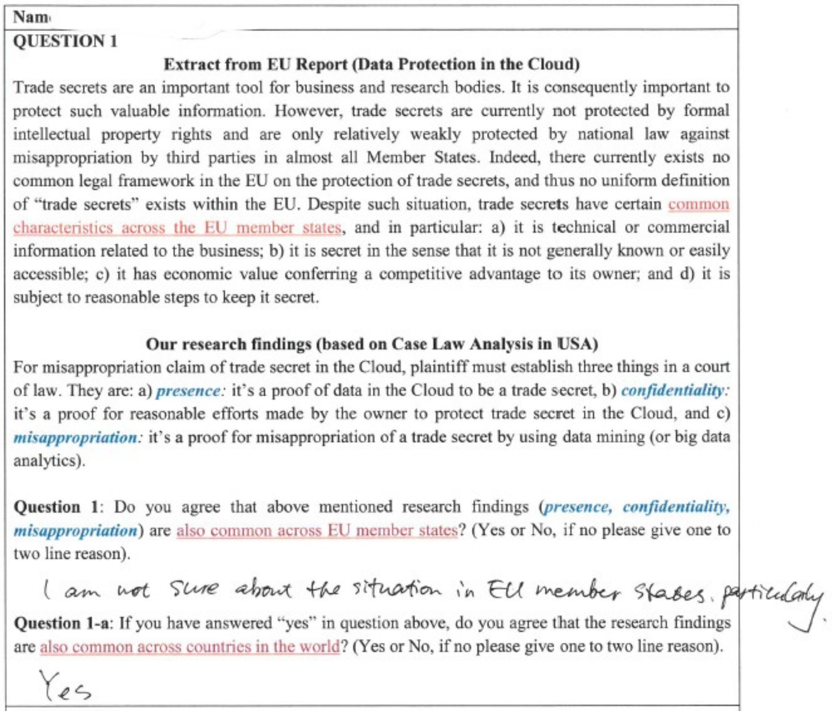
Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason).

Yes

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

yes

Respondent 5



Nam

QUESTION 1

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

For misappropriation claim of trade secret in the Cloud, plaintiff must establish three things in a court of law. They are: a) ***presence**:* it's a proof of data in the Cloud to be a trade secret, b) ***confidentiality**:* it's a proof for reasonable efforts made by the owner to protect trade secret in the Cloud, and c) ***misappropriation**:* it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

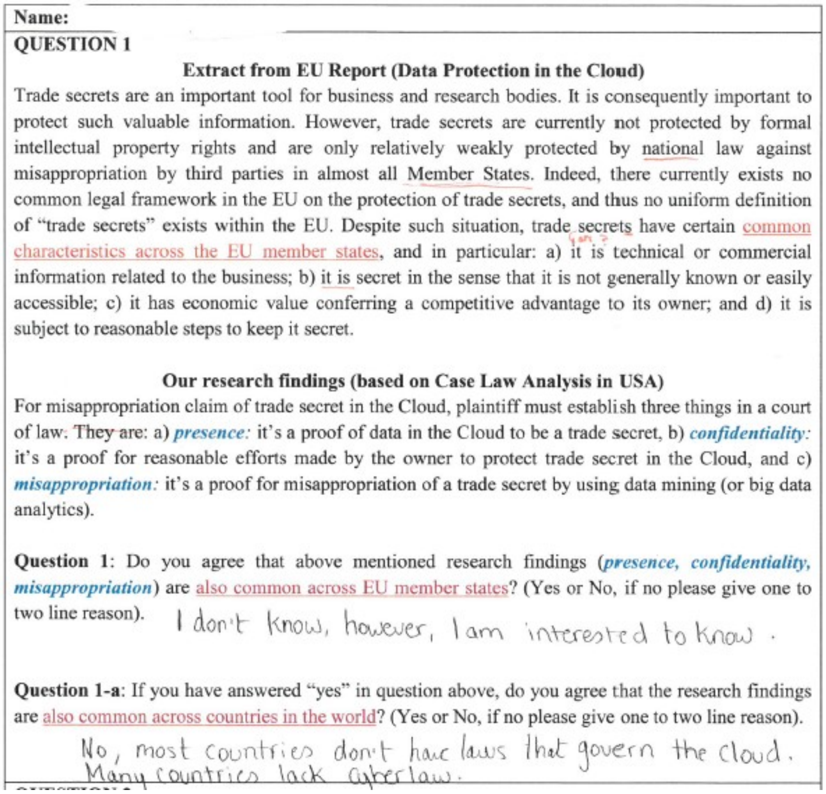
Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason).

I am not sure about the situation in EU member states. particularly.

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

Yes

Respondent 6



Name:

QUESTION 1

Extract from EU Report (Data Protection in the Cloud)

Trade secrets are an important tool for business and research bodies. It is consequently important to protect such valuable information. However, trade secrets are currently not protected by formal intellectual property rights and are only relatively weakly protected by national law against misappropriation by third parties in almost all Member States. Indeed, there currently exists no common legal framework in the EU on the protection of trade secrets, and thus no uniform definition of "trade secrets" exists within the EU. Despite such situation, trade secrets have certain common characteristics across the EU member states, and in particular: a) it is technical or commercial information related to the business; b) it is secret in the sense that it is not generally known or easily accessible; c) it has economic value conferring a competitive advantage to its owner; and d) it is subject to reasonable steps to keep it secret.

Our research findings (based on Case Law Analysis in USA)

For misappropriation claim of trade secret in the Cloud, plaintiff must establish three things in a court of law~~. They are~~: a) ***presence:*** it's a proof of data in the Cloud to be a trade secret, b) ***confidentiality:*** it's a proof for reasonable efforts made by the owner to protect trade secret in the Cloud, and c) ***misappropriation:*** it's a proof for misappropriation of a trade secret by using data mining (or big data analytics).

Question 1: Do you agree that above mentioned research findings (***presence, confidentiality, misappropriation***) are also common across EU member states? (Yes or No, if no please give one to two line reason).

I don't know, however, I am interested to know.

Question 1-a: If you have answered "yes" in question above, do you agree that the research findings are also common across countries in the world? (Yes or No, if no please give one to two line reason).

No, most countries don't have laws that govern the cloud. Many countries lack cyberlaw.

3.6 Summary

This chapter proposes a model for an online broker that embeds legal protection as preventive measure to reduce burden of proof in the court of law. The underlying concept in proposed model is built upon the notion of factor analysis from the domain of Unsupervised Machine Learning. For evaluation of proposed model, a two-stage procedure was implemented. In stage one; the proposed model showed how to assess structural significance of criteria and in stage two, a comparative analysis was performed between the proposed model and its counterpart to show how results of stage one can be used to reduce burden of proof in the court of law. A real time QoS based dataset for seven different cloud storage provider's i.e Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, SugarSync, and Zip Cloud, was used for evaluation. The simulation results showed better results of proposed model as compared to its counterparts in the field i.e. AHP and TOPSIS.

For legal validation of the research findings of table 3.1, using Delphi Sampling, questions were sent to law and ICT experts as part of a questionnaire. There were total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). There were two rounds. After two rounds the sample **(5 out of 6 respondents)** agreed that our research findings (*presence, confidentiality, misappropriation*) are common across EU member states and other countries in the world.

Chapter 4 : Generalization of Proposed Model

This chapter presents generalization of the model proposed in chapter 3. This is one of the major requirements of the second PhD degree “PhD in Informatics (Informatique)” at University of Luxembourg, Luxembourg. Furthermore, the dataset used in chapter 3 for evaluation of the proposed model was secondary data (data that was collected by someone other than the user). This chapter takes the evaluation one step further and tests the proposed model in cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). Section 4.1 presents the context (or scenario) in which the generalization of the proposed model (in chapter 3) was applied. Afterwards, section 4.2 evaluates the generalized model in emulated cloud brokerage architecture. Section 4.3 legally validates the results of preceding section by using Delphi Sampling. Finally, section 4.4 summarizes the discussion and findings in the chapter.

4.1 The Context and Generalization

Multi-criteria decision analysis (MCDA), as discussed in section 3.2, is one of the prevalent branches of operations research, aims to design mathematical and computational tools for selecting the best alternative among several choices [75]. It prescribes a methodology that deals with the most important components in the process of decision making and aims at supplying reliable information to take an unbiased decision. These components include a pre-established *goal* achievable under given constraints. Constraints are *criteria* used to rank potential *alternatives*. An unbiased ranking of alternatives is based upon selection of relevant criteria by a decision maker which strongly relates to his/her profound knowledge of the subject matter [75, 76]. Hence, the approach is termed ineffective when the decision maker has insufficient subject knowledge [77, 78]. For example, let’s assume a startup called Moogle is using cloud based brokerage architecture (online broker) to buy online storage service for data backups. The goal of online broker is to select a service provider with best QoS from the list: carbonite, dropbox, ibackup, justcloud, sos online backup, sugarsync, and zip cloud. A ranking of these service providers is generated by online broker using following QoS based criteria: availability, response time, price, speed, ease of use, technical support, and customer services. However, Moogle as per its insufficient domain knowledge for cloud based storage environment includes an additional criterion of *storage space* to the list. As a result, the ranking generated by online broker for service provider is off by a certain amount and consequently, Moogle bypasses an optimal choice for online storage service in the cloud.

Since most common MCDA methods used by online brokers fail to operate without customer interference, a self-regulated MCDA to deal with misspecification of criterion owing to insufficient knowledge of a customer is needed [79-81]. This chapter proposes self-regulated MCDA (generalization of model proposed in chapter 3), which resolves *misspecification for criterion owning to its statistical relevance that is estimated using notion of communality.* Communality belongs to broader concept of factor analysis from the field of statistics [52, 74]. Numerically, it is a measure of a relationship between a criterion and a goal [52]. Its high value indicates strong correlation between the two and hence, endorses the criterion as *relevant* with reference to a goal. In the example of Moogle, except for the additional criterion of *storage space*, all other criteria have strong correlation with QoS and hence, relevant to generate QoS based ranking of service providers.

Communality is estimated by using structural equation modeling (SEM). SEM is a statistical approach used to examine association between a latent variable and an observed variable [52, 74]. Latent variable, as mentioned in section 3.3, is a theoretical construct that is inferred from the variables that are observed during a test or survey. In the example of Moogle, QoS is a latent variable since it represents intent of a customer and is inferred from the variables (availability, response time, price, speed, ease of use, technical support, and customer service) that are observed during the test or survey.

In SEM, as mentioned in section 3.3, the most popular and frequently used methods to estimate communality are Principal Factor Analysis (PFA) and

Maximum Likelihood (ML) [52, 74]. Considering that ML estimation assumes normal distribution of observed variables and this research is dealing with observed variables without making any prior assumption, PFA was used to estimate communality. The vector notation in PFA that is used to calculate communality (ς) between n observed variables and a goal is given in equation 1. For summarized discussion on derivation of ς see following text box at the end of this section.

$$\varsigma\ = \begin{bmatrix} (u_1)^2 \\ (u_2)^2 \\ \vdots \\ (u_n)^2 \end{bmatrix} \Theta \qquad (1)$$

In the equation, eigenvector contains estimated unit-scaled loadings or weights (u_i) that are associated with each observed variable. The eigenvalue Θ is a shared variance among all the observed variables that represent the latent variable. Communality is obtained by multiplying squared value of u_i with Θ, which represents the relationship of latent variable with observed variable. The strong correlation between the two is identified by using the condition $\varsigma >\ \omega$. Where, ω is a controlled variables (or constant) same one that was discussed in section 3.3. $\omega = 0.60$ ensures that a criterion which contributes less than 60% to the goal is not selected for further processing. In the example of Moogle, *storage space* was one such example. Accordingly, equation 1 can be rewritten as:

$$\varsigma = \left(\begin{bmatrix} (\mathrm{u}_1)^2 \\ (\mathrm{u}_2)^2 \\ \vdots \\ (\mathrm{u}_\mathrm{n})^2 \end{bmatrix} \Theta \right) > \omega \qquad (2)$$

Derivation (Read Section 3.3 for detailed understanding)

In PFA, the relationship vector $\Lambda = (\lambda_1 \lambda_2 \ldots \lambda_\mathrm{n})'$ between a latent variable F and observed variable vector $\mathrm{Y} = (\mathrm{y}_1 \mathrm{y}_2 \ldots \mathrm{y}_\mathrm{n})'$ is expressed in a variance-covariance matrix notation as:

$$\mathrm{cov}(\mathrm{Y}) = \mathrm{cov}(\Lambda \mathrm{F}) + \psi$$

ψ is a vector that represent uniqueness of observed variables not shared with the latent variable. By using covariance property $\mathrm{cov}(\mathrm{AZ}) =$

$\mathrm{A\,cov(Z)\,A^T}$, $\mathrm{cov}(\Lambda\mathrm{F})$ in the right hand side of above equation can be expanded to $\Lambda\,\mathrm{cov(F)}\,\Lambda^\mathrm{T} + \psi$. Moreover, since F being an identity matrix has $\mathrm{cov(F)} = 1$, $\Lambda\,\mathrm{cov(F)}\,\Lambda^\mathrm{T}$ can be further reduced to: $\Lambda\Lambda^\mathrm{T} + \psi$ and the equation becomes:

$$\mathrm{cov(Y)} = \Lambda\Lambda^\mathrm{T} + \psi$$

If Y is not commensurate i.e. observed variables are measured in different units and scales, then standardized Y is used. After standardization, covariance becomes correlation (r) and subsequently, covariance matrix $\mathrm{cov(Y)}$ becomes a correlation matrix R.

$$\mathrm{R} = \Lambda\Lambda^\mathrm{T} + \psi$$

we can expand above equation as:

$$\begin{bmatrix} 1 & \cdots & \mathrm{r}_{1\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathrm{r}_{\mathrm{n}1} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_\mathrm{n} \end{bmatrix} [\lambda_1\ \lambda_2 \ldots \lambda_\mathrm{n}] + \begin{bmatrix} \psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_\mathrm{n} \end{bmatrix}$$

Bringing ψ to left hand side and preforming subtraction,

$$\begin{bmatrix} 1-\psi_1 & \cdots & \mathrm{r}_{1\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathrm{r}_{\mathrm{n}1} & \cdots & 1-\psi_\mathrm{n} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_\mathrm{n} \end{bmatrix} [\lambda_1\ \lambda_2 \ldots \lambda_\mathrm{n}]$$

Subtracting unique variance from the one $(1-\psi_\mathrm{i})$ will yield shared variance of an observed variable for the latent variable, which is equal to square of λ_i. Respectively, $(\lambda_\mathrm{i})^2$ can replace $1-\psi_\mathrm{i}$ and above equation will become:

$$\begin{bmatrix} (\lambda_1)^2 & \cdots & \mathrm{r}_{1\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathrm{r}_{\mathrm{n}1} & \cdots & (\lambda_\mathrm{n})^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_\mathrm{n} \end{bmatrix} [\lambda_1\ \lambda_2 \ldots \lambda_\mathrm{n}] \qquad (1)$$

Where left hand side,

$$\begin{bmatrix} (\lambda_1)^2 & \cdots & \mathrm{r}_{1\mathrm{n}} \\ \vdots & \ddots & \vdots \\ \mathrm{r}_{\mathrm{n}1} & \cdots & (\lambda_\mathrm{n})^2 \end{bmatrix} = \mathrm{R} - \psi$$

Accordingly, in a reduce form, equation 1 becomes:

$$\mathrm{R} - \psi = \Lambda\Lambda^\mathrm{T} \qquad (2)$$

$R - \psi$ is a 'reduced correlation matrix' with $(\lambda_i)^2$ on the diagonal. If $R - \psi$ is positive semi-definite matrix i.e. it satisfy $R - \psi = (R - \psi)^T$, then this implies that left hand side in equation 2 is symmetric and has a following spectral decomposition.

$$R - \psi = UDU^T \qquad (3)$$

Spectral decomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvectors to identify latent variable and corresponding eigenvalues to show strength of identified latent variable. In equation 3, U is the matrix of eigenvectors of $R - \psi$ and D is the diagonal matrix of corresponding eigenvalues $\Theta_1\ \Theta_2 \ldots \Theta_n$.

$$D = \begin{bmatrix} \Theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Theta_n \end{bmatrix}$$

The important property of a positive semi-definite matrix is that its eigenvalues are always positive or null. Hence, $\Theta_i \geq 0$ and consequently, D can be factored into $D^{1/2}D^{1/2}$ and right hand side in equation 3 becomes:

$$R - \psi = \left(UD^{\frac{1}{2}}\right)\left(D^{\frac{1}{2}}U^T\right) \qquad (4)$$

Equation 4 is in the form of equation 2 and accordingly, following can be deduced for Λ.

$$\Lambda = \left(UD^{\frac{1}{2}}\right)$$

In an expanded form, right hand side in above equation can be written as:

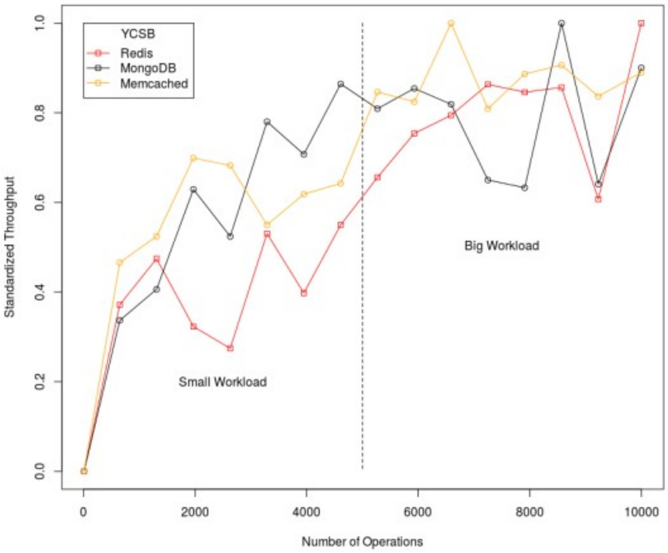
$$\Lambda = \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nn} \end{bmatrix} \times \begin{bmatrix} \sqrt{\Theta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{\Theta_n} \end{bmatrix}$$

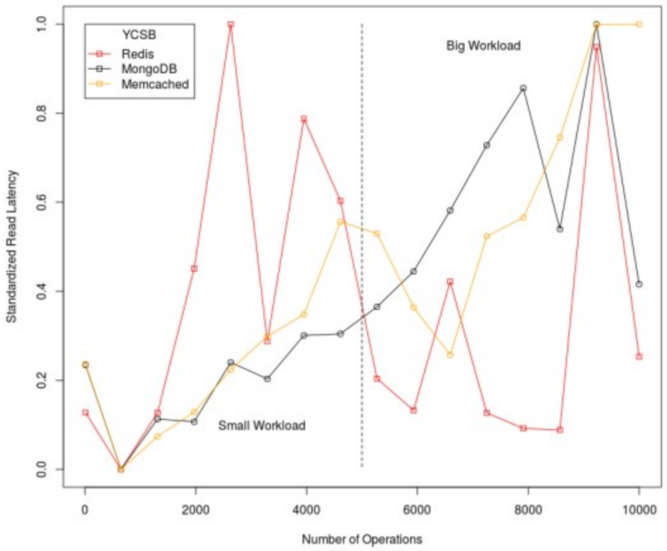
It can be observed that Λ (or $UD^{1/2}$) is $[n \times n]$ matrix, however, for single latent variable F, Λ must be $[n \times 1]$ matrix as $\Lambda = (\lambda_1 \lambda_2 \ldots \lambda_n)'$. Hence, from the right hand side of above equation we take the largest eigenvalue Θ_i and corresponding eigenvector U_i for calculation of Λ i.e., $\Lambda = U_i\sqrt{\Theta_i}$. Whereas, using Λ, communality (ς) is calculated as:

$$\varsigma = \Lambda^2 = \begin{bmatrix} (u_1)^2 \\ (u_2)^2 \\ \vdots \\ (u_n)^2 \end{bmatrix} \Theta_i$$

4.2 Technical Evaluation and Results

Same as section in section 3.4, a two-stage procedure was implemented in order to evaluate self-regulated MCDA in an online cloud environment. In stage one; relevance of criterion was assessed by using equation 2. In stage two, a comparative analysis was performed between two types of MCDA based online brokers. Only one type was equipped with self-regulated MCDA. The dataset used during these stages comprised of "*feedback from servers*" on QoS of cloud storage providers. The data was generated from cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). More specifically, a virtual machine in HPC cluster together with docker (a software container platform) was used to emulate three cloud storage providers running NoSQL databases: Redis, MongoDB, and Memcached. Each of these service providers were operating under a workload comprising of operations ranging from 0 to 10,000, records ranging from 0 to 10,000, and threads ranging from 0 to 100.


YCSB
Redis
MongoDB
Memcached
Standardized Throughput
Number of Operations
Small Workload
Big Workload
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000


YCSB
Redis
MongoDB
Memcached
Standardized Read Latency
Number of Operations
Small Workload
Big Workload
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000

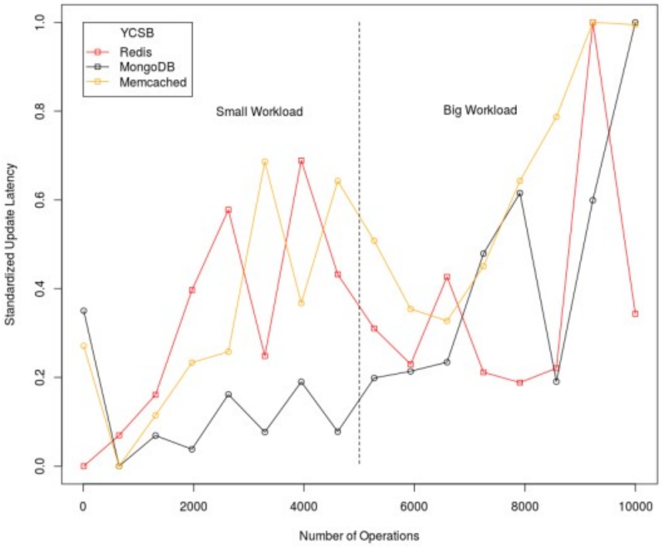


Fig. 4.1. Descriptive Statistics of Server Feedbacks

Yahoo Cloud Service Benchmark (YCSB) was deployed at the customer end i.e., second virtual machine in HPC cluster, to continuously monitor QoS of these storage providers in terms of throughput (operations per second), read latency (time to read data from database), and update latency (time to update data in database).

For eight simulation runs with small workload (number of operations < 5000) and big workload (number of operations > 5000), Figure 4.1 depicts descriptive statistics of three storage providers in terms of standardized values of throughput, read latency, and update latency. Based on these statistics, none of the storage provider can be classified “more superior” as compared to others.

The data analysis, scripting, and visualizations tools used during the two-stage procedure include: Python, R/R Studio, Arena Rockwell Input analyzer, STATA – Data Analysis and Statistical Software, IBM Statistical Analysis Software Package (SPSS), and Microsoft Excel. The scripts for setting up service providers (Redis, MongoDB, and Memcached Servers) with Docker and YCSB are given in appendix A.

4.2.1 Structural Significance of Criteria

Using equation 1 and steps presented in section 3.4.1, the communality of each criterion in the dataset was calculated, it was: 0.379 for Throughput, 0.463 for Read Latency, and 0.338 for Update Latency. Using equation 2, the significance of each criterion in the datasets was assessed. Using the opinion of substantive specialist in reference to emulated cloud environment, the value of ω was set to 0.30 (30%). Based on the condition $\varsigma > \omega$ and the communality (Throughput: 0.379, Read Latency: 0.463, and Update Latency: 0.338), none of the criteria was omitted from further processing.

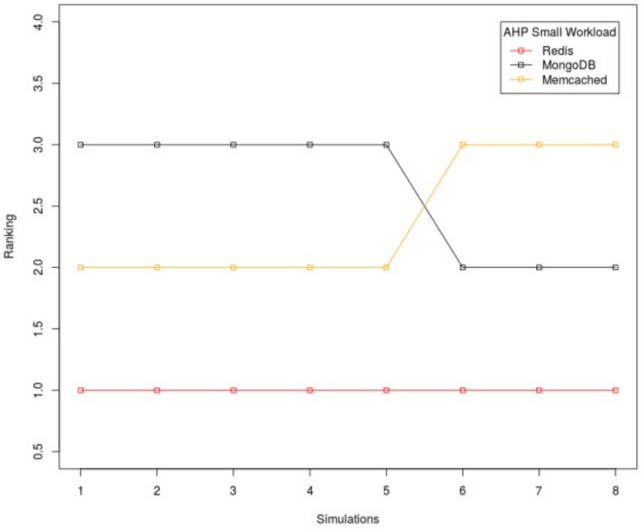
4.2.2 Comparative Analysis

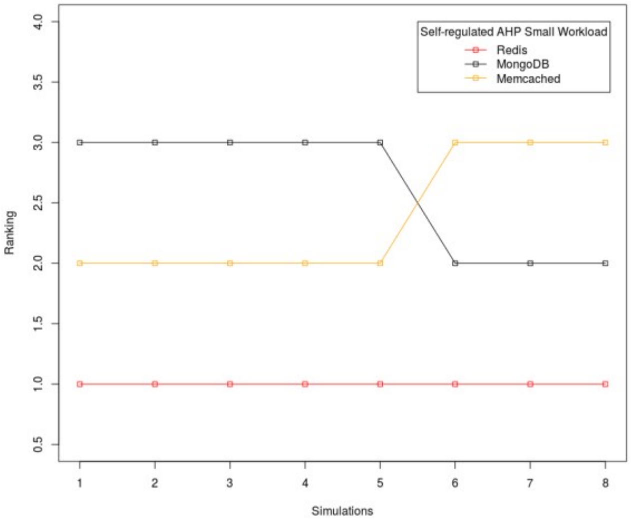
In this stage, following two comparative analyses are performed between MCDA based online broker that is using self-regulated MCDA and the one that is not.

1. It's a comparison between AHP and AHP based upon proposed model i.e. Self-regulated AHP. AHP performs series of pair-wise comparisons for three QoS based criteria using priorities provided by the customer (experts at HPC @ Uni.lu). As there was no omission of criterion based on the condition $\varsigma > \omega$, Self-regulated AHP uses the same three QoS based criteria with priorities assigned based on the communality. However, based on the fact that Self-regulated AHP in this dataset was only using "priorities assigned objectively", it was expected that it might not produce better results as compared to AHP. This is true when priorities assigned by the customer in AHP are not substantially different from priorities in Self-regulated AHP.
2. A similar setting was also applied for comparison between TOPSIS and Self-regulated TOPSIS.

Same as in section 3.4.2, the motivation for performing two pairs of comparative assessment (AHP v. Self-regulated AHP and TOPSIS v. Self-regulated TOPSIS) for each dataset was to produce results for both certain and uncertain online cloud environment. High degree of randomness was induced by using random probability distribution to simulate uncertainty in the datasets for TOPSIS v. Self-regulated TOPSIS.

Figure 4.2 presents results for comparative assessment of AHP v. Self-regulated AHP for dataset with feedback from servers. The assessment was performed for two workloads (small load and big load, see figure 4.1). For big load, the priorities assigned by the customer (experts at HPC @ Uni.lu) in AHP (Update Latency was given highest priority followed by Read Latency and Throughput) were substantially different from priorities in Self-regulated AHP (Read Latency was given highest priority followed by Throughput and Update Latency). Hence, Self-regulated AHP produced better results as compared to AHP. However, for small load, the priorities were not substantially different and therefore, the results of Self-regulated AHP were same as AHP.


AHP Small Workload
Redis
MongoDB
Memcached
Ranking
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Simulations
1
2
3
4
5
6
7
8


Self-regulated AHP Small Workload
Redis
MongoDB
Memcached
Ranking
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Simulations
1
2
3
4
5
6
7
8

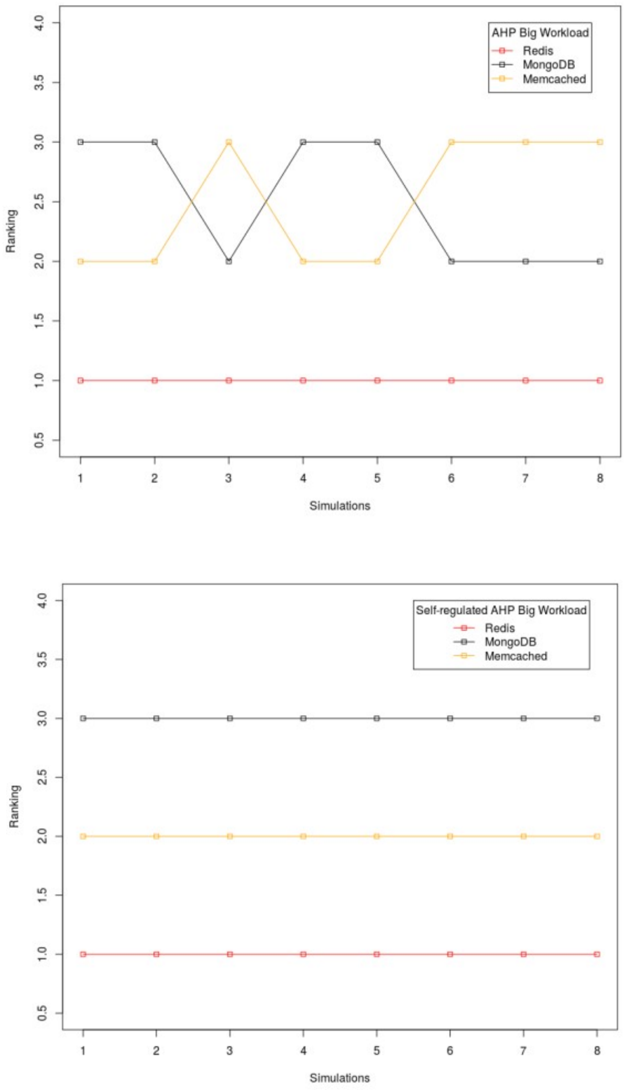
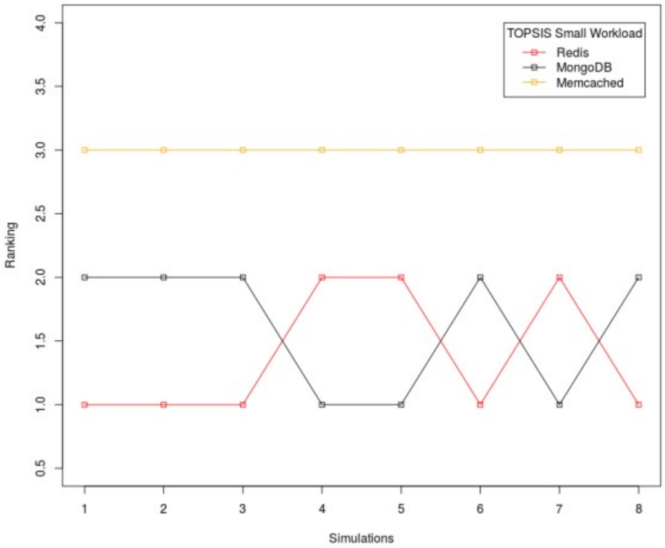
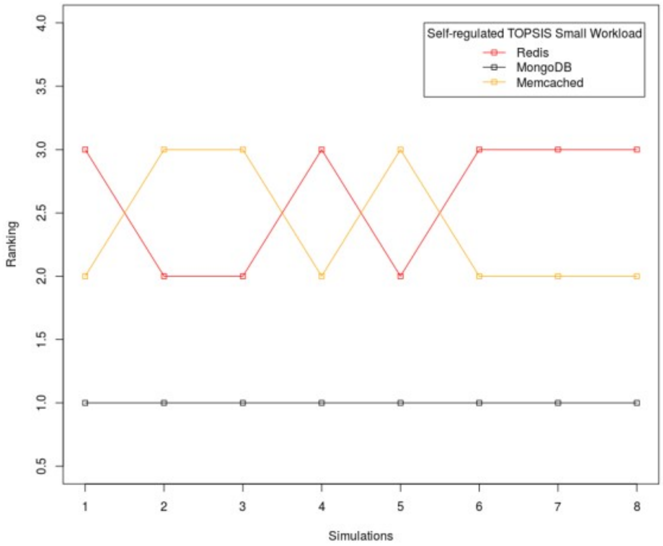


Fig. 4.2. Comparative Assessment AHP (small load) v. Self-regulated AHP (small load) and AHP (big load) v. Self-regulated AHP (big load)

Figure 4.3 presents results for comparative assessment of TOPSIS v. Self-regulated TOPSIS for dataset with feedback from servers. The assessment was performed for two workloads (small load and big load, see figure 4.1). The results are almost similar to results in figure 3.4 in section 3.4.2 i.e. it is not clear which service provider outperforms the others. These results show the same limitation identified in section 3.4.2 of proposed model and suggest a direction of future research to augment proposed model to deal with uncertainly in the cloud. However, in the stable environment i.e. when uncertainty is low, based on above observations, it can be stated that MCDA based online brokers equipped with Self-regulated AHP or Self-regulated TOPSIS will produces more explicit ranking of service providers in the cloud as compared to it its counterparts using AHP and TOPSIS.


TOPSIS Small Workload
Redis
MongoDB
Memcached
Ranking
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
1
2
3
4
5
6
7
8
Simulations


Self-regulated TOPSIS Small Workload
Redis
MongoDB
Memcached
Ranking
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
1
2
3
4
5
6
7
8
Simulations

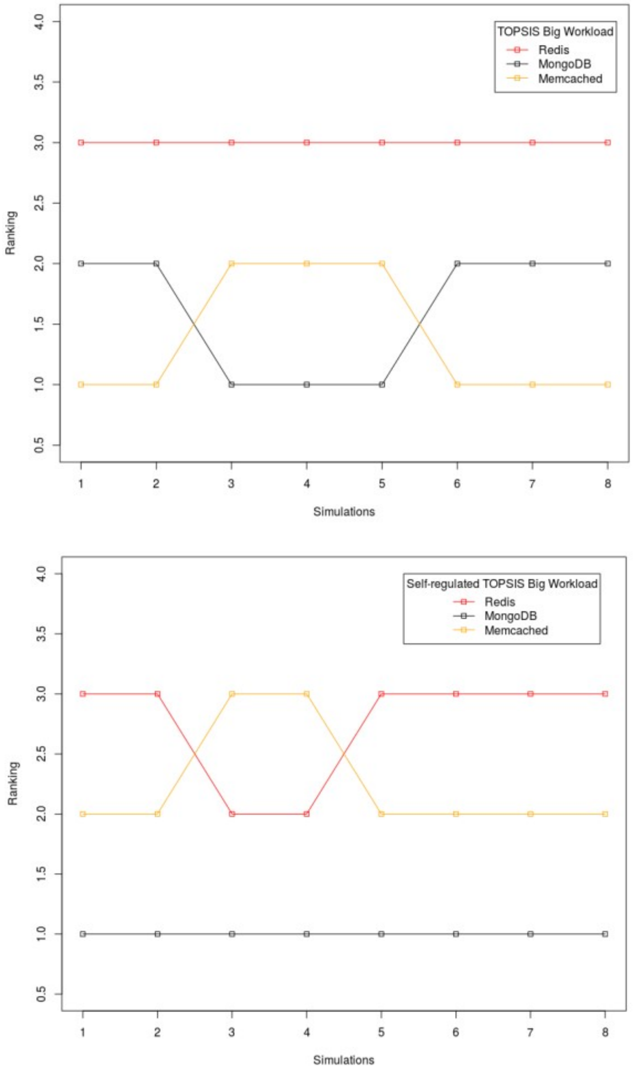


Fig. 4.3. Comparative Assessment TOPSIS (small load) v. Self-regulated TOPSIS (small load) and TOPSIS (big load) v. Self-regulated TOPSIS (big load)

4.3 Legal Validation and Results

For legal validation of research findings in this chapter, using Delphi Sampling, the following question was sent to law and ICT experts as part of a questionnaire.

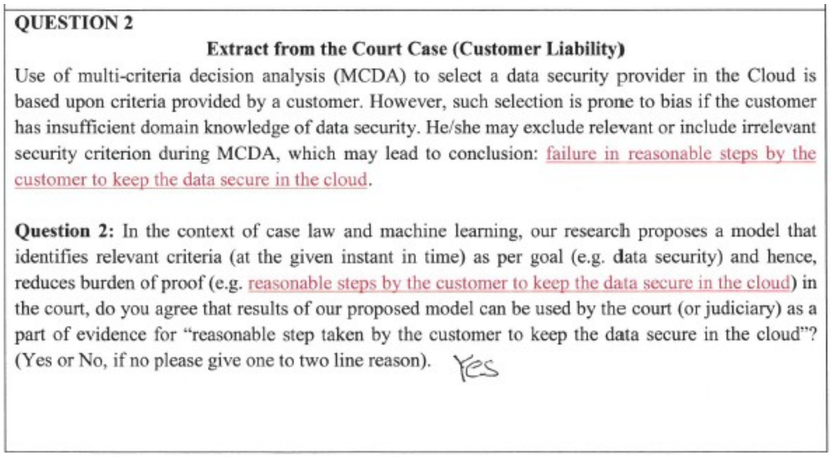
Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason).

The screenshots of responses are presented below. There were total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). There were two rounds. As per requirement of Delphi Sampling i.e. keeping anonymity in following rounds, names (of respondents) are hidden in the screen shorts. After two rounds the sample **(5 out of 6 respondents)** agreed that results of our model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud".

Respondent 1



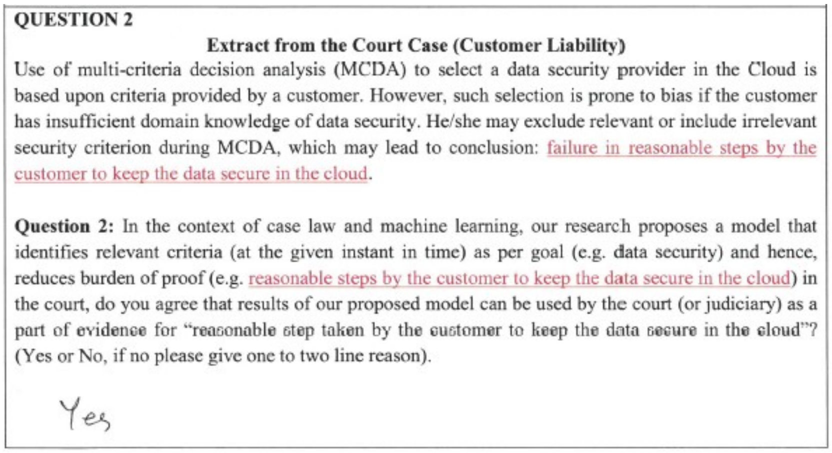
QUESTION 2

Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the Cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason). Yes

Respondent 2



QUESTION 2

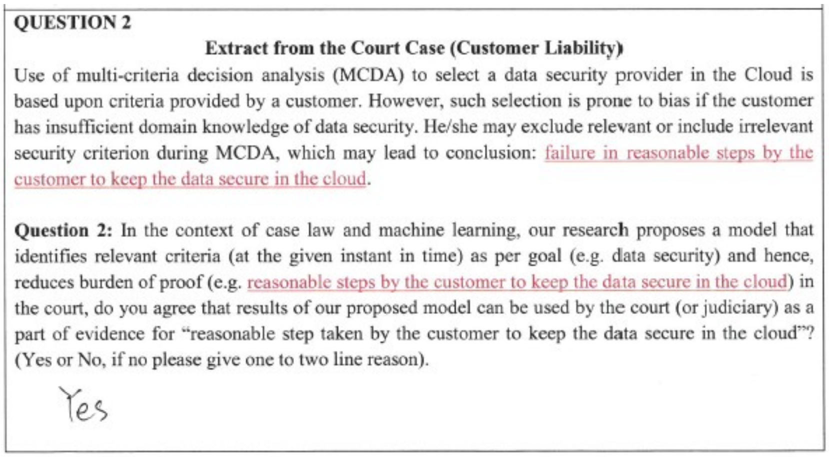
Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the Cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason).

Yes

Respondent 3



QUESTION 2

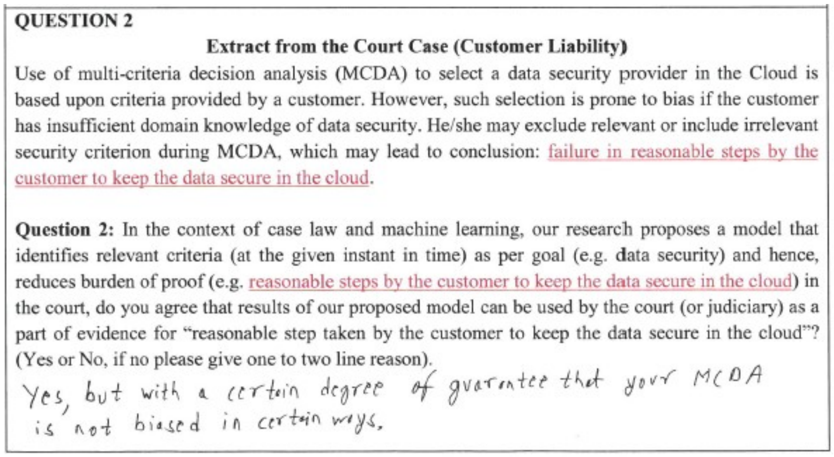
Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the Cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason).

Yes

Respondent 4



QUESTION 2

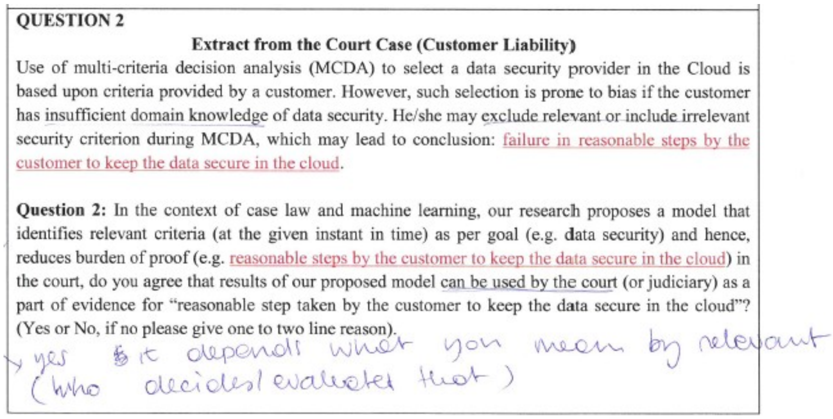
Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the Cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason).

Yes, but with a certain degree of guarantee that your MCDA
is not biased in certain ways.

Respondent 5



QUESTION 2

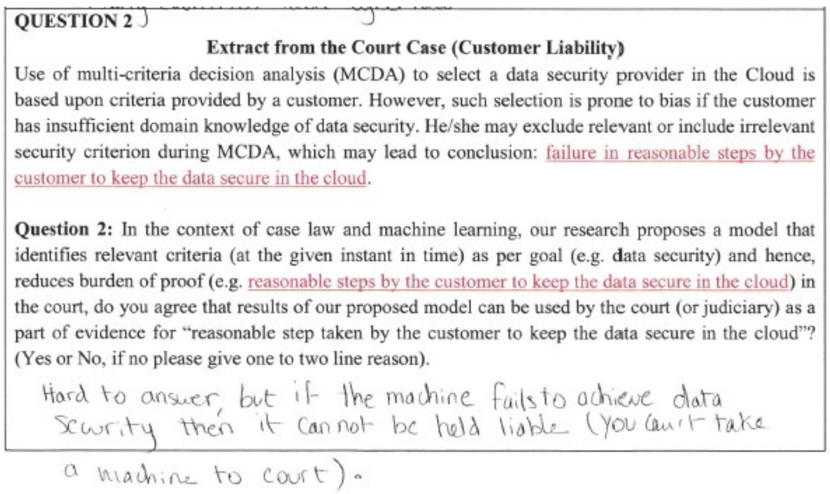
Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the Cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason).

yes it depends what you mean by relevant (who decides/evaluates that)

Respondent 6



QUESTION 2

Extract from the Court Case (Customer Liability)

Use of multi-criteria decision analysis (MCDA) to select a data security provider in the Cloud is based upon criteria provided by a customer. However, such selection is prone to bias if the customer has insufficient domain knowledge of data security. He/she may exclude relevant or include irrelevant security criterion during MCDA, which may lead to conclusion: failure in reasonable steps by the customer to keep the data secure in the cloud.

Question 2: In the context of case law and machine learning, our research proposes a model that identifies relevant criteria (at the given instant in time) as per goal (e.g. data security) and hence, reduces burden of proof (e.g. reasonable steps by the customer to keep the data secure in the cloud) in the court, do you agree that results of our proposed model can be used by the court (or judiciary) as a part of evidence for "reasonable step taken by the customer to keep the data secure in the cloud"? (Yes or No, if no please give one to two line reason).

Hard to answer, but if the machine fails to achieve data Scurity then it can not be held liable (you can't take a machine to court).

4.4 Summary

This chapter proposes self-regulated MCDA (generalization of model proposed in chapter 3). A two-stage procedure was implemented in order to evaluate self-regulated MCDA in an online cloud environment. In stage one; relevance of criterion was assessed by using the proposed model. In stage two, a comparative analysis was performed between two types of MCDA based online brokers. One type was equipped with self-regulated MCDA while the other was not. QoS based dataset was used for evaluation of self-regulated MCDA. The dataset was generated from cloud brokerage architecture that was emulated using high performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu). The simulation runs in the stable environment i.e. when uncertainty was low, showed better results of the proposed model as compared to its counterparts in the field. In particular, the results have implications for enterprises that view insufficient domain knowledge as a limiting factor for acquisition of cloud services.

For legal validation of research findings in this chapter, using Delphi Sampling, a question was sent to law and ICT experts as part of a questionnaire. There were total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). There were two rounds. After two rounds the sample **(5 out of 6 respondents)** agreed that results of our model can be used by the court (or judiciary) as a part of evidence for “reasonable step taken by the customer to keep the data secure in the cloud”.

Chapter 5 : Blockchain Evolution and Law

As discussed in section 1.1, for an R&D based enterprise that employee online broker, the answer to primary research question in chapter 3 will benefit it in negotiating a contract with service providers to minimize trade secret misappropriation in the cloud. However, if the enterprise starts using decentralized architecture in the cloud e.g. blockchains, the contract could emerge towards a smart contract, an autonomous software program running over blockchains. In this context, a well negotiated contract is not the solution to minimize trade secret misappropriation. In fact, in such case it is particularly relevant to instantiate role of judiciary over a blockchain.

This chapter describes the background information on essential concepts necessary to understand notion of blockchain; its evolution i.e. blockchain 1.0 (bitcoin), blockchain 2.0 (smart contract), and blockchain 3.0 (innovations based on smart contracts); and related key legal issues. Section 4.1 presents the concept and evolution of blockchain. Afterwards, section 4.2 presents the key legal issue related to smart contracts and section 4.3 concludes the chapter by presenting summary of the discussion and findings of the chapter.

5.1 Blockchain: Concept and Evolution

In 2009, blockchain evolution started with the developing concept of "*peer to peer economy*" on the Internet, which is known as Bitcoin [82]. The bitcoin is supplied and supported not by a central authority e.g. Bank or enterprise like PayPal, but by automated consent among networked users. Its uniqueness, however, is based on the fact that it did not require the users to trust each other [82, 83]. Through algorithmic self-policing, any malevolent effort to cheat the system is prohibited. Technically, *Bitcoin is digital cash that is transacted via the internet in a decentralized trustless system using a public ledger called the blockchain. It combines BitTorrent peer-to-peer file sharing with public key cryptography* [82].

The benefits of the blockchain are more than just *peer to peer economy;* they extend into political, environmental, medical domains etc. [17]. For example [83],

- *"To counter oppressive political systems, blockchain technology can be used to enact in a decentralized cloud functions that previously needed administration by jurisdictionally bound organizations. This is obviously useful for organizations like WikiLeaks (where national governments prevented credit card processors from accepting donations in the sensitive Edward Snowden situation) as well as organizations that are transnational in scope and neutral in political outlook, like Internet standards group ICANN and DNS services. Beyond these situations in which a public interest must transcend governmental power structures, other industry sectors and classes can be freed from skewed regulatory and licensing schemes subject to the hierarchical power structures and influence of strongly backed special interest groups on governments, enabling new disintermediated business models. Even though regulation spurred by the institutional lobby has effectively crippled consumer genome services, newer sharing economy models like Airbnb and Uber have been standing up strongly in legal attacks from incumbents".*
- *"Coordination, record keeping, and irrevocability of transactions using blockchain technology are features that could be as fundamental for forward progress in society as the Magna Carta or the Rosetta Stone. In this case, the blockchain can serve as the public records repository for whole societies, including the registry of all documents, events, identities, and assets. In this system, all property could become smart property; this is the notion of encoding every asset to the blockchain with a unique identifier such that the asset can be tracked, controlled, and exchanged (bought or sold) on the blockchain. This means that all manner of tangible assets (houses, cars) and digital assets could be registered and transacted on the blockchain. As an example, we can see the worldchanging potential of the blockchain in its use for registering and protecting intellectual property (IP). The emerging digital art industry offers services for privately registering the exact*

> *contents of any digital asset (any file, image, health record, software, etc.) to the blockchain. The blockchain could replace or supplement all existing IP management systems. How it works is that a standard algorithm is run over a file (any file) to compress it into a short 64-character code (called a hash) that is unique to that document. No matter how large the file (e.g., a 9-GB genome file), it is compressed into a 64-character secure hash that cannot be computed backward. The hash is then included in a blockchain transaction, which adds the timestamp¬the proof of that digital asset exiting at that moment. The hash can be recalculated from the underlying file (stored privately on the owner's computer, not on the blockchain), confirming that the original contents have not changed. Standardized mechanisms such as contract law have been revolutionary steps forward for society, and blockchain IP (digital art) could be exactly one of these inflection points for the smoother coordination of large-scale societies, as more and more economic activity is driven by the creation of ideas".*

Above mentioned benefits of the blockchain can be categorize into three categories: Blockchain 1.0, 2.0, and 3.0 [83]. Following subsections briefly discuss each of the categories.

5.1.1 Blockchain 1.0 (Bitcoins)

Bitcoin is a digital currency. It was created in 2009 by an anonymous entity using the name *Satoshi Nakamoto* [82]. Payments using the bitcoins are recorded in a public ledger that is stored on computers connected to bitcoin network. The ledger can be viewed at any time on the internet. Bitcoin is the first and largest decentralized cryptocurrency whereas, other digital currencies include: Altcoin, Litecoin and Dogecoin [84]. Users can send and receive Bitcoins electronically for an optional (or very small) transaction fee using wallet (a software on a personal computer, mobile, or web application). In response to these transactions, new bitcoin are created as a reward for computational processing (known as mining), which is used to verify and record bitcoin transactions into the public ledger [82].

5.1.2 Blockchain 2.0 (Smart Contracts)

In the blockchain, smart contracts go beyond simple transactions of bitcoins, and have more extensive instructions (processing) embedded into them [85]. Formally, a smart contract is a method of using blockchain (or bitcoin transactions) to form agreements between agents.

In general, a contract (discussed in section 2.2.2), it is a promise between two or more agents to do (or not do) work in exchange for something else [86]. Each agent must trust the other agent to fulfill its side of the commitment. Smart contract feature the same kind of settlement to act or not act, but it eliminate the requirement of one agent to trust the other agent(s) [85]. This is because a smart contract is a software code that is executed over a blockchain without any discretion. In fact, two elements of the smart contracts that make them distinctive are: self-enforceability and decentralization [85]. Self-enforceability means that after it is launched, the agents engaged in the smart contract need not be in further contact. Decentralized means that smart contract do not subsist on a single centralized server; they are distributed and self-executing across the blockchain network [17]. The classic illustration of smart contracts in daily life is a vending machine. Unlike a person, the vending machine behaves algorithmically; the same instruction set will be executed every time in every case [85].

An example of a basic smart contract, with more extensive instructions as compared to bitcoins, is an *inheritance gift that becomes available on eighteenth birthday* [17]. A transaction can be created that sits on the blockchain and goes uninitiated until following two conditions are triggered.

1. The program sets the date (18th birthday) on which to initiate the transaction, which includes checking if the transaction has already been executed.
2. The program scans an online death registry database to certify that the entity of inheritance (parent or grandparent) has died. When the smart contract confirms the death, it can automatically transfer the inheritance (e.g. funds).

5.1.3 Blockchain 3.0 (Innovations based on Smart Contracts)

Except for the fact that blockchain is reinventing almost all the categories of financial services or transactions, it might also offer similar reconfiguration possibilities to all industries, and even more broadly, to nearly all areas of human endeavors [87]. For example, Northern Trust and IBM uses smart contract to help transform private equity administration [88].

- ***Said by Peter Cherecwich, president of Corporate & Institutional Services at Northern Trust:*** *"Current legal and administrative processes that support private equity are time consuming and expensive. A lack of transparency and efficient market practices leads to lengthy, duplicative and fragmented investment and administration processes. Northern Trust's solution is designed to deliver a significantly enhanced and efficient approach to private equity administration".*
- ***Said by Bridget van Kralingen, Senior Vice President, IBM Industry Platforms:*** *"Smart contract is an ideal technology to bring innovation to the private equity market, allowing Northern Trust to improve traditional business processes at each stage to deliver greater transparency and efficiency.*
- ***Said by Justin Chapman, global head of market advocacy and research at Northern Trust:*** *"Northern Trust anticipates substantial opportunities to bring improvements to the private equity market by using smart contracts. This is an important first step to connecting participants much more effectively, including investors, managers, administrators, regulators, advisors and auditors."*

Also other projects like ADEPT by IBM, Slock.it, Trans Active Grid, and Filament [89]; are successfully using smart contracts as underlying technology for bringing innovations in to the market. However, like a traditional contracts (discussed in section 2.2.2), smart contracts have also given rise to legal challenges in the domain of contract law, which could damage the reputation of conceived innovations. Breach of contract is one of such challenges, which is discussed in next section.

5.2 Rule of law and Blockchain 3.0

As mentioned in section 5.1.2, smart contracts are self-enforceable i.e. once a smart contract is concluded, its further execution is neither dependent on intend of contractual parties or third party nor does it require any additional approvals or actions from their side [17]. Thus, any malicious intent of the party i.e. breach of contract, and role of third party addressing the malicious intent i.e. judiciary, becomes irrelevant during the execution of a smart contract [18].

In addition to dealing with breaches, contract law also encompasses deviations in pre-defined outcomes. [19]. Even though breach of contract and role of judiciary become irrelevant during the execution of a smart contract, what if an output of a smart contract is considered as a breach by court of law? For example, a court may acknowledge deviation in output of a contract as a breach, if average uptime of a web service is 90% instead of agreed 95%. In such chase, as discussed in section 1.1, an automating role of judiciary over a blockchain becomes necessary. However, current projects mentioned in preceding section (Northern Trust and IBM, ADEPT by IBM, Slock.it, Trans Active Grid, and Filament) have overlooked the need to instantiate such role [89]. One of the major reasons for such gap is initial level of multi-disciplinary research when it comes to provisioning legal protection over a blockchain [12].

5.3 Summary

This chapter discusses essential concepts necessary to understand the part of PhD research that is addressing the secondary research question identified in section 1.1. In this regards, the chapter majorly discusses smart contracts. These contracts over the blockchain, go beyond simple transactions of bitcoins (or peer to peer economy), and have more extensive instructions (processing) embedded into them. They are reinventing almost all the categories of industries by conceiving innovations running over the blockchain. However, like a traditional contracts, smart contracts have also given rise to legal challenges in the domain of contract law, which could damage the reputation of conceived innovations. Breach of contract is one of such challenges and one of the ways to deal with it is by an automating role of judiciary over a blockchain.

Chapter 6 : Related Work and Proposed Model 2.0

This chapter successfully addresses the secondary research question: *what happens when the outcome of a smart contract deviates from the outcome that the law demands?* The answer to this research question will eventually benefit blockchain driven R&D based enterprises to control and stop breach of contract that could potentially lead to trade secret misappropriation. We first present in section 6.1 current models that are successfully using smart contracts as underlying technology for bringing innovations in to the market. Afterwards, section 6.2 presents the proposed model 2.0 (an extension of self-regulated MCDA presented in chapter 4) to automatically issue a court injunction when output of a smart contract breaches the contract, section 6.3 presents evaluation of the proposed model in a simulated cloud environment (same one that was presented in section 4.2), section 6.4 legally validates the results of preceding section by using Delphi Sampling; and finally, section 6.6 summaries the discussion and findings of the chapter.

6.1 Current Models

Following projects are successfully using smart contracts as underlying technology for bringing innovations in to the market. The following text is the extract from official website of the projects.

- *"Northern Trust and IBM: Northern Trust in collaboration with IBM and other key stakeholders has launched the first commercial deployment of blockchain technology for the private equity market. Northern Trust is a leading provider of wealth management, asset servicing, asset management and banking to corporations, institutions, affluent families and individuals. For more than 125 years, Northern Trust has earned distinction as an industry leader for exceptional service, financial expertise, integrity and innovation. IBM is rapidly actively working with companies to make blockchain ready for business. Financial services, supply chains, IoT, risk management, digital rights management and healthcare are some of the areas that are poised for dra-*

matic change using blockchain networks".

- *"ADEPT by IBM: Architecture designed for a dynamic democracy of objects connected to a universal digital ledger, which provides users with secure identification and authentication".*
- *"Slock.it: Architecture designed to address security, identity, coordination and privacy across millions of devices by making them autonomous. It gives connected objects an identity, the ability to receive payments, enter into complex agreements and transact without intermediary, leading to cost savings".*
- *"Trans Active Grid: Architecture designed to allows individuals to produce and exchange their energy locally via a nanogrid , which reduces transportation costs, distribution and energy losses. Specifically, the platform uses blockchain technology and protocols to store consumption / transaction data and optimize energy sharing, even on a very small scale like that of the Brooklyn community".*
- *"Filament: Architecture designed enables devices to hold unique identities on a public ledger and to discover, communicate and interact with each other in an autonomous and distributed manner".*

As mentioned in section 5.2, above mentioned projects have overlooked the need to instantiate role of judiciary over a blockchain and one of the major reasons for such gap is initial level of multi-disciplinary research when it comes to provisioning legal protection over a blockchain. Following section propose solution for such gap.

6.2 Proposed Model 2.0

This part of research proposes an unsupervised machine learning algorithm called as Probability based Factor Model (PFM) to automatically issue a preliminary injunction (or temporary restraining order by court of law) when output of a smart contract breaches the contract. The underlying concept in PFM is built upon Self-Regulated MCDA proposed in chapter 4 and stochastic modeling from the discipline of Data Science [51]. Using past data, it performs two-phase validation process to issue a court injunction. Initially, it assesses significance of a breach to ensure that the breach has a potential to create a substantial damage. Afterwards, if the significance is high, it assesses the probability of the breach. In case the probability is also high i.e. breach was frequently occurring in the past and there is certainty for it to occur in the future, PFM invokes a transaction and executes a function in a smart contract that results in the issue of court injunction. Figure 6.1 presents an example of a smart contract for Quality of Service (QoS) and a context when the contract is implemented with PFM.

Smart Contract for QoS	**PFM based Smart Contract for QoS**
Condition If latency of a cloud service goes beyond a pre-defined threshold or throughput falls below pre-defined threshold, the client machine sends a maintenance request. Transaction For sending the maintenance request, a transaction is sent to the *request_service_function* of the *Service_Smart_Contract* between the client machine and the service provider machine.	Condition (or Breach) If latency of a cloud service goes beyond a pre-defined threshold or throughput falls below pre-defined threshold, PFM at the client machine applies following logical operations to send a injunction request. φ is a high significance of the breach ϑ is a high probability of the breach $\mathbf{INJ}$ is a court injunction $(\neg\varphi \vee (\varphi \wedge \neg\vartheta) \rightarrow \neg\mathbf{INJ}) \wedge (\varphi \wedge \vartheta \rightarrow \mathbf{INJ})$ Transaction For sending the injunction request, a transaction is sent to the *request_service_function* of the *Breach_Service_Smart_Contract* between the client machine, the service provider, and the court of law.

Fig. 6.1 PFM enabled Smart Contract

6.2.1 Assessing Significance of Breach

To assess significance of breach, PFM uses notion of communality that was presented in section 4.2.1.

6.2.2 Assessing Probability of Breach

To assess probability of breach $\mathrm{P}(x)$, PFM uses notion of stochastic modeling. A stochastic model predicts a random event weighted by its probability [90]. PFM, based on the distribution modeling of the previous breaches $(x_{t-1}, x_{t-2}, .., x_{t-n})$, suggests a stochastic model with minimum "square error" to find $\mathrm{P}(x)$. In distribution modeling, square error as criteria with the minimum value indicates best possible approximation (stochastic model) for the data. However, the best possible approximation also requires verification in terms of accuracy i.e. how precisely a stochastic model can represent the data.

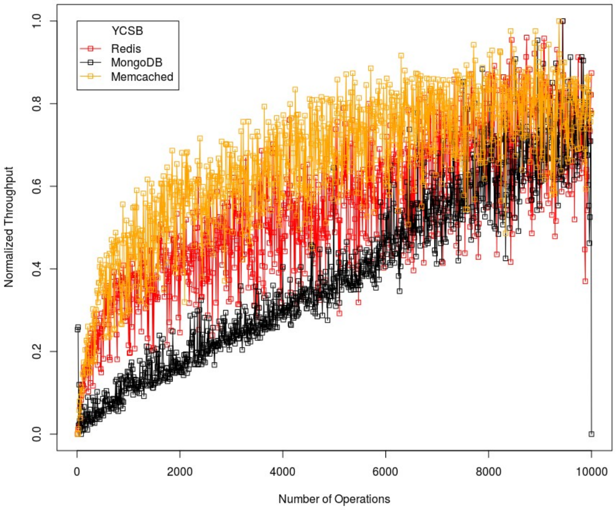
For example, during the distribution analysis, if PFM observes previous beaches are lognormal increasing with minimum square error, then the stochastic model in equation 1 will be used by PFM to calculate probability of breach $\mathrm{P(x)}$.

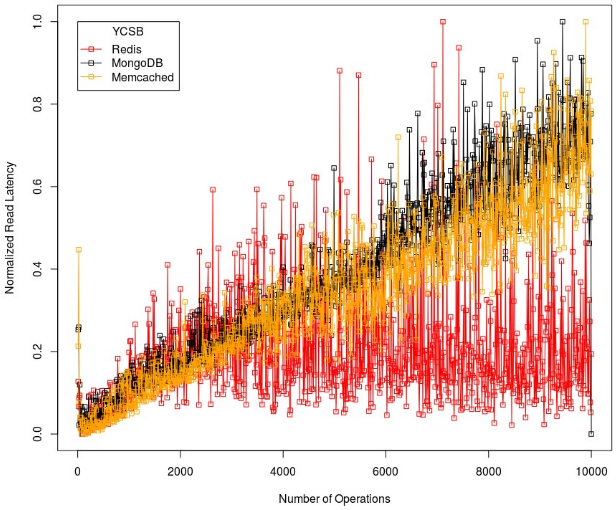
$$\mathrm{P}(x) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} e^{-(ln(x)-\mu)^2/(2\sigma^2)} & if\ (x_{\mathrm{t}-1}, \ldots, x_{t-n}) \sim LOGN(\mu, \sigma) \end{cases} \quad (1)$$

To verify the accuracy of above model, PFM performs a Paired Sample T-Test. In the test, it determines whether the mean difference between two samples i.e., previous breaches and random data generated using $\mathrm{LOGN}(\mu, \sigma)$ in equation 1, is zero or not. For later case i.e. $\neq 0$, PFM dismisses the use of stochastic model in equation 1.

6.3 Technical Evaluation and Results

For evaluation of PFM, this part of research uses the same emulated environment (of three cloud storage providers running NoSQL databases: Redis, MongoDB, and Memcached) presented in section 4.2. Figure 6.2 presents YCSB monitoring of service providers in terms of unit-scaled throughput, read latency, and update latency (see section 4.2 for details on YCSB).


YCSB
Redis
MongoDB
Memcached
Normalized Throughput
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000
Number of Operations


YCSB
Redis
MongoDB
Memcached
Normalized Read Latency
0.0
0.2
0.4
0.6
0.8
1.0
0
2000
4000
6000
8000
10000
Number of Operations

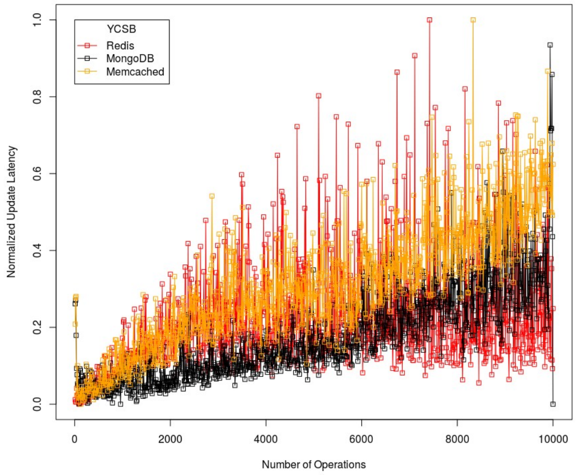


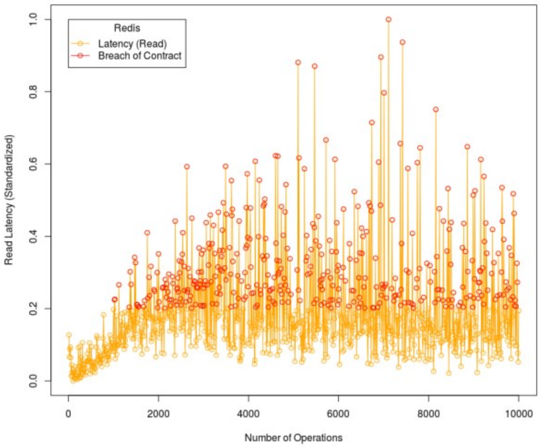
Fig. 6.2 YCSB (V 0.12.0) Monitoring of Redis, MongoDB, and Memcached

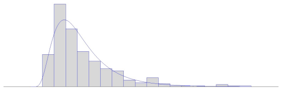
The YCSB data of all three service providers was used by PFM to calculate communality for throughput (0.38), read latency (0.46), and update latency (0.33). It can be observed that read latency has highest value and consequently, the strongest relationship with QoS. Therefore, the related breach i.e. read latency > threshold, is significant and most likely to create substantial damage.

For each service provider, (a) the threshold was set to average read latency, which was calculated from its YCSB data, (b) based on the condition i.e. read latency > average read latency, previous breaches $(x_{t-1}, x_{t-2}, \ldots, x_{t-n})$ were identified, (c) distribution modeling of previous breaches was performed using PFM, (d) afterwards, stochastic model with minimum square error was identified, and further verified for accuracy using Paired Sample T-Test.

The stochastic models for read latency of Redis and Memcached successfully passed the T-Test. However, for MongoDB (as it failed the prior T-Test) the procedure in preceding paragraph was repeated for throughput (with second highest communality value of 0.38) and stochastic model identified successfully passed the T-Test.

Table 6.1 Implementation and Results of PFM – Redis Server





Distribution**: Lognormal**

Stochastic Model: 0.12 + LOGN(0.204, 0.117)
where,
LOGN(LogMean μ, LogStd σ)
LogMean μ = 0.204, LogStd σ = 0.117, Offset = 0.12

Square Error: 0.007417 and **p-value (t-test)**: 0.5449 (>0.05)

Equation:

$$P(\mathrm{x}) = \frac{1}{\sigma \mathrm{x}\sqrt{2\pi}} \mathrm{e}^{-(\ln(\mathrm{x})-\mu)^2/(2\sigma^2)}$$

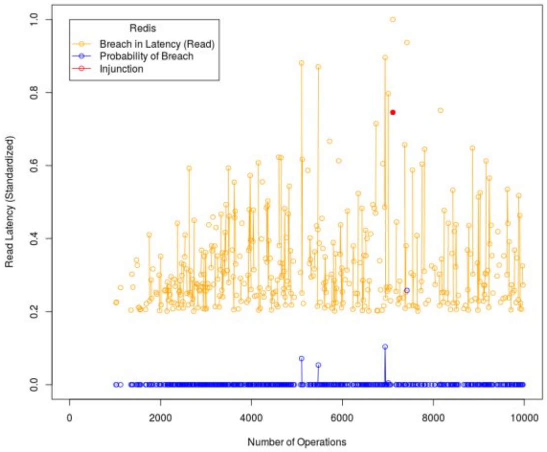
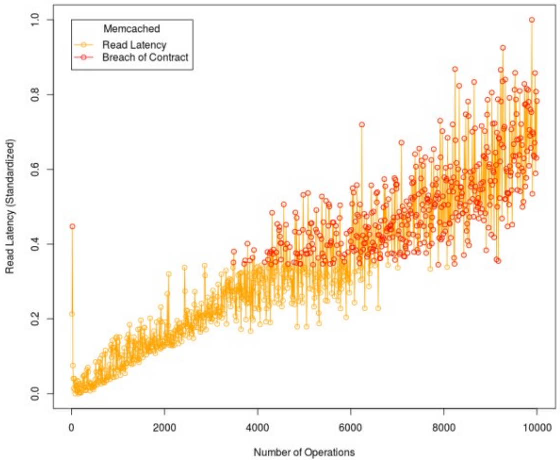
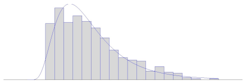


Table 6.2 Implementation and Results of PFM – Memecached Server



 **Distribution:** Lognormal	**Stochastic Model**: 0.27 + LOGN(0.245, 0.137) where, LOGN(LogMean μ, LogStd σ) LogMean μ = 0.245, LogStd σ = 0.137, Offset = 0.27 **Square Error**: 0.003444 **and p-value (t-test)**: 0.8258 (>0.05) **Equation**: $P(\mathrm{x}) = \frac{1}{\sigma\mathrm{x}\sqrt{2\pi}} \mathrm{e}^{-(\ln(\mathrm{x})-\mu)^2/(2\sigma^2)}$

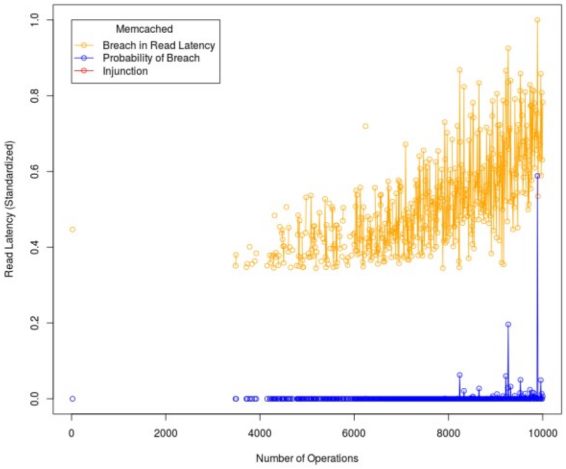
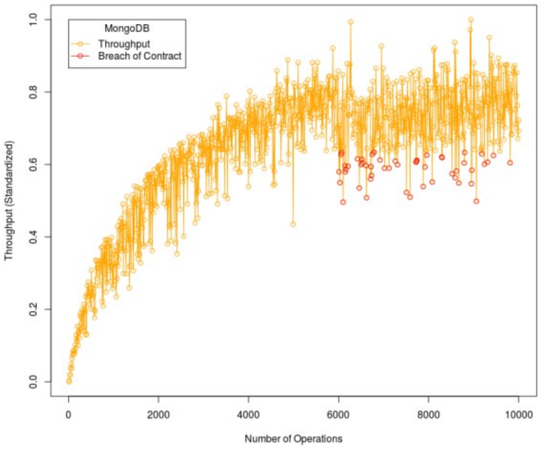
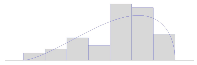


Table 6.3 Implementation and Results of PFM – MongoDB





Distribution: Beta

Stochastic Model: 0.48 + 0.17 * BETA(2.49, 1.48)
where,
BETA(Beta β, Alpha α) or BETA(Alpha1, Alpha2)
β (Alpha1) = 2.49, α (Alpha2) = 1.48, Offset = 0.48 + (0.17 * BETA)
Square Error: 0.018634 and **p-value (t-test)**: 0.4788 (>0.05)
Equation:

$$P(\mathrm{x}) = \frac{\mathrm{x}^{\beta-1}(1-\mathrm{x})^{\alpha-1}}{\int_0^1 \mathrm{t}^{\beta-1}(1-\mathrm{t})^{\alpha-1}\mathrm{dt}}$$

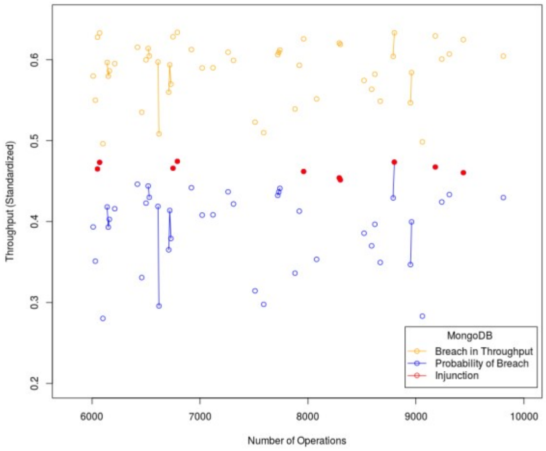


Table 6.1, 6.2, and 6.3 presents the implementation and results of PFM. Row 1 of table 6.1 and 6.2 shows previous breaches based on two conditions: "read

latency > average read latency" for Redis and Memcached. Row 1 of table 6.3 shows previous breaches based on the condition: "throughput < average throughput" for MongoDB. Row 2 of each table shows distribution modeling results. It can be observed that for Redis and Memecached, previous breaches in read latency are lognormal increasing and for MongoDB, previous breaches in throughput are beta increasing.

Row 3 of each table presents stochastic models for each service provider with minimum square error (Redis: 0.007417, Memcashed: 0.003444, and MongoDB: 0.018634). Moreover, as p-values of Paired Sample T-Test (Redis: 0.5449, Memcashed: 0.8258, and MongoDB: 0.4788) are greater than 0.05, the null hypothesis (the two samples are same) is accepted as compared to alternate hypothesis (the two samples are different). Hence, the stochastic models for Redis (read latency) i.e., 0.12 + LOGN(0.204, 0.117), Memcached (read latency) i.e., 0.27 + LOGN(0.245, 0.137), and MongoDB (throughput) i.e-0.48 + 0.17 * BETA(2.49, 1.48), can be used by PFM to find probability of breach $\mathrm{P}(x)$.

Last row in each table shows lognormal $\mathrm{P}(x)$ for Redis and Memcached and beta $\mathrm{P}(x)$ for MongoDB. It also shows issued injunctions. Based on the opinion of substantive specialist in the field and communality, for Redis and Memcached the injunction was issued based on the condition: $\mathrm{P}(x) > 0.70$, whereas, for MongoDB the condition was: $\mathrm{P}(x) > 0.45$. It can be observed that court injunction(s) was issued only for Redis and MongoDB Servers. Technically, this difference could be attributed to the fact that Memcached is simply used for caching and therefore, it is less prone to breach of contract. Whereas, Redis and MongoBD as databases and message brokers are performing more complex operations and are more likely to cause a breach. Overall, these results shows that this part of the research has successfully addressed the secondary research question (*what happen when the outcome of a smart contract deviates from the outcome that the law demands?*) and have implemented notion of confidentiality by design over the blockchain.

6.4 Legal Validation and Results

For legal validation of research findings in this chapter, using Delphi Sampling, the following question was sent to law and ICT experts as a part of questionnaire.

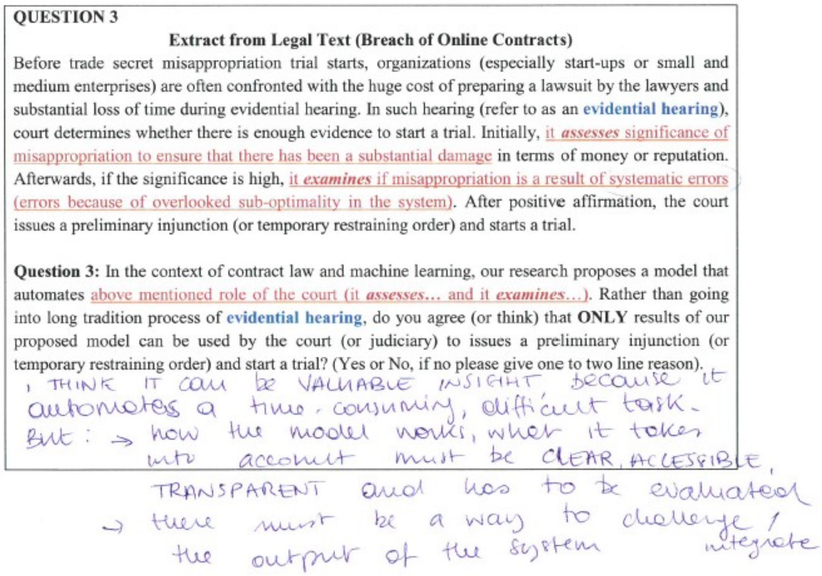
Extract from Legal Text (Breach of Online Contracts)

Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines***…). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issues a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

The screenshots of responses are presented below. There were a total of six respondents (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). There were two rounds. As per requirement of Delphi Sampling i.e. keeping anonymity in following rounds, names (of respondents) are hidden in the screen shorts. After two rounds the sample **(4 out of 6 respondents)** disagreed for **ONLY** using the results of our proposed model by the court (or judiciary) to issue a preliminary injunction (or temporary restraining order) and starts a trial.

Respondent 1



QUESTION 3

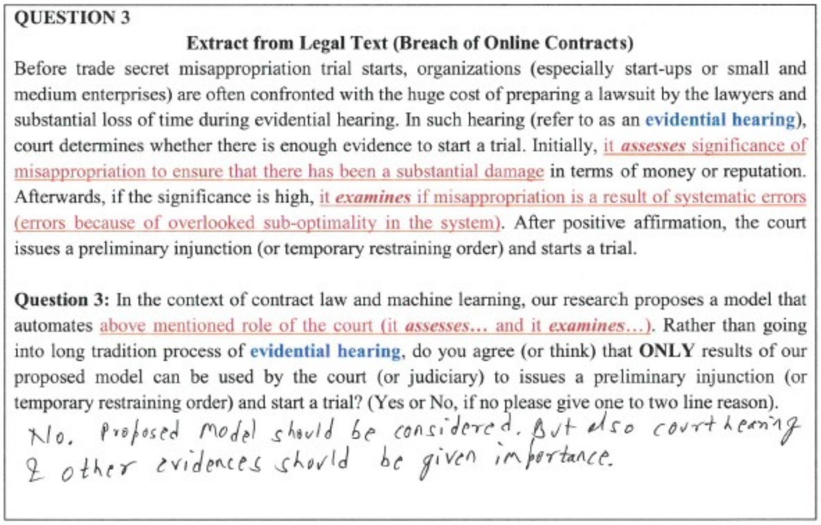
Extract from Legal Text (Breach of Online Contracts)

Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines…***). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issues a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

I THINK IT CAN be VALUABLE INSIGHT because it automates a time-consuming, difficult task.
But: → how the model works, what it takes into account must be CLEAR, ACCESSIBLE, TRANSPARENT and has to be evaluated
→ there must be a way to challenge / integrate the output of the system

Respondent 2



QUESTION 3

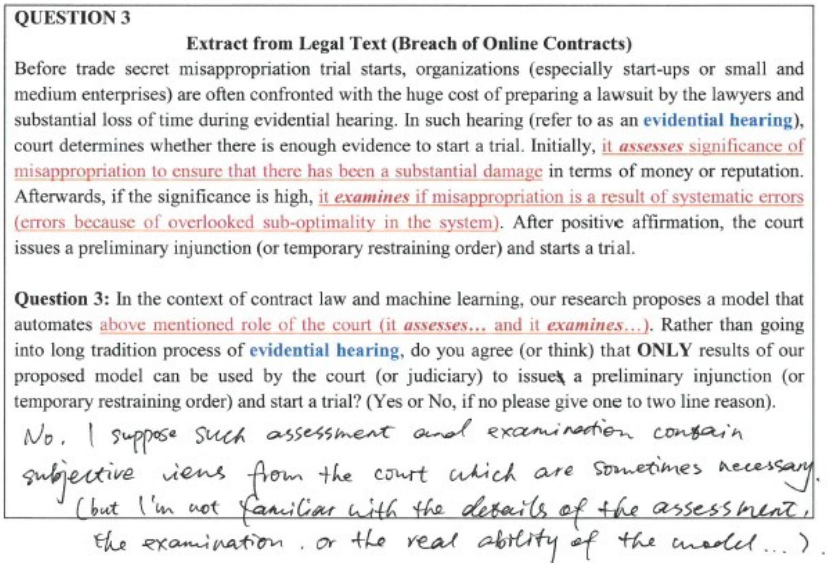
Extract from Legal Text (Breach of Online Contracts)

Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines***…). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issues a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

No. Proposed Model should be considered. But also court hearing & other evidences should be given importance.

Respondent 3



QUESTION 3

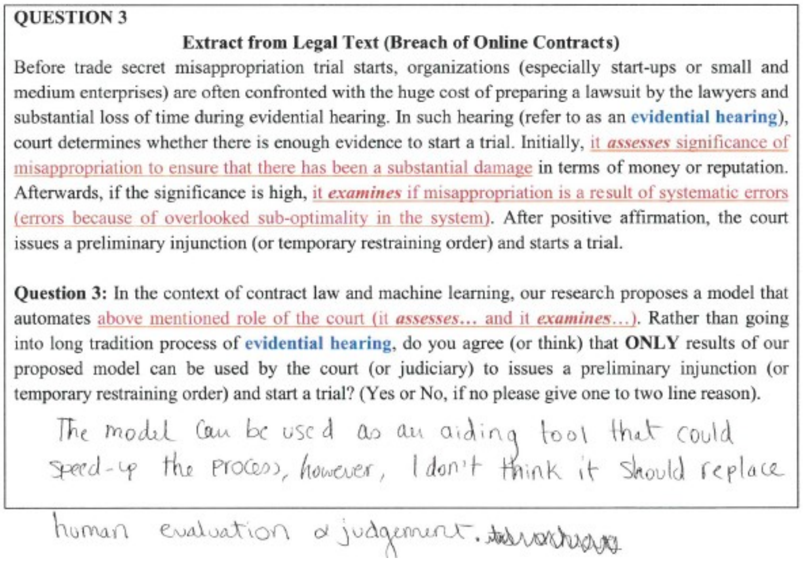
Extract from Legal Text (Breach of Online Contracts)

Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines***…). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issue~~s~~ a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

No. I suppose such assessment and examination contain subjective views from the court which are sometimes necessary. (but I'm not familiar with the details of the assessment, the examination, or the real ability of the model…).

Respondent 4



QUESTION 3

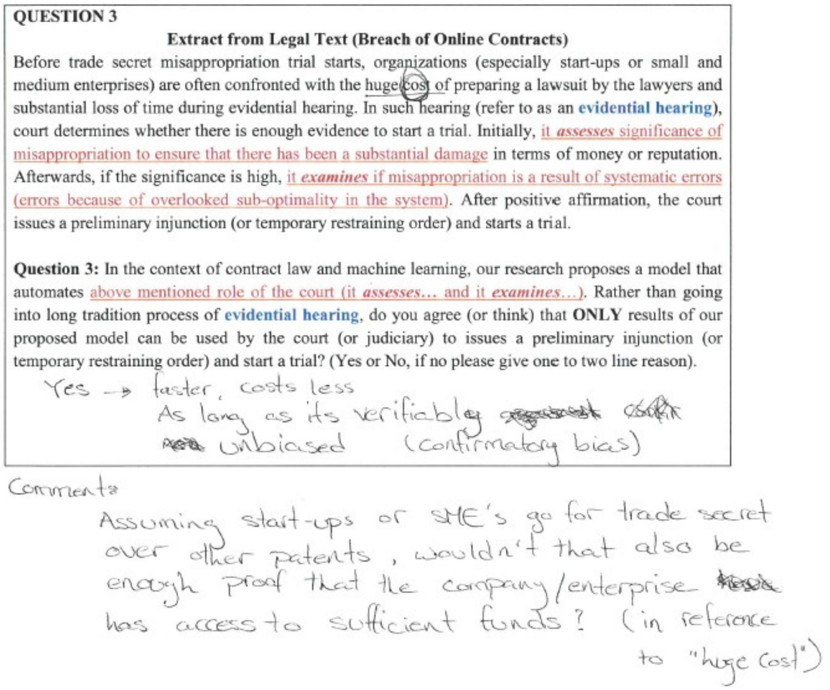
Extract from Legal Text (Breach of Online Contracts)

Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines…***). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issues a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

The model can be used as an aiding tool that could speed-up the process, however, I don't think it should replace human evaluation & judgement.

Respondent 5



QUESTION 3

Extract from Legal Text (Breach of Online Contracts)

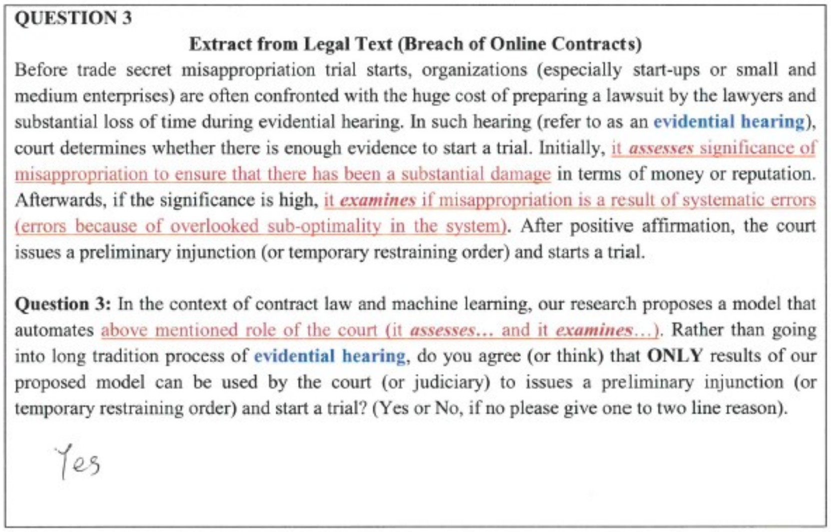
Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines…***). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issues a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

Yes → faster, costs less
As long as its verifiably
unbiased (confirmatory bias)

Comments:
Assuming start-ups or SME's go for trade secret over other patents, wouldn't that also be enough proof that the company/enterprise has access to sufficient funds? (in reference to "huge cost")

Respondent 6



QUESTION 3

Extract from Legal Text (Breach of Online Contracts)

Before trade secret misappropriation trial starts, organizations (especially start-ups or small and medium enterprises) are often confronted with the huge cost of preparing a lawsuit by the lawyers and substantial loss of time during evidential hearing. In such hearing (refer to as an **evidential hearing**), court determines whether there is enough evidence to start a trial. Initially, it ***assesses*** significance of misappropriation to ensure that there has been a substantial damage in terms of money or reputation. Afterwards, if the significance is high, it ***examines*** if misappropriation is a result of systematic errors (errors because of overlooked sub-optimality in the system). After positive affirmation, the court issues a preliminary injunction (or temporary restraining order) and starts a trial.

Question 3: In the context of contract law and machine learning, our research proposes a model that automates above mentioned role of the court (it ***assesses…*** and it ***examines***…). Rather than going into long tradition process of **evidential hearing**, do you agree (or think) that **ONLY** results of our proposed model can be used by the court (or judiciary) to issues a preliminary injunction (or temporary restraining order) and start a trial? (Yes or No, if no please give one to two line reason).

Yes

6.5 Summary

The secondary aim of this research was to develop a model that can be implemented over the blockchain to automatically issue preliminary injunction (or temporary restraining order by court of law) for the breach. This part of the research proposes an unsupervised machine learning algorithm called as Probability based Factor Model (PFM) to automatically issue a court injunction when output of a smart contract breaches the contract. The underlying concept in PFM is built upon Self-Regulated MCDA proposed in chapter 4 and stochastic modeling from the discipline of Data Science. High performance computing (HPC) cluster at University of Luxembourg (HPC @ Uni.lu) and docker (a software container platform) were used to emulate contractual environment of three service providers: Redis, MongoDB, and Memcached Servers. The breach of contract was emulated by increasing the workload on these providers. The results showed that the court injunction(s) was issued only for Redis and MongoDB Servers. Technically, this difference could be attributed to the fact that Memcached is simply used for caching and therefore, it is less prone to the breach of contract. Whereas, Redis and MongoDB as databases and message brokers are performing more complex operations and are more likely to cause a breach.

For legal validation of research findings in this chapter, using Delphi Sampling, a question was sent to law and ICT experts as a part of a questionnaire. There were total of six experts (two from the field of ICT, two from the field of law, and two from the field of ICT and Law). There were two rounds. After two rounds the sample **(4 out of 6 respondents)** disagreed for **ONLY** using the results of our proposed model by the court (or judiciary) to issue a preliminary injunction (or temporary restraining order) and starts a trial.

Chapter 7 : Conclusions and Future Directions

This chapter concludes the thesis by presenting the main research findings, academic contribution, and directions for future research. Section 7.1 presets research findings based on the primary research (chapter 2, 3, and 4). Afterwards, section 7.2 presents research findings based on the secondary research (chapter 5 and 6); section 7.3 presents academic contributions in the field of law and ICT; and finally, section 7.4 presents directions for future research.

7.1 Conclusions based on Primary Research (Chapter 2, 3, and 4)

Following are the conclusions based on the PhD research dealing with the primary research question i.e. *how an online broker can embed legal protection as preemptive measure to reduce burden of proof in a court of law?* The answer to this research question will benefit R&D based enterprises to negotiate a contract with service providers that will minimize trade secret misappropriation in the cloud.

- In chapter 2, it was concluded that the trust between the stakeholders in the cloud (consumer, service provider, auditor, and broker) is critical for planning, delivering, and consumption of cloud computing service and deployment models. One of the major issues that could aid or impair such trust is data protection. For an enterprise, data protection is protection of its business data or trade secrets in the cloud. Despite of the fact that contract can provide greater certainty for trade secrets protection in the cloud, the jurisdictional problems do not completely go away and may result in failure of legal protection of trade secrets in the cloud.
- In chapter 3, based on case law analysis in section 3.1, it was concluded that If an online broker can (1) inspect contract (or electronic contract) for compliance with non-disclosure regulations and (2) assess structural significance of criteria, then it is successfully providing legal protection for trade secrets in the cloud and subsequently, reducing

burden of proof in a court of law.

- In chapter 3, based on systematic review of literature in section 3.2, it was concluded that an online broker is still in its infancy stage when placed under the capacity to provide legal protection. Such protection is subjected to capability of online broker to ensure confidentiality that in a court of law is partially related to selection of relevant criteria for security of trade secret (goal). Statistically, relevance of criteria as per goal is its structural significance.
- In chapter 3, based on systematic review of literature in section 3.2, it was also concluded that AHP and TOPSIS in the domain of MCDA are the most prevalent techniques used by online brokers in the cloud. Both of these techniques deal with objective, criteria, and alternatives to reach a pre-established goal while assuming structural significance for criteria owning to the subjective judgments of the decision maker. This research is first in line to propose model for online brokers to assess structural significance of criteria objectively and in doing so, it uses notion of "factor loading" that belongs to broader concept of factor analysis from the domain of Unsupervised Machine Learning.
- In chapter 3, a two-stage technical evaluation procedure was implemented in section 3.4. In stage one; proposed model showed how to assess structural significance of criteria and in stage two, a comparative analysis was performed between the proposed model and its counterparts to show how results of stage 1 can be used to reduce burden of proof in a court of law. A real time QoS based dataset for seven different cloud storage providers i.e Carbonite, Dropbox, iBackup, JustCloud, SOS Online Backup, SugarSync, and Zip Cloud, was used for evaluation. It was concluded that the simulation runs in the stable environment i.e. when uncertainty is low, shows better results of proposed model as compared to its counterparts in the field.
- In chapter 3, based on Delphi Sampling in section 3.5, it was concluded that the experts in the field of ICT and law agreed that following re-

search findings (based on “case law analysis in USA”) are also common across EU member states and other countries in the world.

> *For misappropriation claim of trade secret in the cloud, plaintiff must establish three things in a court of law. They are: a)* ***presence****: it’s a proof of data in the cloud to be a trade secret, b)* ***confidentiality****: it’s a proof for reasonable efforts made by the owner to protect trade secret in the cloud, and c)* ***misappropriation****: it’s a proof for misappropriation of a trade secret by using data mining (or big data analytics).*

- In chapter 4, based on Delphi Sampling in section 4.3, it was concluded that the experts in the field of ICT and law agreed to our research findings in chapter 4 i.e. results of our model can be_used by the court (or judiciary) as a part of evidence for “reasonable step taken by the customer to keep the data secure in the cloud”.

7.2 Conclusions based on Secondary Research (Chapter 5 and 6)

Following are the conclusions based on the PhD research dealing with seconday research question i.e. *what happens when the outcome of a smart contract deviates from the outcome that the law demands?* The answer to this research question will eventually benefit blockchain driven R&D based enterprises to control and stop breach of contract that could potentially lead to trade secret misappropriation.

- In chapter 5, it was concluded that the smart contracts over the blockchain, go beyond simple transactions of bitcoins (or peer to peer economy), and have more extensive instructions (processing) embedded into them. They are reinventing almost all the categories of industries by conceiving innovations running over the blockchain. However, like a traditional contracts, smart contracts have also given rise to legal challenges in the domain of contract law, which could damage the reputation of conceived innovations. Breach of contract is one of such challenges and one of the ways to deal with it is by automating role of judiciary over a blockchain.

- In chapter 6, the research proposes an unsupervised machine learning algorithm called as Probability based Factor Model (PFM) to automatically issue a preliminary injunction (or temporary restraining order by court of law) when output of a smart contract breaches the contract. However, during legal validation using Delphi Sampling, experts in ICT and law domain disagreed for **ONLY** using the results of PFM by the court (or judiciary) to issue a preliminary injunction.

7.3 Academic Contributions in the Field of Law and ICT

Following list briefly presents research contributions of this multidisciplinary Ph.D. research in the domain of Law.

- This research is first in-line to uses ICT (unsupervised machine learning) to help owner of a trade secret to reduce burden of proof in the court. In doing so, it is first in-line to focus on “legal protection” for trade secrets in the cloud as compared to well-established similar concept of “information security”, which provides technical protection for trade secrets in the cloud e.g. encryption, hashing etc.
- The underlying notion in this PhD research is also about incorporating law into ICT architecture. However, unlike Privacy by Design (PbD) that focuses on privacy of a physical person, this research focuses on confidentiality of a legal person (an enterprise) and proposes a new concept of Confidentiality by Design (CbD). CbD includes the idea that ICT architecture should scale down burden of proof in the court of law, which could help in proving trade secret misappropriation, see chapter 3. Unlike PbD, CbD is a novel area of inter-disciplinary research whose body of knowledge is not yet well established. This PhD research is first in-line to implement notion of CbD in an online cloud environment.
- This research is first in-line to use case law together with newly proposed method of Delphi Sampling (see section 1.2, 1.3, and 1.6.3) to provide legal protection for trade secrets in the cloud. In this regards,

in the domain of case law, precedents set by previous court rulings on trade secret misappropriation (in United States of America - USA) were identified, see table 3.1. Afterwards, using Delphi Sampling, it was established that identified precedents are applicable in any jurisdiction (or most of them) around the globe and hence, they are also applicable to the cloud, see section 3.5.

- By defying the myth that “smart contracts are unbreachable” and in the context of contract law, This PhD research is first in-line to automate role of the judiciary over blockchains. In this regards, it uses unsupervised machine learning and stochastic modeling together with smart contract.

Following list briefly presents research contributions of this multidisciplinary Ph.D. research in the domain of ICT.

- In the context of online cloud environment, this PhD research is first in-line to propose self-regulated Multi-criteria Decision Analysis (MCDA) that operates without decision maker interference and well suited for the context where automation of decision making is required, see chapter 4.
- Real-world data, which is the input for data processing and analytics, are affected by many factors; among them, the presence of noise is a main factor. It is an unavoidable problem, which influence data processing and analytics. Noisy data in MCDA generally means that the decision making take account of insignificant correlations (or criteria), which could result in selection of sub-optimal or least optimal alternative. Using unsupervised machine learning (or factor analysis); this PhD research is first in-line to identify and analyze noisy data in MCDA, see sections 3.4.1 and 4.2.1.
- Communication of normative and empirical research results between the disciplines of law and ICT is one of the barriers in achieving genuine interdisciplinary validation. The proposed method of Delphi Sampling is an approximation technique for universal validation of multi-

disciplinary research results. Sections 3.5, 4.3, and 6.4 present use of Delphi Sampling to seek inter-disciplinary (ICT and law) validation of the results in this PhD research.

7.4 Future Work

In the context of this PhD research, the following list presents proposed directions for future research.

1. We plan to increase scope of literature review by also including databases other than ACM Digital Library, Google Scholar, IEEE Xplore, ScienceDirect, and SpringerLink.
2. We plan to enhance proposed models (in chapter 3 and 6) to deal with uncertainty in the system.
3. We plan to implement and test proposed models (in chapter 3 and 6) in Amazon cloud and monitor data streams for information security.

Appendix A

Script: Docker Run and Image Loading

Docker runs processes in isolated containers. A container is a process which runs on a host. The host may be local or remote. When an operator executes docker run, the container process that runs is isolated in that it has its own file system, its own networking, and its own isolated process tree separate from the host.

The basic docker run command takes this form:

```
$ docker run [OPTIONS] IMAGE[:TAG|@DIGEST] [COMMAND] [ARG...]
```

The docker run command must specify an *IMAGE* to derive the container from.

Script: Running Yahoo! Cloud System Benchmark (YCSB)

1. *Download the latest release of YCSB:*
2. curl -O --location https://github.com/brianfrankcooper/YCSB/releases/download/0.12.0/ycsb-0.12.0.tar.gz
3. tar xfvz ycsb-0.12.0.tar.gz
   cd ycsb-0.12.0

4. Set up a database to benchmark. There is a README file under each binding directory.

5. Run YCSB command.

   On Linux:

   ```
   bin/ycsb.sh load basic -P workloads/workloada
   bin/ycsb.sh run basic -P workloads/workloada
   ```

   On Windows:

   ```
   bin/ycsb.bat load basic -P workloads\workloada
   bin/ycsb.bat run basic -P workloads\workloada
   ```

Running the ycsb command without any argument will print the usage.
See https://github.com/brianfrankcooper/YCSB/wiki/Running-a-Workload for a detailed documentation on how to run a workload.
See https://github.com/brianfrankcooper/YCSB/wiki/Core-Properties for the list of available workload properties.

Building from source

YCSB requires the use of Maven 3; if you use Maven 2, you may see errors such as these.
To build the full distribution, with all database bindings:

```
mvn clean package
```

To build a single database binding:

```
mvn -pl com.yahoo.ycsb:mongodb-binding -am clean package
```

Script: Running MongoDB and YCSB

1. Start MongoDB

First, download MongoDB and start mongod. For example, to start MongoDB on x86-64 Linux box:

```
wget http://fastdl.mongodb.org/linux/mongodb-linux-x86_64-x.x.x.tgz
tar xfvz mongodb-linux-x86_64-*.tgz
mkdir /tmp/mongodb
cd mongodb-linux-x86_64-*
./bin/mongod --dbpath /tmp/mongodb
```

Replace x.x.x above with the latest stable release version for MongoDB.
See http://docs.mongodb.org/manual/installation/ for installation steps for various operating systems.

2. Install Java and Maven

Go to http://www.oracle.com/technetwork/java/javase/downloads/index.html and get the url to download the rpm into your server. For example:

```
wget http://download.oracle.com/otn-pub/java/jdk/7u40-b43/jdk-7u40-linux-x64.rpm?AuthParam=11232426132 -o jdk-7u40-linux-x64.rpm
rpm -Uvh jdk-7u40-linux-x64.rpm
```

Or install via yum/apt-get

```
sudo yum install java-devel
```

Download MVN from http://maven.apache.org/download.cgi

```
wget http://ftp.heanet.ie/mirrors/www.apache.org/dist/maven/maven-3/3.1.1/binaries/apache-maven-3.1.1-bin.tar.gz
sudo tar xzf apache-maven-*-bin.tar.gz -C /usr/local
cd /usr/local
sudo ln -s apache-maven-* maven
sudo vi /etc/profile.d/maven.sh
```

Add the following to maven.sh

```
export M2_HOME=/usr/local/maven
export PATH=${M2_HOME}/bin:${PATH}
```

Reload bash and test mvn

```
bash
mvn -version
```

3. Set Up YCSB

Download the YCSB zip file and compile:

```
curl -O --location https://github.com/brianfrankcooper/YCSB/releases/download/0.5.0/ycsb-0.5.0.tar.gz
tar xfvz ycsb-0.5.0.tar.gz
cd ycsb-0.5.0
```

4. Run YCSB

Now you are ready to run! First, use the asynchronous driver to load the data:

```
./bin/ycsb load mongodb-async -s -P workloads/workloada > outputLoad.txt
```

Then, run the workload:

```
./bin/ycsb run mongodb-async -s -P workloads/workloada > outputRun.txt
```

Similarly, to use the synchronous driver from MongoDB Inc. we load the data:

```
./bin/ycsb load mongodb -s -P workloads/workloada > outputLoad.txt
```

Then, run the workload:

```
./bin/ycsb run mongodb -s -P workloads/workloada > outputRun.txt
```

See the next section for the list of configuration parameters for MongoDB.

Log Level Control

Due to the mongodb driver defaulting to a log level of DEBUG, a logback.xml file is included with this module that restricts the org.mongodb logging to WARN. You can control this by overriding the logback.xml and defining it in your ycsb command by adding this flag:

```
bin/ycsb run mongodb -jvm-args="-Dlogback.configurationFile=/path/to/logback.xml"
```

MongoDB Configuration Parameters

- mongodb.url
    - This should be a MongoDB URI or connection string.
        - See http://docs.mongodb.org/manual/reference/connection-string/ for the standard options.
        - For the complete set of options for the asynchronous driver see:
            - http://www.allanbank.com/mongodb-async-driver/apidocs/index.html?com/allanbank/mongodb/MongoDbUri.html
        - For the complete set of options for the synchronous driver see:
            - http://api.mongodb.org/java/current/index.html?com/mongodb/MongoClientURI.html
    - Default value is mongodb://localhost:27017/ycsb?w=1

  - Default value of database is ycsb
- mongodb.batchsize
  - Useful for the insert workload as it will submit the inserts in batches inproving throughput.
  - Default value is 1.
- mongodb.upsert
  - Determines if the insert operation performs an update with the upsert operation or a insert. Upserts have the advantage that they will continue to work for a partially loaded data set.
  - Setting to true uses updates, false uses insert operations.
  - Default value is false.
- mongodb.writeConcern
  - **Deprecated** - Use the w and journal options on the MongoDB URI provided by the mongodb.url.
  - Allowed values are :
    - errors_ignored
    - unacknowledged
    - acknowledged
    - journaled
    - replica_acknowledged
    - majority
  - Default value is acknowledged.
- mongodb.readPreference
  - **Deprecated** - Use the readPreference options on the MongoDB URI provided by the mongodb.url.
  - Allowed values are :
    - primary
    - primary_preferred
    - secondary
    - secondary_preferred
    - nearest

    - Default value is primary.
- mongodb.maxconnections
    - **Deprecated** - Use the maxPoolSize options on the MongoDB URI provided by the mongodb.url.
    - Default value is 100.
- mongodb.threadsAllowedToBlockForConnectionMultiplier
    - **Deprecated** - Use the waitQueueMultiple options on the MongoDB URI provided by the mongodb.url.
    - Default value is 5.

For example:

```
./bin/ycsb load mongodb-async -s -P workloads/workloada -p mongodb.url=mongodb://localhost:27017/ycsb?w=0
```

To run with the synchronous driver from MongoDB Inc.:

```
./bin/ycsb load mongodb -s -P workloads/workloada -p mongodb.url=mongodb://localhost:27017/ycsb?w=0
```

Script: Running Memcached and YCSB

1. Install and start memcached service on the host(s)

Debian / Ubuntu:

```
sudo apt-get install memcached
```

RedHat / CentOS:

```
sudo yum install memcached
```

2. Install Java and Maven

See step 2 in ../mongodb/README.md.

3. Set up YCSB

Git clone YCSB and compile:

```
git clone http://github.com/brianfrankcooper/YCSB.git
cd YCSB
mvn -pl com.yahoo.ycsb:memcached-binding -am clean package
```

4. Load data and run tests

Load the data:

```
./bin/ycsb load memcached -s -P workloads/workloada > outputLoad.txt
```

Run the workload test:

```
./bin/ycsb run memcached -s -P workloads/workloada > outputRun.txt
```

5. memcached Connection Parameters

A sample configuration is provided in conf/memcached.properties.

Required params

- memcached.hosts

  This is a comma-separated list of hosts providing the memcached interface. You can use IPs or hostnames. The port is optional and defaults to the memcached standard port of 11211 if not specified.

Optional params

- memcached.shutdownTimeoutMillis

  Shutdown timeout in milliseconds.

- memcached.objectExpirationTime

  Object expiration time for memcached; defaults to Integer.MAX_VALUE.

- memcached.checkOperationStatus

  Whether to verify the success of each operation; defaults to true.

- memcached.readBufferSize

  Read buffer size, in bytes.

- memcached.opTimeoutMillis

  Operation timeout, in milliseconds.

- memcached.failureMode

  What to do with failures; this is one of net.spy.memcached.FailureMode enum values, which are currently: Redistribute, Retry, or Cancel.

- memcached.protocol Set to 'binary' to use memcached binary protocol. Set to 'text' or omit this field to use memcached text protocol

You can set properties on the command line via -p, e.g.:

```
./bin/ycsb load memcached -s -P workloads/workloada \
  -p "memcached.hosts=127.0.0.1" > outputLoad.txt
```

Script: Running Redis and YCSB

1. Start Redis (same as MongoDB and Memcached)

2. Install Java and Maven (same as MongoDB and Memcached)

3. Set Up YCSB

Git clone YCSB and compile:

```
git clone http://github.com/brianfrankcooper/YCSB.git
cd YCSB
mvn -pl com.yahoo.ycsb:redis-binding -am clean package
```

4. Provide Redis Connection Parameters

Set the host, port, and password (do not redis auth is not turned on) in the workload you plan to run.

- redis.host
- redis.port
- redis.password

Or, you can set configs with the shell command, EG:

```
./bin/ycsb load redis -s -P workloads/workloada -p "redis.host=127.0.0.1" -p "redis.port=6379" > outputLoad.txt
```

5. Load data and run tests

Load the data:

```
./bin/ycsb load redis -s -P workloads/workloada > outputLoad.txt
```

Run the workload test:

```
./bin/ycsb run redis -s -P workloads/workloada > outputRun.txt
```

Script: Python Monitoring Redis, MongoDB, and Memcached using YCSB

```
import subprocess
```

```
from subprocess import check_output
from subprocess import call
import csv
import time
from datetime import datetime
import os

host = "redis.host=127.0.0.1"
port = "redis.port=32768"
exName = "YSCB-Redis"
server = "redis"
workLoad = "workloads/workloada"
No_oper = str(1000)
No_reco = str(1000)
tpar = 10 #parallel connections
num = 0
d = 3
x = 1
nq = 100
data = str(32) #data size
P = "redis" #protocol, by default Redis
par = 1000
par2 = 1000
test = str(10) #num of threads/connection
#Threads = "threadcount="+test

csvfile = open("YCSB_output.csv", 'w') ####################CHANGE

fieldnames = ['count','Date','ExpName', 'WorkLoad','No Operations','No Records','No_Threads','Runtime(ms)','Thr(ops/sec)','cleanup_lat(us)','readFail_lat(us)','Read_ReturnOk','Read_ReturnErr','read_lat(us)','update_lat(us)']
```

```
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()

    ##./bin/ycsb load redis -s -P workloads/workloada -p "redis.host=127.0.0.1" -p
"redis.port=6379" -p "threadcount=10"
    ## mvn -pl com.yahoo.ycsb:redis-binding -am clean package
    ## recordcount=1000
    ## operationcount=1000

    os.chdir("/home/spark/broker/YCSB") ####################CHANGE
    print "==================Directory Changed to YCSB
!!=======================================\n\n"
    draft = call(["mvn", "-pl", "com.yahoo.ycsb:redis-binding", "-am",
"clean","package"])
    print "=============================Maven
Called====================================\n\n"
    subprocess.call(["./bin/ycsb", "load", server,"-s", "-P", workLoad,"-p",host,"-
p",port,"-threads",test,"-p","recordcount="+No_reco,"-
p","operationcount="+No_oper],stdout=subprocess.PIPE)
    print "=============================YCSB TEST
LOADED====================================\n\n\n"
    print "=============================YCSB TEST will
RUN====================================\n\n\n"
    for num in range (0,100):
        print "================= New Run ================== \n\n"
        if num % 10 == 0 :
          tpar = tpar + 20
        No_oper = str(par)
        No_reco = str(par2)
        test = str(tpar)
        #data = str(d)
    #c = str(cpar)
    test_out = check_output(["./bin/ycsb", "run", server,"-s", "-P", workLoad,"-
p",host,"-p",port,"-threads",test,"-p","recordcount="+No_reco,"-
p","operationcount="+No_oper])
    #print test_out
    a = test_out.split('\n')
    list = []
    listb = []
    for g in range (0, len(a)):
      e = a[g].split(', ')
      if (e[0] == '[OVERALL]' and e[1] == 'RunTime(ms)') or (e[0] ==
'[OVERALL]' and e[1] == 'Throughput(ops/sec)') or (e[0] == '[CLEANUP]' and
e[1] == 'AverageLatency(us)') or (e[0] == '[READ]' and e[1] == 'AverageLatency(us)') or (e[0] == '[UPDATE]' and e[1] == 'AverageLatency(us)') or (e[0] ==
'[READ-FAILED]' and e[1] =='AverageLatency(us)') or (e[0] =='[READ]' and
e[1] == 'Return=OK') or (e[0] =='[READ]' and e[1] == 'Return=ERROR'):
        list.append(g)
       listb.append(float(e[2]))
  ### listb[0] = runtime(ms) listb[1] = thr(ops/sec) lisbt[2] = cleanup_lat(us)
listb[3] = readFail_lat(us) listb[4] = Read_ReturnOk listb[5] = Read_ReturnErr
listb[6] = read_lat(us) listb[7] = update_lat(us)
    #print len(listb),"   ===  ", len(list)
     if len(listb) == 6:
      print
num,"=====","n_oper=",No_oper,"=====","t=",test,"====",list[0],"====",list[1
],"====", list[2] , "====" , list[3], "=====", list[4],"====",list[5]
      writer.writerow({'count':num,'Date':time.strftime("%d/%m/%Y"),'ExpName':exName,
'WorkLoad':workLoad,'No Operations':No_oper,'No Records':No_reco,'No_Threads':test,'Runtime(ms)':listb[0],'Thr(ops/sec)':listb[1],'clea
```

```
nup_lat(us)':listb[2],'readFail_lat(us)':'NULL','Read_ReturnOk':listb[4],'Read_Ret
urnErr':'NULL','read_lat(us)':listb[3],'update_lat(us)':listb[5]})
        elif len(listb) < 6 and workLoad == "workloads/workloadc" :
          print
num,"=====","n_oper=",No_oper,"=====","t=",test,"====",list[0],"====",list[1
],"====", list[2] , "====" , list[3], "=====", list[4]
          writer.writerow({'count':num,'Date':time.strftime("%d/%m/%Y"),'ExpName':exName,
'WorkLoad':workLoad,'No Operations':No_oper,'No Records':No_reco,'No_Threads':test,'Runtime(ms)':listb[0],'Thr(ops/sec)':listb[1],'clea
nup_lat(us)':listb[2],'readFail_lat(us)':'NULL','Read_ReturnOk':listb[4],'Read_Ret
urnErr':'NULL','read_lat(us)':listb[3],'update_lat(us)':'NULL'})
        elif len(listb) < 8 and workLoad == "workloads/workloadc" :
           print
num,"=====","n_oper=",No_oper,"=====","t=",test,"====",list[0],"====",list[1
],"====", list[2] , "====" , list[3], "=====", list[4],"====",list[5],"====",list[6]
          writer.writerow({'count':num,'Date':time.strftime("%d/%m/%Y"),'ExpName':exName,
'WorkLoad':workLoad,'No Operations':No_oper,'No Records':No_reco,'No_Threads':test,'Runtime(ms)':listb[0],'Thr(ops/sec)':listb[1],'clea
nup_lat(us)':listb[3],'readFail_lat(us)':listb[2],'Read_ReturnOk':listb[5],'Read_Retu
rnErr':listb[6],'read_lat(us)':listb[4],'update_lat(us)':'NULL'})
        else :
           print
num,"=====","n_oper=",No_oper,"=====","t=",test,"====",list[0],"====",list[1
],"====", list[2] , "====" , list[3], "=====",
list[4],"====",list[5],"====",list[6],"====",list[7]
          writer.writerow({'count':num,'Date':time.strftime("%d/%m/%Y"),'ExpName':exName,
'WorkLoad':workLoad,'No Operations':No_oper,'No Records':No_reco,'No_Threads':test,'Runtime(ms)':listb[0],'Thr(ops/sec)':listb[1],'clea
```

```
nup_lat(us)':listb[3],'readFail_lat(us)':listb[2],'Read_ReturnOk':listb[5],'Read_ReturnErr':listb[6],'read_lat(us)':listb[4],'update_lat(us)':listb[7]})
        #d = d + 1
        par = par + 2000
        par2 = par2 + 2000
        time.sleep(10)

    print "YCSB !! Benchmark is finished !!"
```

References

[1] P. J. Stumbo, "Funding nutrition software development: The Small Business Innovation Research (SBIR) program," *Journal of Food Composition and Analysis,* vol. 14, pp. 329-332, Jun 2001.

[2] M. I. P. Silaghi, D. Alexa, C. Jude, and C. Litan, "Do business and public sector research and development expenditures contribute to economic growth in Central and Eastern European Countries? A dynamic panel estimation," *Economic Modelling,* vol. 36, pp. 108-119, Jan 2014.

[3] M. Falk, "What drives business research and development (R&D) intensity across organisation for economic co-operation and development (OECD) countries?," *Applied Economics,* vol. 38, pp. 533-547, Mar 20 2006.

[4] C. M. Sweet and D. S. E. Maggio, "Do Stronger Intellectual Property Rights Increase Innovation?," *World Development,* vol. 66, pp. 665-677, Feb 2015.

[5] E. Ottoz and F. Cugno, "Choosing the scope of trade secret law when secrets complement patents," *International Review of Law and Economics,* vol. 31, pp. 219-227, Dec 2011.

[6] I. Daizadeh, D. Miller, A. Glowalla, M. Leamer, R. Nandi, and C. I. Numark, "A general approach for determining when to patent, publish, or protect information as a trade secret," *Nature Biotechnology,* vol. 20, pp. 1053-1054, Oct 2002.

[7] D. D. Friedman, W. M. Landes, and R. A. Posner, "Some Economics of Trade Secret Law," *Journal of Economic Perspectives,* vol. 5, pp. 61-72, Win 1991.

[8] J. R. Kalyvas and M. R. Overly, *Big Data: A Business and Legal Guide*: CRC Press, 2014.

[9] Z. H. Zhou, N. V. Chawla, Y. C. Jin, and G. J. Williams, "Big Data Opportunities and Challenges: Discussions from Data Analytics

Perspectives," *Ieee Computational Intelligence Magazine,* vol. 9, pp. 62-74, Nov 2014.

[10] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *Journal of Parallel and Distributed Computing,* vol. 74, pp. 2561-2573, Jul 2014.

[11] E. Casalicchio and M. Palmirani, "A Cloud Service Broker with Legal-Rule Compliance Checking and Quality Assurance Capabilities," *Procedia Computer Science,* vol. 68, pp. 136-150, 2015.

[12] Q. Dupont and B. Maurer, "Ledgers and Law in the Blockchain," *Kings Review (23 June 2015) http://kingsreview. co. uk/magazine/blog/2015/06/23/ledgers-and-law-in-the-blockchain,* 2015.

[13] K. Delmolino, M. Arnett, A. Kosba, A. Miller, and E. Shi, "Step by step towards creating a safe smart contract: Lessons and insights from a cryptocurrency lab," in *International Conference on Financial Cryptography and Data Security*, 2016, pp. 79-94.

[14] M. Swan, *Blockchain: Blueprint for a new economy*: " O'Reilly Media, Inc.", 2015.

[15] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker*, et al.*, "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 conference on Internet measurement conference*, 2013, pp. 127-140.

[16] X. Xu, C. Pautasso, L. Zhu, V. Gramoli, A. Ponomarev, A. B. Tran*, et al.*, "The blockchain as a software connector," in *Software Architecture (WICSA), 2016 13th Working IEEE/IFIP Conference on*, 2016, pp. 182-191.

[17] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the internet of things," *IEEE Access,* vol. 4, pp. 2292-2303, 2016.

[18] A. Savelyev, "Contract law 2.0:'Smart'contracts as the beginning of the end of classic contract law," *Information & Communications Technology Law,* vol. 26, pp. 116-134, 2017.

[19] C. L. Knapp, N. M. Crystal, and H. G. Prince, *Problems in Contract Law: cases and materials*: Wolters Kluwer Law & Business, 2016.

[20] H. A. Linstone and M. Turoff, *The Delphi method: Techniques and applications* vol. 29: Addison-Wesley Reading, MA, 1975.

[21] G. Barzilai, *Communities and law: Politics and cultures of legal identities*: University of Michigan Press, 2010.

[22] J. Cave, N. Robinson, S. Kobzar, and H. R. Schindler, "Regulating the cloud: more, less or different regulation and competing agendas," 2012.

[23] S. Biggs and S. Vidalis, "Cloud computing: The impact on digital forensic investigations," in *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for*, 2009, pp. 1-6.

[24] J.-H. Morin, J. Aubert, and B. Gateau, "Towards cloud computing SLA risk management: issues and challenges," in *System Science (HICSS), 2012 45th Hawaii International Conference on*, 2012, pp. 5509-5514.

[25] J. Steiner, *The foundations of deliberative democracy: Empirical research and normative implications*: Cambridge University Press, 2012.

[26] W. H. Newell, J. Wentworth, and D. Sebberson, "A theory of interdisciplinary studies," *Issues in Interdisciplinary Studies,* 2001.

[27] L. Qian, Z. Luo, Y. Du, and L. Guo, "Cloud computing: An overview," *Cloud computing,* pp. 626-631, 2009.

[28] S. K. Sandeen, "Lost in the cloud: Information flows and the implications of cloud computing for trade secret protection," *Va. JL & Tech.,* vol. 19, p. 1, 2014.

[29] R. L. Krutz and R. D. Vines, *Cloud security: A comprehensive guide to secure cloud computing*: Wiley Publishing, 2010.

[30] L. Kaplow, "Burden of proof," *The Yale Law Journal,* pp. 738-859, 2012.

[31] A. Cavoukian and J. Jonas, *Privacy by design in the age of big data*: Information and Privacy Commissioner of Ontario, Canada, 2012.

[32] O. W. Holmes, *The common law*: Harvard University Press, 2009.

[33] J. H. Merryman and R. Pérez-Perdomo, *The civil law tradition: an introduction to the legal systems of Europe and Latin America*: Stanford University Press, 2007.

[34] D. Svantesson and R. Clarke, "Privacy and consumer risks in cloud computing," *Computer law & security review,* vol. 26, pp. 391-397, 2010.

[35] S. S. Diamond, L. E. Bowman, M. Wong, and M. M. Patton, "Efficiency and cost: The impact of videoconferenced hearings on bail decisions," *The Journal of Criminal Law and Criminology (1973-),* vol. 100, pp. 869-902, 2010.

[36] M. J. Hutter, "Trade secret misappropriation: a lawyer's practical approach to the case law," *W. New Eng. L. Rev.,* vol. 1, p. 1, 1978.

[37] S. Khaddaj, "Cloud computing: service provisioning and user requirements," in *Distributed Computing and Applications to Business, Engineering & Science (DCABES), 2012 11th International Symposium on*, 2012, pp. 191-195.

[38] C.-Y. Yu, *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* vol. 30: University of California, Los Angeles Los Angeles, 2002.

[39] T. Sathyanarayana and L. M. I. Sheela, "Data security in cloud computing," in *Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on*, 2013, pp. 822-827.

[40] M. D. Ryan, "Cloud computing security: The scientific challenge, and a survey of solutions," *Journal of Systems and Software,* vol. 86, pp. 2263-2268, 2013.

[41] P. De Hert and V. Papakonstantinou, "The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals," *Computer Law & Security Review,* vol. 28, pp. 130-142, 2012.

[42] P. A. Baistrocchi, "Liability of intermediary service providers in the EU Directive on Electronic Commerce," *Santa Clara Computer & High Tech. LJ,* vol. 19, p. 111, 2002.

[43] M. Schnjakin, R. Alnemr, and C. Meinel, "Contract-based cloud architecture," in *Proceedings of the second international workshop on Cloud data management*, 2010, pp. 33-40.

[44] P. Wieder, J. M. Butler, W. Theilmann, and R. Yahyapour, *Service level agreements for cloud computing*: Springer Science & Business Media, 2011.

[45] D. D. Lamanna, J. Skene, and W. Emmerich, "SLAng: A language for service level agreements," 2003.

[46] G. Di Vita, "The TRIPs agreement and technological innovation," *Journal of Policy Modeling,* vol. 35, pp. 964-977, Nov-Dec 2013.

[47] S. K. Sandeen, "The limits of trade secret law: Article 39 of the TRIPS Agreement and the Uniform Trade Secrets Act on which it is based," *Law and Theory of Trade Secrecy: A Handbook of Contemporary Research,* pp. 537-567, 2011.

[48] L. F. Dong, "Issues and strategies of China IP protection after the TRIPS Agreement," *Trips and Developing Countries: Towards a New Ip World Order?,* pp. 39-71, 2014.

[49] "Study on Trade Secrets and Confidential Business Information in the Internal Market " European Commission2013.

[50] M. Z. M. Nomani and F. Rahman, "Intellection of Trade Secret and Innovation Laws in India," *Journal of Intellectual Property Rights,* vol. 16, pp. 341-350, Jul 2011.

[51] M. Verbeek, *A guide to modern econometrics*: John Wiley & Sons, 2008.

[52] R. J. Rummel, *Applied factor analysis*: Northwestern University Press, 1988.

[53] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: comparing public cloud providers," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 2010, pp. 1-14.

[54] S. Sundareswaran, A. Squicciarini, and D. Lin, "A Brokerage-Based Approach for Cloud Service Selection," in *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, 2012, pp. 558-565.

[55] S. Gong and K. M. Sim, "CB-Cloudle: A Centroid-based Cloud Service Search Engine," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2014.

[56] J. Kang and K. M. Sim, "Cloudle: an ontology-enhanced cloud service search engine," in *Web Information Systems Engineering–WISE 2010 Workshops*, 2011, pp. 416-427.

[57] X. Wang, J. Cao, and Y. Xiang, "Dynamic cloud service selection using an adaptive learning mechanism in multi-cloud computing," *Journal of Systems and Software,* vol. 100, pp. 195-210, 2015.

[58] D. Ergu, G. Kou, Y. Peng, Y. Shi, and Y. Shi, "The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment," *The Journal of Supercomputing,* vol. 64, pp. 835-848, 2013.

[59] C.-H. Su, G.-H. Tzeng, and H.-L. Tseng, "Improving cloud computing service in fuzzy environment—combining fuzzy DANP and fuzzy VIKOR with a new hybrid FMCDM model," in *Fuzzy Theory and it's Applications (iFUZZY), 2012 International Conference on*, 2012, pp. 30-35.

[60] B. Y. Ooi, H. Y. Chan, and Y.-N. Cheah, "Resource selection using fuzzy logic for dynamic service placement and replication," in *TENCON 2011-2011 IEEE Region 10 Conference*, 2011, pp. 128-132.

[61] J. Rao, Y. Wei, J. Gong, and C.-Z. Xu, "DynaQoS: model-free self-tuning fuzzy control of virtualized resources for QoS provisioning," in *Quality of Service (IWQoS), 2011 IEEE 19th International Workshop on*, 2011, pp. 1-9.

[62] Y. O. Yazir, C. Matthews, R. Farahbod, S. Neville, A. Guitouni, S. Ganti*, et al.*, "Dynamic resource allocation in computing clouds using

distributed multiple criteria decision analysis," in *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*, 2010, pp. 91-98.

[63] O. Boutkhoum, M. Hanine, T. Agouti, and A. Tikniouine, "A decision-making approach based on fuzzy AHP-TOPSIS methodology for selecting the appropriate cloud solution to manage big data projects," *International Journal of System Assurance Engineering and Management,* pp. 1-17, 2017.

[64] R. Garg, M. Heimgartner, and B. Stiller, "Decision Support System for Adoption of Cloud-based Services," 2016.

[65] L. H. Nunes, J. C. Estrella, C. Perera, S. Reiff-Marganiec, and A. C. Botazzo Delbem, "Multi-criteria IoT resource discovery: a comparative analysis," *Software: Practice and Experience,* 2016.

[66] S. K. Garg, S. Versteeg, and R. Buyya, "A framework for ranking of cloud computing services," *Future Generation Computer Systems,* vol. 29, pp. 1012-1023, 6// 2013.

[67] M. Sun, T. Zang, X. Xu, and R. Wang, "Consumer-centered cloud services selection using AHP," in *Service Sciences (ICSS), 2013 International Conference on*, 2013, pp. 1-6.

[68] C.-T. Chen and K.-H. Lin, "A decision-making method based on interval-valued fuzzy sets for cloud service evaluation," in *New Trends in Information Science and Service Science (NISS), 2010 4th International Conference on*, 2010, pp. 559-564.

[69] C.-C. Lo, D.-Y. Chen, C.-F. Tsai, and K.-M. Chao, "Service selection based on fuzzy TOPSIS method," in *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, 2010, pp. 367-372.

[70] D.-Y. Cheng, K.-M. Chao, C.-C. Lo, and C.-F. Tsai, "A user centric service-oriented modeling approach," *World Wide Web,* vol. 14, pp. 431-459, 2011.

[71] L. Qu, Y. Wang, and M. Orgun, "Cloud service selection based on the aggregation of user feedback and quantitative performance assessment,"

in *Services Computing (SCC), 2013 IEEE International Conference on*, 2013, pp. 152-159.

[72] H. Wu, Q. Wang, and K. Wolter, "Methods of cloud-path selection for offloading in mobile cloud computing systems," in *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, 2012, pp. 443-448.

[73] P. Gulia and S. Sood, "Automatic Selection and Ranking of Cloud Providers using Service Level Agreements," *International Journal of Computer Applications,* vol. 72, 2013.

[74] A. C. Rencher, *Methods of multivariate analysis* vol. 492: John Wiley & Sons, 2003.

[75] J. Figueira, S. Greco, and M. Ehrgott, *Multiple criteria decision analysis: state of the art surveys* vol. 78: Springer Science & Business Media, 2005.

[76] C. Macharis and A. Bernardini, "Reviewing the use of Multi-Criteria Decision Analysis for the evaluation of transport projects: Time for a multi-actor approach," *Transport Policy,* vol. 37, pp. 177-186, 2015.

[77] I. Ivlev, J. Vacek, and P. Kneppo, "Multi-criteria decision analysis for supporting the selection of medical devices under uncertainty," *European Journal of Operational Research,* vol. 247, pp. 216-228, 2015.

[78] H. Broekhuizen, C. G. Groothuis-Oudshoorn, J. A. van Til, J. M. Hummel, and M. J. IJzerman, "A Review and Classification of Approaches for Dealing with Uncertainty in Multi-Criteria Decision Analysis for Healthcare Decisions," *PharmacoEconomics,* vol. 33, pp. 445-455, 2015.

[79] V. Belton and T. Stewart, *Multiple criteria decision analysis: an integrated approach*: Springer Science & Business Media, 2002.

[80] A. Ishizaka and P. Nemery, *Multi-criteria decision analysis: methods and software*: John Wiley & Sons, 2013.

[81] J. Bilcke, P. Beutels, M. Brisson, and M. Jit, "Accounting for methodological, structural, and parameter uncertainty in decision-analytic

models a practical guide," *Medical Decision Making,* vol. 31, pp. 675-692, 2011.

[82] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," ed, 2008.

[83] D. Tapscott and A. Tapscott, *Blockchain Revolution: How the technology behind Bitcoin is changing money, business, and the world*: Penguin, 2016.

[84] A. Hayes, "What factors give cryptocurrencies their value: An empirical analysis," *Browser Download This Paper,* 2015.

[85] H. Watanabe, S. Fujimura, A. Nakadaira, Y. Miyazaki, A. Akutsu, and J. Kishigami, "Blockchain contract: Securing a blockchain applied to smart contracts," in *Consumer Electronics (ICCE), 2016 IEEE International Conference on*, 2016, pp. 467-468.

[86] S. Omohundro, "Cryptocurrencies, smart contracts, and artificial intelligence," *AI matters,* vol. 1, pp. 19-21, 2014.

[87] J. L. Zhao, S. Fan, and J. Yan, "Overview of business innovations and research opportunities in blockchain and introduction to the special issue," *Financial Innovation,* vol. 2, p. 28, 2016.

[88] W. A. Kaal, "Blockchain Innovation for Private Investment Funds," 2017.

[89] J. J. Sikorski, J. Haughton, and M. Kraft, "Blockchain technology in the chemical industry: Machine-to-machine electricity market," *Applied Energy,* vol. 195, pp. 234-246, 2017.

[90] H. M. Taylor and S. Karlin, *An introduction to stochastic modeling*: Academic press, 2014.