Assessing Data Breach Risk in Cloud Systems

Yogachandran Rahulamathavan*, Muttukrishnan Rajarajan†, Omer F. Rana‡, Malik S. Awan§,
Pete Burnap¶, and Sajal K. Das‖
*†School of Engineering and Mathematics and Computer Science, City University London, U.K.
‡§¶School of Computer Science and Informatics, Cardiff University, U.K.
‖Department of Computer Science, Missouri University of Science and Technology, Rolla, USA
Email: {*yogachandran.rahulamathavan.1, †r.muttukrishnan}@city.ac.uk, {‡ranaof,§malik.s.awan,¶BurnapP}@cardiff.ac.uk
‖sdas@mst.edu

***Abstract*—The emerging cloud market introduces a multitude of cloud service providers, making it difficult for consumers to select providers who are likely to be a *low risk* from a security perspective. Recently, significant emphasis has arisen on the need to specify Service Level Agreements that address security concerns of consumers (referred to as SecSLAs) – these are intended to clarify security support in addition to Quality of Service characteristics associated with services. It has been found that such SecSLAs are not consistent among providers, even though they offer services with similar functionality. However, measuring security service levels and the associated risk plays an important role when choosing a cloud provider. Data breaches have been identified as a high priority threat influencing the adoption of cloud computing. This paper proposes a general analysis framework which can compute risk associated with data breaches based on pre-agreed SecSLAs for different cloud providers. The framework exploits a tree based structure to identify possible attack scenarios that can lead to data breaches in the cloud and a means of assessing the use of potential mitigation strategies to reduce such breaches.**

I. Introduction

Despite the advantages and rapid growth of cloud computing, existing cloud environments are still not seen to be sufficiently trustworthy by consumers. This framework enables consumers to specify which security parameters are most significant for them, enabling a subjective view to be formed of different cloud providers.

There have been significant recent advances in cloud computing – enabling providers to differentiate themselves on a number of different factors primarily centered around capabilities and costs. Security remains an important concern for many users, particularly prevention of data breaches at the cloud provider and the ability of a provider to interrogate data stored at their systems. Many providers are consequently responding to this security challenge by improving the types of security mechanisms they support. Traditionally customers choose the cloud provider based on metrics such as number and types of CPUs (e.g. large vs. small instances), number and types of virtual machines (e.g., Debian, CentOS), and storage space, etc. Recent security breaches across the globe have changed this trend and has prompted cloud providers to include security attribute as part of above metrics. Various recent efforts have attempted to specify these security parameters [2], [5]–[7], [15], enabling customers to monitor whether particular security constraints are being met (although this is not always possible to measure, requiring the customer to often rely on the advertised capability from the provider).

One of the drawbacks of this approach is that the customer does not know the level of risk associated with security service levels agreed by the cloud provider. Cloud providers generally offer different levels of security. For example one cloud provider offers a key size of 1024 bits for RSA encryption and 256 bits for AES encryption, while another cloud provider offers a key size of 2048 bits for RSA encryption and 128 bits for AES encryption. Consequently, the risk associated with the use of these two approaches are different. Cloud providers and consumers (or clients) have advantages and disadvantages with both of these combinations. The clients may be charged differently depending on what they decide to use. However, it would be more appropriate if this process was more closely driven by consumer requirements.

In this paper we attempt to provide a framework to respond to this challenge, by utilising a tree based model to assess risk from a customer's perspective. The framework identifies the weaknesses in security provisions made by cloud providers – by identifying possible attack routes or attack scenarios for adversaries based on available security services. The framework can be used as a decision support mechanism to analyze the capability of the attack source and estimate the likely impact of an attack route. The proposed work assigns a number of attributes such as technical difficulty, cost to break the system, and attack discovery to each leaf node of the tree and aggregates them via multiple utility functions. Users can prioritize on particular security attributes that are significant for them by using a weighting/utility function which combines the attributes of leaf nodes. Finally a route which shows high risk is identified and notified to the customer.

Recently, Cloud Security Alliance[a] has identified a number of vulnerabilities to the cloud computing and placed data breach as the top threat to the cloud. A data breach is a security incident in which sensitive, protected or confidential data is copied, transmitted, viewed, stolen, or lost. Hence this paper focuses on data breach threat in the cloud. In our tree based architecture, data breach is the root node while a number of branches and leaf nodes are defined to capture different types

[a]The Notorious Nine, Cloud Computing Top Threats in 2013, http://tinyurl.com/mgt85e2

of data breach scenarios.

II. Related Works

The adoption of cloud computing (like other internet technologies) is often based on perceived risk by a customers [29] – as essentially by outsourcing their systems to a cloud service provider (CSP), the customer is placing trust in the provider to deliver its advertised capability (both in terms of QoS and security). In recent years, significant work has focused on defining and deriving security service level agreements (SecSLAs) [5], [16], [17], primarily quantifying such security services into multiple levels and enforcing a strategy to monitor whether the agreed level of services to a client are actually being delivered. Here, establishing "trust" in a provider remains important [30], [31], with trust being classified into the following five categories: provision, access, delegation, identity, and context. These categories define trust relationships between a consumer and: (i) a service provider, (ii) resources made available by the provider, (iii) a third-party arbitrator, (iv) signed attributes, and (v) supporting transactions.

Trust in cloud systems is often subjective and may be calculated using a centralized or distributed approach. When using a centralized approach, a single authority or trust broker collects all ratings from consumers, computes a reputation score for every participant, and makes all scores publicly available. When using a distributed approach, there can be distributed storages where ratings can be submitted, requiring interaction between storages to compute a single trust value for a given provider. A broker based trust model was proposed in [1] based on SLA violation and user experience where the authors exploited SLA and cloud characteristic parameters such as CPU, number of virtual machines (VMs), and service down time, for evaluating the trustworthiness of providers. This approach is also robust against malicious group of entities performing reputation based attacks. Recently, Ghosh et. al. [26] proposed SelCSP, a risk model which enables clients to select the most reliable CSP by using trustworthiness and competence of each CSP to estimate the provider reliability. SelCSP focuses on metrics such as number of CPUs and VMs, down time and interaction.

Our focus in this paper is primarily on security related metrics such as authentication, data confidentiality, access control etc, using a model based on an attack tree [32]. Attack tree analysis is a process of analyzing how systems fail, enabling the study of possible vulnerabilities within systems, visualize those vulnerabilities and assign various weights to determine which scenarios are most likely to occur.

Attack tree analysis has been extensively used in several areas including software design models [33], Internet security [34], computer security [35]. We use an attack tree based approach to identify risk associated with a data breach in cloud systems. The key contributions of this paper are two folds: (i) a framework and an associated model to compute data breach risk for multiple CSPs; and (ii) the use of the model to enable consumers to prioritize (using weights/utlity) a set of security protocols based on their particular use of cloud services.

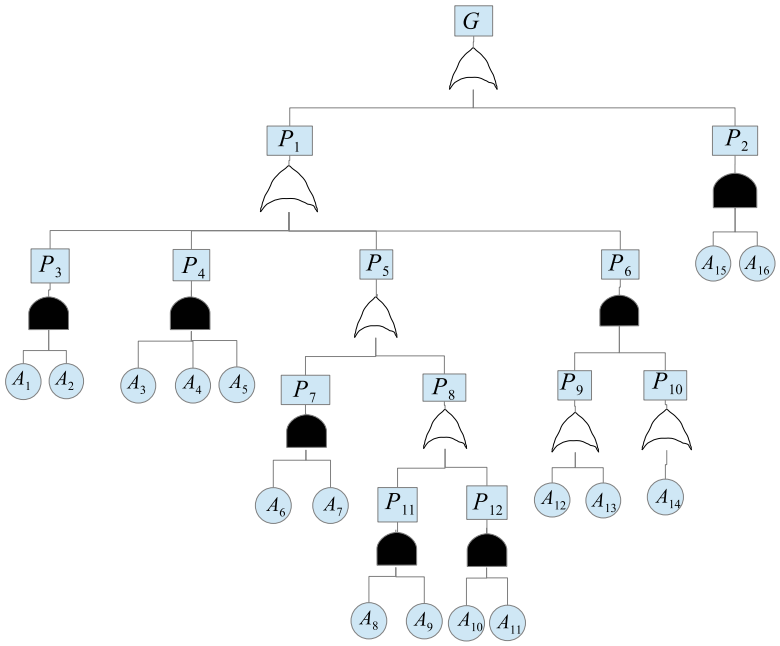
III. Tree based Framework

We make use of a tree-based technique, in which the event at the root of the tree can be referred to as the (attack) goal – i.e., the intended outcome from an attackers perspective. The attack goal considered in this paper is data breach in the cloud. This section focuses on identifying potential events that are likely to contribute to such a data breach. Events can be linked with an OR/AND relationship, where an OR-gate shows that the output event occurs only if one or more of the input events occur. The AND-gate shows that the output event occurs only if all the input events occur. If one of those events cannot be divided further, it is a leaf node of the underlying tree. Otherwise, those events are gate nodes that are treated as sub-goals separately and can be divided continually until all the events become leaf nodes. Whether an event can be further subdivided (i.e., an event may be a leaf node in an attack tree, but a gate node in another one) depends on the knowledge and experience of the security analyst who handles the attack tree. In order to build a tree, we identify and derive the causes for data breach into multiple types of events. At a high level, data breach in the cloud can broadly be either "active" or "passive". If an adversary compromises the user's data via malicious activities then this is classified as an active data breach. Conversely, if user data is lost due to reasons other than direct malicious activities of an adversary, then this can be classified as a passive data breach.

A. Active Data Breaches

Data in a cloud environment may go through the following stages: (i) reside in storage for direct access/use; (ii) be in transmission within and outside the cloud network; (iii) be subject to contract establishment/negotiation with the cloud provider; (iv) reside in non-production/use area – e.g., for backup, design and test purposes. The cloud provider must protect user data in all of these stages to avoid a data breach. Fig. 1 depicts this scenario and denotes the required protection mechanisms as P_3, P_4, P_5, and P_6. An adversary can try to compromise the data through a vulnerability in the protection mechanism. It is necessary for the cloud provider to consider several levels of sub-protection, as described in the following subsections.

1) P_3: During data production: Most production environments have established security and access restrictions to protect against data breaches. Standard security measures can be applied at the network level, the application level and the database level. Physical entry access controls can be extended by implementing multi-factor authentication schemes, such as key tokens or even biometrics. However, these protective measures cannot be simply replicated across every environment because the methods that protect data in production may not necessarily meet the unique requirements for protecting non-production environments.



G	Data Breach	P_7	physical protection	A_1	data in non production	A_9	key management
P_1	active data breach	P_8	virtual protection	A_2	data sanitization	A_{10}	secure server virtualization
P_2	passive data breach	P_9	physical destruction	A_3	authentication	A_{11}	multi-tenant architecture
P_3	during data production	P_{10}	crypto shredding	A_4	data encryption	A_{12}	overwrite
P_4	during data transmission	P_{11}	within host	A_5	network protection	A_{13}	use strong magnet
P_5	during data storage	P_{12}	between hosts	A_6	unauthorized access	A_{14}	delete keys
P_6	after data deletion			A_7	storage encryption	A_{15}	disaster management
				A_8	access control	A_{16}	retention policies

Fig. 1. Tree based model to access the of risk of data breach in cloud computing.

Non-production environments are used for design, development, and test activities internally within an organization. The risk of an unauthorized user getting access to the non-production environment is high. The user's data become vulnerable when it is moved to a non-production environment for testing purposes e.g., the user or service provider updates an application and uses real data to test the new functionality. This vulnerability can be mitigated or reduced if the cloud provider makes use of one of the following mechanisms: avoid the use of real user data for testing or anonymise the data before its use. As shown in Fig. 1, these two aspects constitute the leaf nodes of P_3.

2) P_4: During data transmission: User data may need to be moved between different entities/sites (e.g., different data center locations) over time. A user may access the data on a regular basis or third-party service providers may have permission to access user data. Data will also move within a cloud providers infrastructure, e.g., the user retrieves the data from a storage server and performs operation via a locally provisioned (i.e in the same data center) virtual machine. The cloud provider needs to authenticate users who request data access. During transmission, the data also needs to be encrypted. As shown in Fig. 1, authentication, data encryption, and network protection can be leaf nodes of P_4.

3) P_5: During data storage: The cloud provider may: control physical access (P_7) and/or digital access (P_8) to the data server. Limiting digital access is often complicated and requires further breakdown as follows: how to protect the data within a host machine (P_{11}), i.e., if an adversary and user are residing on the same host and protection is carried out between hosts (P_{12}), i.e., adversary and user are resident on different hosts. As shown in Fig. 1, leaf nodes of P_{11} are access control and key management techniques while those of P_{12} are secure server virtualization and focus on execution over a multi-tenant architecture.

4) P_6: After data deletion: We assume that data is considered completely destroyed when deleted from the drive, and they cannot be recovered by any means – thereby assuming that content discovery tools can no longer read data in the archive. We consider two data removal/ destruction techniques here although the analysis identified here could be generalised to other techniques. A popular techniques currently used is *Crypto Shredding*[b]. This methodology relies less on physical access to storage, but instead involves deliberate destruction

[b]http://tinyurl.com/oxl55y6

of all encryption keys for the data and the destruction of the encryption protocol itself. The keys are made unrecoverable by rotating the key for active storage and shredding it. It follows that archival data is also destroyed once the keys become unavailable. Another secure data destruction methodology is disk/free space wiping and physical destruction. This option is available if the cloud backup vendor enables a user to have limited access to the physical storage or includes this service as part of the procedure for management of the data drives. The software tool must be used to overwrite the data one to three times. *Degaussing* or the use of strong magnets is then used for scrambling data in hard drives so that data becomes unrecoverable. Complete destruction of the physical storage devices and shredding the magnetic media are also undertaken in some instances.

B. Passive Data Breaches

A passive data breach involves data loss due to natural (fire, earthquakes, flood) and/or man made (terrorism) disasters. Cloud providers usually implement several capabilities to reduce the risk of data loss. In general, data centers are built in clusters and located in various global regions. In case of failure, automated processes move user data traffic away from the affected area. Deploying a disaster recovery programme is also important to mitigate the risk of outages or data loss in the cloud. The basic requirement for data centers is to feature fire protection systems such as smoke detectors (passive protection) and fire sprinkler systems or a clean agent (active protection). To protect against earthquakes, the data center's racks need to be bolted down and use seismic restraints, and the facility must have multiple layers of redundancy. Emergency backup generators need to be made flood proof, and there needs to be enough fuel stored to last for days. A cloud provider using a varied set of protection mechanisms against diasters will have greater chance of reducing data loss. This is captured by leaf nodes A_{15} and A_{16} in Fig. 1. It is important that the data center operators clearly identifies the particular strategy they make use of.

IV. MATHEMATICAL MODEL

Fig. 1 shows the required protection mechanisms in the leaf nodes and how they are related. There are 16 leaf nodes, hence 16 technically different security protocols are required to stop the adversary to reach the root of the tree. Leaf and intermediate nodes are connection with AND-/OR-gates to reach the root. Consider a path which is composed of multiple leaf nodes as an attack scenario. It is obvious that the adversary's strategy could be to exploit the weakest scenario to compromise the data. Let us assign weights w_1, w_2, w_3, w_4, w_5, and w_6 to the intermediate nodes P_1, P_2, P_3, P_4, P_5, and P_6, respectively in Fig. 1. These weights will be used by clients to prioritize their interest. Section V provides the effect of these weights, suggesting how they will be used to combine the attack risk probabilities based on the following conditions:

$$w_1, w_2, w_3, w_4, w_5, w_6, \quad \geq \quad 0, \tag{1}$$

$$w_1 + w_2 \quad = \quad 1, \tag{2}$$

$$w_3 + w_4 + w_5 + w_6, \quad = \quad 1. \tag{3}$$

lower case letters represent attack risk probability at leaf or intermediate nodes and root node. For example, g denotes the risk probability that an attack reaches the root node G. Similarly, p_1 denotes the risk probability that an attack reaches the intermediate node P_1, while a_1 denotes the risk probability that an attack compromises the security protocol in leaf node A_1. Using OR/AND-gate we can combine the risk probabilities. The output of the attack risk probability of OR-gate is maximum of inputs while the output of AND-gate is the multiplication of inputs. According to these rules and $w > 0$ and $w_1 + w_2 = 1$, the attack risk probability for data breach is given by:

$$g = \mathsf{max}(w_1 p_1, w_2 p_2), \tag{4}$$

$$p_1 = \mathsf{max}(w_3 p_3, w_4 p_4, w_5 p_5, w_6 p_6) \tag{5}$$

$$p_2 = a_{15} a_{16} \tag{6}$$

$$p_3 = a_1 a_2 \tag{7}$$

$$p_4 = a_3 a_4 a_5 \tag{8}$$

$$p_5 = \mathsf{max}(p_7, p_8) = \mathsf{max}[a_6 a_7, \mathsf{max}(p_{11}, p_{12})], \\ = \mathsf{max}(a_6 a_7, a_8 a_9, a_{10} a_{11}), \tag{9}$$

$$p_6 = p_9 p_{10} = \mathsf{max}(a_{12} a_{14}, a_{13} a_{14}), \tag{10}$$

$$p_8 = \mathsf{max}(p_{11}, p_{12}) = \mathsf{max}(a_8 a_9, a_{10} a_{11}). \tag{11}$$

From (4)–(11), g can be obtained in terms of $a_i \ \forall i$ as follows:

$$\begin{aligned} g &= \mathsf{max}(w_1 w_3 p_3, w_1 w_4 p_4, w_1 w_5 p_5, w_1 w_6 p_6 w_2 a_{15} a_{16}), \\ &= \mathsf{max}(w_1 w_3 a_1 a_2, w_1 w_4 a_3 a_4 a_5, w_1 w_5 a_6 a_7, w_1 w_5 a_8 a_9, \\ &\qquad w_1 w_5 a_{10} a_{11}, w_1 w_6 p_9 p_{10}, w_2 a_{15} a_{16}), \\ &= \mathsf{max}(w_1 w_3 a_1 a_2, w_1 w_4 a_3 a_4 a_5, w_1 w_5 a_6 a_7, w_1 w_5 a_8 a_9, \\ &\qquad w_1 w_5 a_{10} a_{11}, w_1 w_6 a_{12} a_{14}, w_1 w_6 a_{13} a_{14}, w_2 a_{15} a_{16}). \end{aligned} \tag{12}$$

TABLE I
PROBABILITIES FOR POSSIBLE DATA BREACH SCENARIOS.

Data Breach Scenarios	Leaf nodes	Probability
S_1	A_1, A_2	$w_1 w_3 a_1 a_2$
S_2	A_3, A_4, A_5	$w_1 w_4 a_3 a_4 a_5$
S_3	A_6, A_7	$w_1 w_5 a_6 a_7$
S_4	A_8, A_9	$w_1 w_5 a_8 a_9$
S_5	A_{10}, A_{11}	$w_1 w_5 a_{10} a_{11}$
S_6	A_{12}, A_{14}	$w_1 w_6 a_{12} a_{14}$
S_7	A_{13}, A_{14}	$w_1 w_6 a_{13} a_{14}$
S_8	A_{15}, A_{16}	$w_2 a_{15} a_{16}$

From (12) we observe eight possible data breach scenarios, and Table IV illustrates these attack scenarios at the corresponding leaf nodes and identifies their associated risk probabilities. Consider N cloud providers with the attack risk probability associated with the nth provider being g^n where

$n = 1, \ldots, N$. A consumer must therefore chose a provider that offers the minimum of the maximum attack risk. This is given by

$$RiskLessCloud = \underset{n}{\textit{argmin}} \quad \{\mathtt{max}(g^n)\}. \tag{13}$$

More than one leaf node must be compromised to reach the root node of the tree. To compromise the leaf node many aspects such as the possibility to succeed, attack cost, difficulty of the required technique, risk of being detected, and so on have to be considered. In this paper, we calculate the attack risk probability by assigning leaf nodes three attributes: attack cost, technical difficulty and the probability to be discovered. In Table II, we divide each attribute into one of five levels. However, we use different non-linear utility functions (shown later) to convert these linear gradings into more practical non-linear gradings.

TABLE II
PROBABILITIES FOR POSSIBLE DATA BREACH SCENARIOS.

Cost to break		Difficulty		Discovery	
Cost (1000)	Grade	Difficulty	Grade	Probability of Detection	Grade
> 10	5	quite difficult	5	quite difficult	1
$6 - 10$	4	difficult	4	difficult	2
$3 - 6$	3	medium	3	medium	3
$0.5 - 3$	2	simple	2	simple	4
< 0.5	1	quite simple	1	quite simple	5

A cloud provider can offer different types of security techniques for each leaf node e.g., for leaf node A_1, a cloud provider can either allow or disallow the data to be made available in a non-production environment. Similarly for leaf node A_2, different anonymization techniques such as k-anonymity, l-diversity, t-closeness, and differential privacy could be used by the cloud provider. A great deal of efforts has gone into defining and quantifying techniques for each leaf node in literature. Several standardization bodies such as Cloud Security Alliance, Cloud Standard Customer Council, ENISA, IEEE Cloud Computing Standard Study Group (IEEE CCSSG), ITU Cloud Computing Focus Group, Distributed Management Task Force (DMTF), Storage Networking Industry Association (SNIA), Open Grid Forum (OGF), Open Cloud Consortium (OCC), and Organization for the Advancement of Structured Information Standards (OASIS) are working on standardizing security SLAs for cloud. In this paper, we have combined their efforts in Table III where we categorize these techniques into two directions: vertically we assigned each technique into one of the leaf nodes while horizontally we assign gradings in ascending order.

We emphasize here that the contribution of this paper is to come up with a model to evaluate the overall risk in a CSP and the potential risk for a data breach. The tree model proposed in Fig. 1 will clearly need to evolve over time as the number of leaf nodes and the number of attack scenarios increase. Hence the details on Table III will be adapted accordingly. Our emphasis here is on the proposed methodology, which can be generalised through the identification of additional potential risks and mitigation strategies. After associating a technique with a leaf node, multi-attribute utility theory can be adopted to aggregate attack risk probability [27]. Accordingly to such a multi-attribute utility, let us define the weight and utility function for each attribute. Let z_i denote the weights and $u_i[A_l(x)]$ denote utility functions, where $i = 1, 2, 3$, $l = 1, \ldots, 16$; and let x represent grades associated with the leaf node A_l. Hence the aggregated attack risk probability is given by:

$$\begin{aligned} a_l \quad = \quad & z_1 \times u_1[A_l(x)] + z_2 \times u_2[A_l(x)] + z_3 \times u_3[A_l(x)], \\ & l = 1, \ldots, 16, x \in \{1, 2, 3, 4, 5\}, \end{aligned} \tag{14}$$

where

$$z_1 + z_2 + z_3 = 1. \tag{15}$$

If there are N cloud providers then each of them offers one of the techniques from Table III for each leaf node. A client can obtain gradings for all three attributes using Table II and calculate the attack risk probability for each leaf node using (14). Risk probability for a data breach is then calculated using (4). This procedure will be repeated for all N cloud providers using the same values of z_i and w_i. A cloud which is then less vulnerable to a potential data breach can be selected using (13).

V. NUMERICAL ANALYSIS

In this section we evaluate the proposed model by considering different values for the weights – based on a priority set by a user. Let us introduce utility functions which map the linear gradings given in the Table II into a non-linear domain. Utility functions can be defined in many ways but in this paper we consider the following three functions for attributes cost, technical difficulty, and detection, respectively (and represent risk probability associated with a provider):

$$u_1(\alpha) = 10^{1-\alpha}; u_2(\alpha) = \alpha^{-1}; u_3(\alpha) = 10^{1-\alpha} \tag{16}$$

where $\alpha \in \{1, 2, 3, 4, 5\}$. Fig. 2 shows the relation between three functions for different gradings. From Fig. 2, the risks associated with attributes cost, detection, and technical difficulty are monotonically increasing for the same grading, i.e., $u_1(2) < u_3(2) < u_2(2)$.

Let us consider an example scenario with five different cloud service providers (CSPs) to evaluate the model. The gradings (based on technical difficulty) for each leaf node for each CSP are provided in Table IV. CSP 1 offers the worst security SLA in terms of technical difficulties while CSP 5 offers the best security SLA. The other three CSPs offer intermediate security SLAs.

If we consider only the technical difficulty attribute for leaf nodes, then $z_1 = 1, z_2 = 0$, and $z_3 = 0$. For this simple case let us simulate the attack risk probability by assigning equal values for other weights, i.e., $w_i = 1, \ \forall i$. Fig. 3 shows the risk probabilities for all attack scenarios. Risk probabilities for CSP 1 is high in all attack scenarios while the risk reduces gradually for all other CSPs. For CSP 2, the risk associated

TABLE III

DIFFERENT OPTIONS AVAILABLE FOR A CLOUD PROVIDER. THE GRADES IN THE TABLE REPRESENT THE RELATIVE TECHNICAL DIFFICULTY ($1-$ EASY TO BREAK AND $5-$VERY DIFFICULT TO BREAK).

Grade	1	2	3	4	5
A_1	Yes				No
A_2	$k-$anonimity	$l-$diversity	$t-$closeness		differential privacy
A_3	service provider assertion	authentication federation	user name and TLS client certificate	user name and two factor authentication	limited access over dedicated link
A_4	legacy SSL and TLS - $128-$bits	TLS (Version 1.2 or above)- $128-$bits		legacy SSL and TLS - $256-$bits	TLS (Version 1.2 or above)- $256-$bits
A_5	community WAN service	encrypted community WAN service	IPsec VPN gateway	bonded fibre optic connections	bonded fibre optic connections with TLS
A_6	video surveillance		two-factor biometric authentication		monthly access reviews
A_7	only physical protection		keys are with cloud provider	keys are with third-party server	keys are with client
A_8	coarse-grained role-based access control	fine-grained role-based access control	attribute-based access control (ABAC)	ABAC and community network	ABAC and private network
A_9	keys are managed by cloud provider		keys are managed by third-party		keys are managed by users
A_{10}	application virtualization	desktop virtualization	user virtualization	storage virtualization	hardware virtualization
A_{11}	database-based segmentation		hypervisor-based segmentation	VM Introspection	penetration test certificates
A_{12}	reuse without caution	overwrite multiple times	free space wiping and physical destruction		pass content discovery test
A_{13}	No		Yes		pass content discovery test
A_{14}	no key-rotation				delete keys
A_{15}	no protection	fire protection	fire and water protection	fire, water, and earthquake protection	fire, water, earthquake, and anti-terrorism protection
A_{16}	backup at same location	backup at multiple locations	backup at multiple locations everyday	backup at multiple locations every hour	instant backup at multiple locations

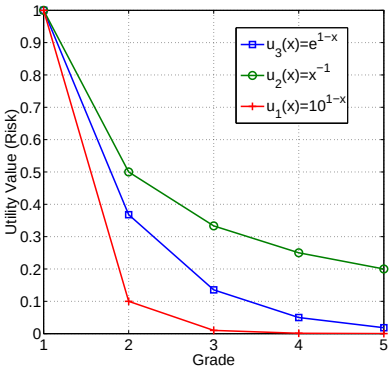


Fig. 2. Risk values for various utility functions for grades change between 1–5.

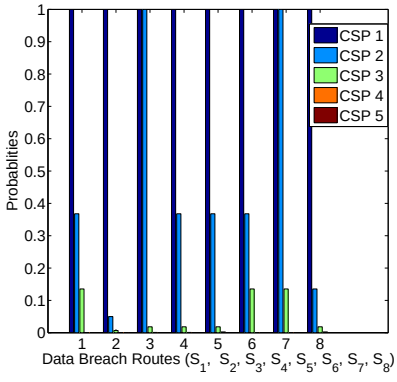


Fig. 3. Risk probabilities for attack scenarios when only one attribute (technical difficulty) considered for leaf nodes.

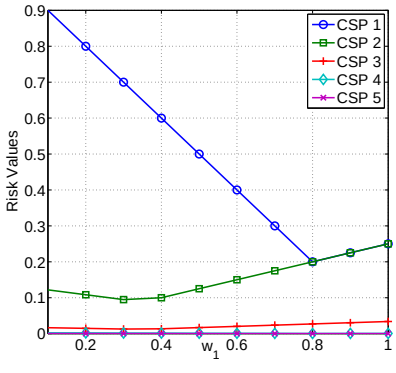
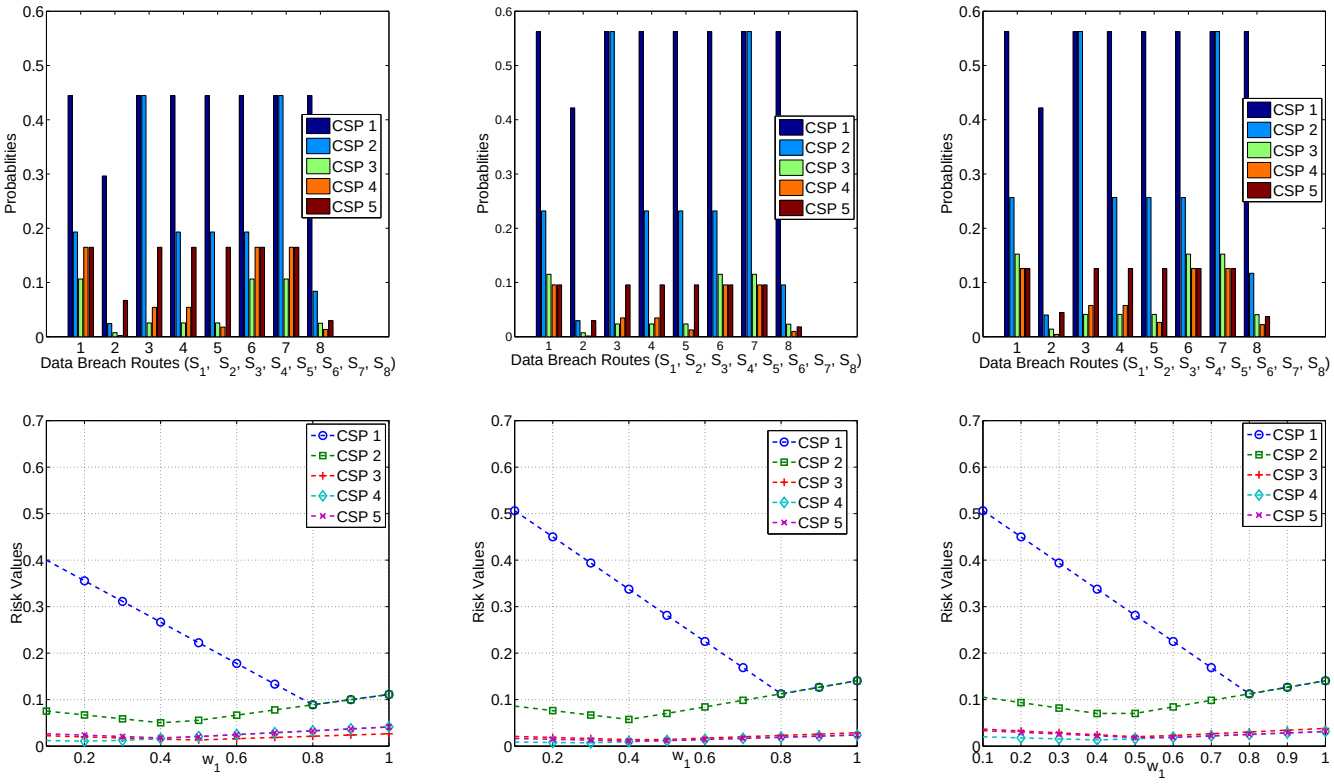


Fig. 4. Attack risk probability for different weights.

with scenarios 3 and 4 is high. Hence, according to Table IV, the attacker will exploit leaf nodes 6 and 7, or leaf nodes 13 and 14 to compromise the data via scenario 4 or scenario 7, respectively.

Let us now evaluate the effect of weights w_i for the cases considered above. Fig. 4 shows the attack risk probabilities when w_1 varies between 0 and 1, w_2 varies between 1 and 0, while $w_3 = w_4 = w_5 = w_6 = 0.25$, based on the preferences



(a) $z_1 = 1/3, z_2 = 1/3, z_3 = 1/3$ (b) $z_1 = 1/2, z_2 = 1/4, z_3 = 1/4$ (c) $z_1 = 1/4, z_2 = 1/2, z_3 = 1/4$

Fig. 5. Risk probability (without scaling) values for five CSPs against the eight data breach routes shown in the Table II.

TABLE IV
FIVE DIFFERENT CSPS HAVE BEEN CHOSEN TO EVALUATE THE PROPOSED MODEL. GRADE VALUES HAVE BEEN CHOSEN FROM TABLE III.

	CSP 1	CSP 2	CSP 3	CSP 4	CSP 5
A_1	1	1	1	5	5
A_2	1	2	3	5	5
A_3	1	2	3	4	5
A_4	1	2	2	4	5
A_5	1	2	3	4	5
A_6	1	1	3	5	5
A_7	1	1	3	4	5
A_8	1	2	3	4	5
A_9	1	1	3	5	5
A_{10}	1	2	3	4	5
A_{11}	1	1	3	4	5
A_{12}	1	2	3	5	5
A_{13}	1	1	3	5	5
A_{14}	1	1	1	5	5
A_{15}	1	2	3	4	5
A_{16}	1	2	3	4	5

identified by a user. Fig. 4 clearly demonstrates that the risk values are constantly changing. When $w_1 > 0.8$ (i.e. high priority for active attack and data loss due to passive attack being unimportant for the client) then the risks associated with CSP 1 and CSP 2 are the same. In this case, the client can choose CSP 2 since they need to pay less for CSP 2 than CSP 1 for the same level of risk. However, the risk associated with CSP 1, CSP 2, and CSP 3 is much lower than CSP 4 and CSP 5.

The risk probability will be different if we consider a greater number of attributes for each leaf node. Let us now simulate the proposed model by considering all three attributes, with w_i being a constant and $0 < z_i < 1$. Fig. 5 shows the risk probabilities for three different combinations of z_i. The first row of Fig. 5 shows the individual probability for each attack scenario while the second row shows the data breach risk (i.e., maximum value out of eight scenarios). In Table II, it is assumed that the probability of discovering an attack on a highly secure protocol is low [28]. CSP 4 and CSP 5 provide highly secure protocols compared to other CSPs. Hence when z_3 is high (i.e., high priority given for detecting an attack), the risk associated with CSPs 4 and 5 are high compared to others. This clearly shows that the risk cannot be measured based on a single attribute.

From the second row in Fig. 5, the risk of data breach associated with CSPs 1 and 2 are always higher than others CSPs. It clearly shows that a security protocol which is difficult to compromise is always important for a cloud provider. Both CSPs 4 and 5 use such protocols, however, CSP 5 has greater protection than that of CSP 4 in terms of technical difficulty in compromising data (see Table IV). Even though the risks associated with CSPs 4 and 5 are almost the same and much

lower than other CSPs, in some instances the risk associated with CSP 5 is higher than CSP 4 (see second row of Fig. 5). The proposed model can be used by a user/ client to choose the right CSP based on their requirement (i.e. by consideration of different weights to obtain a model similar to Fig. 5).

VI. Conclusion

We propose a risk model to characterise data breach for cloud service providers, enabling users to identify parameters of most concern to them and provide a weighting function to prioritise them. Multi-attribute utility theory is used to aggregate risk probabilities across the measured parameters. This model enable a user/ client to choose a cloud provider who offers lower risk compared to others based on the same criteria. More importantly, our model enables a client to prioritize security parameters that are most important for their application by assigning different weights. We use numerical analysis to demonstrate that the data breach risk will be varying based on security SLAs as well as based on clients requirements. Our model can be extended as additional monitoring tools become available, enabling extension of the attack tree parameters considered in this work. The analysis proposed here also enables a better comparison to be carried out across multiple cloud providers, enabling an objective assessment by a user.

References

[1] P. S. Pawar, M. Rajarajan, S. K. Nair, and A. Zisman: Trust model for optimized cloud services. In *Trust Management VI (pp. 97–112), 2012, Springer* Berlin Heidelberg.

[2] Y. Rahulamathavan, P. S. Pawar, P. Burnap, M. Rajarajan, O. Rana, and G Spanoudakis, Analysing Security requirements in Cloudbased Service Level Agreements, In Proc *The 7th International Conference on Security of Information and Networks (SIN'14)*, Sept. 2014, Glasgow, UK.

[3] Y. Rahulamathavan, R. C. W. Phan, S. Veluru, K. Cumanan, and M. Rajarajan: Privacy-Preserving Multi-Class Support Vector Machine for Outsourcing the Data Classification in Cloud. In *IEEE Trans. Dependable and Secure Computing*, vol. 11, no. 5, pp. 467– 479 . (2014)

[4] Cloud Security Alliance Cloud Controls Matrix (CCM) version 3.0 @ ONLINE http://www.cloudsecurityalliance.org. 2013.

[5] P. R. Barbosa, R. R. Righi, and D. L. Kreutz. Defining metrics to Sec-SLA agreements in conformance to international security standards. In *Latin American Informatics Conference, San Jos, Costa Rica*, pp. 36–47, Jun. 2007.

[6] K. Bernsmed, M. G. Jaatun, P. H. Meland, and A. Undheim. Thunder in the clouds: Security challenges and solutions for federated clouds. In *IEEE Fourth Intl Conf. Cloud Computing Technology and Science (CloudCom)*, pp. 113–120, Dec. 2012.

[7] N. S. Chauhan, A. Saxena, and J. V. R. Murthy. An approach to measure security of cloud hosted application. In *IEEE Intl Conf. Cloud Computing in Emerging Markets (CCEM)*, pp. 1–6, Oct. 2013.

[8] S. A. D. Chaves, C. B. Westphall, and F. R. Lamin. SLA perspective in security management for cloud computing. In *Sixth Intl Conf. Netw. Services (ICNS)*, pp. 212–217, Mar. 2010.

[9] F. Li, Y. Rahulamathavan, and M. Rajarajan. LSD-ABAC: Lightweight static and dynamic attributes based access control scheme for secure data access in mobile environment. In *Proc. 39th Annual IEEE Conf. Local Computer Networks (LCN)*, Sep. 2014.

[10] F. Li, Y. Rahulamathavan, M. Rajarajan, and R. C.-W. Phan. Low complexity multi-authority attribute based encryption scheme for mobile cloud computing. In *Proc. IEEE 7th Intl Symp. Service Oriented System Engineering (SOSE)*, pp. 573–577, Mar. 2013.

[11] A. Mantelero. Cloud computing, trans-border data flows and the european directive 95/46/ec: applicable law and task distribution. *European Journal of Law and Technology*, 3(2), 2012.

[12] P. S. Pawar, M. Rajarajan, T. Dimitrakos, and A. Zisman. Trust model for cloud based on cloud characteristics. In *IFIP Advances in Information and Communication Technology*, pp. 239–246, 2013.

[13] Y. Rahulamathavan, V. Moonsamy, L. Batten, S. Shunliang, and M. Rajarajan. An analysis of tracking service settings in blackberry 10 and windows phone 8 smartphones. In *Proc. 19th Australasian Conference on Information Security and Privacy (ACISP)*, Jul. 2014.

[14] M. Rak, L. Liccardo, and R. Aversa. A SLA-based interface for security management in cloud and GRID integrations. In *Seventh Intl Conf. Information Assurance and Security (IAS)*, pp. 378–383, Dec. 2011.

[15] M. Rak, N. Suri, J. Luna, D. Petcu, V. Casola, and U. Villano. Security as a service using an SLA-based approach via SPECS. In *IEEE Fifth Intl Conf. Cloud Computing Technology and Science (CloudCom)*, volume 2, pp. 1–6, Dec. 2013.

[16] R. R. Righi, D. L. Kreutz, and C. B. Westphall. SEC-MON: An architecture for monitoring and controlling security service level agreements. In *XI Workshop on Managing and Operating Networks and Services, SBC Press, Porto Alegre*, pp. 73–84, 2006.

[17] R. R. Righi, F. Pelissari, and C. Westphall. SEC-SLA: Specification and validation of metrics to security service level agreements. In *IV Workshop on Computer System Security, SBC Press, Porto Alegre*, pp. 199–210, 2004.

[18] C. Rong, S. T. Nguyen, and M. G. Jaatun. Beyond lightning: A survey on security challenges in cloud computing. *Computers & Electrical Engineering*, 39(1): pp. 47 – 54, 2013.

[19] S. Subashini and V. Kavitha. A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1): pp. 1 – 11, 2011.

[20] Q. Zhang, L. Cheng, and R. Boutaba. Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1): pp. 7–18, 2010.

[21] J. Zhengwei, D. Ran, L. Zhigang, W. Xihong, and L. Baoxu. A meta-synthesis approach for cloud service provider selection based on secsla. In *Fifth Intl Conf. Computational and Information Sciences (ICCIS)*, pp. 1356–1360, Jun. 2013.

[22] G. Zhien and D. Yiqi. Security SLAs for IMS-based cloud services. In *Seventh ChinaGrid Annual Conference (ChinaGrid)*, pp. 57–60, Sep. 2012.

[23] J. Weis, and J. Alves-Foss. Securing Database as a Service: Issues and Compromises, In *IEEE Security & Privacy*, vol. 9, no. 6, pp. 49–55, Nov.-Dec. 2011.

[24] D. Martin. Implementing Effective Controls in a Mobile, Agile, Cloud-Enabled Enterprise, In *IEEE Security & Privacy*, vol. 11, no. 1, pp. 13–14, Jan.-Feb. 2013.

[25] L. M. Kaufman. Can a Trusted Environment Provide Security?, In *IEEE Security & Privacy*, vol. 8, no. 1, pp. 50–52, Jan.-Feb. 2010.

[26] N. Ghosh, S. K. Ghosh, S. K. Das. SelCSP: A Framework to Facilitate Selection of Cloud Service Providers, In *IEEE Trans. Cloud Computing*, vol. 3, no. 1, pp. 66–79, Jan.-March. 2015.

[27] R. A. Kemmerer: Covert Flow Trees: A Visual Approach to Analyzing Covert Storage Channels, In *IEEE Transaction on Software Engineering*, vol. 11, no.17, pp. 1166–1185, 1991.

[28] D. Ren, S. Du, and H. Zhu: A Novel Attack Tree Based Risk Assessment Approach For Location Privacy Preservation In The VANETs, In *IEEE IEEE Int'l Conf. Communications*,, June, 2011.

[29] U. Onwudebelu, and B. Chukuka: Will Adoption Of Cloud Computing Put The Enterprise at Risk? In *IEEE 4th Int'l Conf. Adaptive Science & Technology (ICAST)*, pp. 82–85, 25-27 Oct. 2012

[30] T. Grandison and M. Sloman: A Survey Of Trust In Internet Applications, In *IEEE Commun. Surv. Tutorials*, vol. 3, no. 4, pp. 2-16, Fourth Quarter 2000.

[31] A. Jsang, R. Ismail, and C. Boyd: A Survey Of Trust And Reputation Systems For Online Service Provision, *Decision Support Sys.*, vol. 43, no. 2, pp. 618-644, Mar. 2007.

[32] B. Schneier: Attack Trees, *Dr. Dobbs journal* vol. 24, no. 12, pp. 21–29, 1999.

[33] M. Frydman, G. Ruiz, E. Heymann, E. Csar, and B. P. Miller: Automating Risk Analysis of Software Design Models. In *The Scientific World Journal*, 2014.

[34] K. H. Chang. Security Threat Assessment Of An Internet Security System Using Attack Tree And Vague Sets. In *The Scientific World Journal*, 2014.

[35] M. Tentilucci, N. Roberts, S. Kandari, D. Johnson, D. Bogaard, B. Stackpole, and G. Markowsky: Crowdsourcing Computer Security Attack Trees. In *10th Annual Symposium on Information Assurance*, June, 2015.