

2 Foundations

In this chapter, we present the theoretical foundations of this thesis. First, we introduce the paradigm of Cloud Computing, along with its core technological concepts and terminology (Section 2.1). Then, we present the work related to the nature of ITSR perception and perceived ITSRs in the context of Cloud Computing (Section 2.2). Finally, the organizational IT security risk management and its five key phases are described (Section 2.3).

2.1 Cloud Computing

Cloud Computing (the Cloud) has recently emerged as a new paradigm for the provision of computing resources, which has the potential to transform large parts of the IT industry (Armbrust et al. 2010; Smith 2014). It is based on the principle that large groups of remote servers are networked to allow centralized access to computing resources, such as data storage, central processing unit (CPU) capacity, and software applications. However, the variety of technologies in the Cloud makes the overall picture confusing (Vaquero et al. 2008). Besides, there is a number of inconsistent definitions of the Cloud (e.g., Armbrust et al. 2010; Buyya et al. 2008; Vaquero et al. 2008; Wang et al. 2008b).

The definition provided by the U.S. National Institute of Standards and Technology (NIST) is most frequently used in IS research and practice. It has also been adopted by the European Network and Information Security Agency (ENISA). NIST universally defines the Cloud as a model “for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (Mell and Grance 2011, p. 2).

By shifting the location of the infrastructure from the users’ side to a central provider, the paradigm reduces the costs associated with the deployment and management of hardware and software resources (Vaquero et al. 2008). In this regard, the development of the Cloud is closely related to the advancements in virtualization technologies, which enables the allocation of IT-based resources to worldwide distributed computers (Buyya et al. 2009). Therefore, the providers have the economic advantage that they can use available resources more effectively and realize provider-sided economies of scale (Wang et al. 2008c). However, the Cloud is also attractive to customer companies as it eliminates the requirement to plan ahead for provisioning, and allows organizations to start from the small and increase computing resources only when there is a rise in capacity demand (Zhang et al. 2010). Hence, it is not surprising that leading market analysts forecast that the Cloud will grab a large share of the overall IT market in the next years (e.g., Gartner 2013). Beside the large providers such as Amazon, Google, and Microsoft, an increasing number of organizations tap into the Cloud market.

Today, industry databases already list thousands of different services (Cloud Showplace 2014).

In general, the “cloud model is composed of five essential characteristics, three service models, and four deployment models” (Mell and Grance 2011, p. 2).

2.1.1 Essential Characteristics

The essential characteristics are:

On-demand self-service: The Cloud technology enables customers to request resources, such as computing and storage capacities, independently from the service provider, so that they can be efficiently adjusted to their current needs without having to depend on interaction with the provider (Mell and Grance 2011). In particular, the use of service management software typically allows customers to scale the provided resources with minimal provider interaction (Marston et al. 2011).

Broad network access: The Cloud model is based on network-access of the provided resources and services in real-time, utilizing standard technologies and mechanism, like the Internet and web interfaces (Mell and Grance 2011). The accessibility by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations) makes the Cloud universally usable for customers (Weinhardt et al. 2009).

Resource pooling: By utilizing a multi-tenant model, the providers’ computing resources, such as storage, processing, memory, and network bandwidth, are pooled in order to enable parallel use by multiple customers. The different physical and virtual resources dynamically assigned to fulfill current consumer demands (Mell and Grance 2011). In particular, the multi-tenant model allows customers to share the same provider-sided hardware resources by offering them a shared application and database instance, which can be configured to their needs as if it runs on a dedicated environment (Bezemer and Zaidman 2010). In this context, a specific characteristic of the Cloud is that the user has neither knowledge nor control over the actual location of the service provisions. However, customers are often able to specify the location of the data processing at a higher level of abstraction (e.g., country, state, or datacenter) (Takabi et al. 2010).

Rapid elasticity: The Cloud technology leverages an elastic provision and release of computing resources, so that provided services can be dynamically reconfigured and quickly scaled to changed customer demands (Mell and Grance 2011). The capability available for provisioning oftentimes create the impression that infinite resources are accessible at any time (Mather et al. 2009; Owens 2010).

Measured service: The Cloud model allows for an automatic control and optimization of the resource use by offering metering (e.g., pay-per-use or charge-per-use) capabilities at some

level of abstraction that is appropriate to the type of provided service (e.g., used storage, processing time, bandwidth, or active user accounts) (Mell and Grance 2011). In this context, the exact monitoring, controlling, and reporting of the resource usage creates transparency for customer companies as well as providers (Fouquet et al. 2009; Weinhardt et al. 2009).

2.1.2 Delivery Models

According to the NIST definition, the Cloud is structured into three consecutive layers, representing different service delivery models (Mell and Grance 2011):

Infrastructure as a Service (IaaS): The resources provided to the customers on the IaaS layer are computing infrastructure, such as storage, networking and processing capabilities (Mell and Grance 2011). In particular, the providers allow the customers to deploy and run arbitrary software – ranging from operating systems to individual software – on their infrastructure by offering a virtual instance of the physical hardware. This virtualization enables the providers to automatically split and dynamically assign their computing resources to customers (Prodan and Ostermann 2009; Vaquero et al. 2010). In this regard, even if the customers do not have direct control of the computing infrastructure, they are typically able to manage and control selected network components (e.g., firewall of the host), operating systems, and storage devices along with the deployed applications (Mell and Grance 2011). However, due to the virtualization of the underlying computing resources, the customers have typically no knowledge about the actual location of the servers and the data processing (Takabi et al. 2010). Widely used examples for IaaS services are Amazon’s EC2, GoGrid’s Cloud Server, SAP HANA Enterprise Cloud, and Microsoft’s Windows Azure (Buyya et al. 2008).

Platform as a service (PaaS): The service provided to the customers on the PaaS layer is the deployment of customer-built or acquired applications onto the Cloud infrastructure (Mell and Grance 2011). In this context, the providers typically offer development tools and application programming interfaces (APIs), which enable the customers to interact with the execution environment and to deploy applications by using the provided programming environments (i.e., programming languages, libraries, services, and tools) (Lenk et al. 2009; Youseff et al. 2008). In particular, due to the ubiquitous network access of the resources, the PaaS layer allows customers to efficiently deploy their own web applications and services (Prodan and Ostermann 2009). When utilizing PaaS services the consumers do not need to invest in the computing infrastructure (e.g., network, servers, operating systems, or storage) nor manage its complexity but have the control over the developed applications and can usually configure the execution environment to their individual needs (Marston et al. 2011). Cloud services on the PaaS layer are typically provided based on Cloud infrastructure services on the IaaS layer (Foster et al. 2008). Examples of established PaaS services include Amazon’s Elastic Bean-

stalk and Simple DB, Google's App Engine, SAP HANA Cloud Platform, and Salesforce's application development platform Force.com (Buyya et al. 2008).

Software as a Service (SaaS): The capability provided on the SaaS layer is the usage of standard software solutions running on the providers' Cloud infrastructure (Mell and Grance 2011). In contrast to traditional software installations, SaaS services typically utilize multi-tenancy architectures, which allow creating multiple instances of software, running on a single (virtual) server. These instances can be used by multiple customers at the same time (Mather et al. 2009). In this regard, services on the SaaS layer are in many cases provisioned based on services on the PaaS or IaaS layer (Prodan and Ostermann 2009). Moreover, SaaS services in general eliminate the need to install the software on the client devices as the provided application is typically accessible over the Internet through a thin client interface (e.g., a web browser) or a dedicated program interface. Accordingly, the applications can be used by the customers from a wide variety of client devices ranging from mobile phones to workstations (Marston et al. 2011). When using SaaS services, the customers do not have to take care of managing the operation and maintenance of the underlying computing infrastructure, such as networks, servers, operating systems, or application resources. However, customers are usually, to a certain extent, able to configure the SaaS services according to their specific needs (Lenk et al. 2009). In comparison to traditional software delivery models, SaaS solutions offer many possibilities for flexible pricing strategies (e.g., pay-per-use approaches), which, for example, allow small and medium-sized businesses to benefit from efficient enterprise resource planning (ERP) software (Al-Roomi et al. 2013). Examples of SaaS services are Apple's iWork, Google's Apps, Microsoft's CRM Online, and SAP BusinessByDesign, or salesforce.com.

Other definitions of delivery models in the context of the Cloud range from specific technology-driven views (e.g., Communications as a Service (CaaS)) to broad business model perspectives (e.g., Everything as a Service (XaaS)) (e.g., Armbrust et al. 2010; Youseff et al. 2008).

2.1.3 Deployment Models

Following Mell and Grance (2011)'s definition, Cloud services can be provisioned using four different deployment models:

Private Cloud: The Cloud infrastructure of a Private Cloud is operated exclusively for a single organization, serving multiple business units as internal customers. Private Clouds are owned and managed by the organization or an external provider, and hosted either internally or externally (Mell and Grance 2011). Similar to the internal data processing centers, the Private Cloud is only accessible by well-defined user groups (e.g., the organization's business units or partner organizations). Accordingly, Private Clouds are frequently said to entail a

higher level of IT security than Public Clouds (e.g., Armbrust et al. 2010; Buyya et al. 2008). Undertaking a Private Cloud project usually requires long-term planning and considerable efforts virtualize the existing infrastructure and business environment but also offers various opportunities to reevaluate and thereby improve current business processes (Takabi et al. 2010). However, regardless whether Private Clouds are hosted internally or not, the cost-intensive computing infrastructure is provided solely for a very limited group of users, so that the model usually does not enable organizations to realize significant economies of scale (Armbrust et al. 2010).

Public Cloud: The Cloud infrastructure of a Public Cloud is rendered over a network that is open for public use (i.e., the Internet). Public Cloud are generally owned, managed, and operated by an external provider. The computing infrastructure is located in data processing centers on the provider premises (Mell and Grance 2011). In general, the Public Cloud enables customers to quickly use and release service capabilities without the need of complex contract negotiations with the provider organizations (Armbrust et al. 2010). Moreover, services in the Public Cloud are in many cases free to use or offer flexible pricing strategies, like pay-per-use models (Al-Roomi et al. 2013). As the Cloud infrastructure is accessible through a public network, the potential ITSR exposure of these services is generally significantly higher than that of Private Cloud models. In particular, the customers have typically no direct control and influence on the safeguarding measures, which are implemented in the Cloud services by the providers (Ackermann et al. 2012). However, efficient sharing (e.g., through virtualization technologies) of the computing infrastructure with many organizations resulting in higher load factors, create in many cases considerable economies of scales, which may significantly reduce the costs of the provided service capabilities (Armbrust et al. 2010).

Community Cloud: The Cloud infrastructure of Community Clouds is provided for the exclusive use of a specific group of customers with similar requirements, i.e., organizations with the same goals, legal frameworks, security concerns, policies, and/or compliance considerations. Community Clouds are owned, managed, and operated on or off premises by one or more of the organizations in the group or by an external provider (Mell and Grance 2011). Since the Community Cloud is only accessible by members of the group through a private or secured network, the level of ITSR is oftentimes considered as lower than in Public Clouds (Zhang et al. 2010). However, as organizations with similar goals in the market are in many cases also potential competitors, there may be additional security and compliance risks associated with the use of Community Clouds (Takabi et al. 2010). Nevertheless, as the members of the group concentrate their resources and share the same infrastructure, the costs of the IT capability provision are oftentimes considerably lower, in comparison with the internal data processing centers of the organizations (Armbrust et al. 2010).

Hybrid Cloud: The Cloud infrastructure of a Hybrid Cloud is composed of the Cloud infrastructures provided by two or more distinct delivery models (i.e., private, public, or community Clouds), which are bound together by standardized or proprietary technology that enables data and application portability (Mell and Grance 2011). This combination enables to combine the benefits of the Private Cloud with those of the Public Cloud, for example, by offering data isolation, high availability, and load balancing at the same time (Buyya et al. 2009). For instance, a Private Cloud service is generally less cost-intensive when the underlying computing infrastructure is only designed for the typical workload of an organization. When the service is operated as Hybrid Cloud, it additionally uses infrastructure of external Public Cloud services to economically handle peaks in demand. However, when data is temporarily swapped to Public Cloud services, it may be exposed to a higher level of ITSR (Takabi et al. 2010).

2.2 IT Security Risk Perception

In economic research, the risk exposure is typically defined as the product of the probability of the occurrence of a negative event and the amount of the losses caused by this event (Boehm 1991; Cunningham 1967; Erb et al. 1996). The definition is primarily focused on the possible damage or potential loss of an investment and does not take potential profits into account. As such, it is categorized as a shortfall-oriented view of risk. The definition is often used by IS research when the focus is on organizational and behavioral aspects of IT security (e.g., Johnston and Warkentin 2010; Liang and Xue 2010; Vance et al. 2012). Another shortfall-oriented risk measure is the Value at risk, which is widely used in financial mathematics and financial risk management to measure the risk of loss in a specific portfolio of financial assets. Related to economic decision-theory, it is based on the knowledge of probabilities and probabilistic distributions of uncertain events in the future. The risk is considered to be higher when the uncertainty of an expected value increases, independent of whether the deviation is positive or negative. The deviation from an expected outcome is usually measured by the characteristics of the distribution, like standard deviation or the variance (see overview of economic risk measures in Ackermann 2013). In IS research, risk distribution characteristic based measures are frequently used by studies, which develop mathematical models to optimize the organizations' IT security levels and investments (e.g., Ackermann et al. 2013; Cavusoglu et al. 2008; Sonnenreich et al. 2006).

Psychological research is usually based on similar shortfall-oriented risk definition but the studies are frequently focused on the probability of a certain negative event, thus neglecting potentially different losses for different persons (see also related psychological research in Subsection 3.1.4 and Subsection 3.1.5) (Weinstein et al. 1990). In contrast to economic research, most of the psychological studies concern general life and health risks (e.g., likelihood

to have a coronary or a car accident), which can be assumed to be associated with comparable potential losses for everyone (e.g., Dillard et al. 2012; Rose 2010; Shepperd et al. 2013).

However, whenever the utilized risk measures are not based on long-term historical data, they generally capture the risk perceived by a person (e.g., the risk perception of a study participant) and not the actual risk of a negative event (Slovic 1987). Therefore, it is also not the actual risk that is central to risk related decisions (i.e., protection behavior) but the risk perceived by a person (Gigerenzer 2004).

2.2.1 *The Nature of Perceived Risks*

Risk perception is the subjective judgment that people make about the likelihood and the severity of a risk (e.g., natural hazards and threats to the environment or health risk). Psychological research generally distinguishes between two different measures of risk perception, which were found to be independently associated with behavioral intentions and other downstream beliefs (Dillard et al. 2012). In particular, absolute risk perceptions (e.g., a person's perception of his/or her risk to get a serious disease) were in many studies found to directly predict behavioral intentions, whereas comparative perceptions (e.g., a person's perception of his/or her risk to get a serious disease in comparison to the average person) were revealed to have more indirect effects (e.g., influence downstream beliefs, which facilitates protection behavior) on people's behavior (Shepperd et al. 2013). Absolute and comparative risk perceptions are frequently measured using numerical (e.g., from 0% to 100%) or categorical (e.g., from low to high) scales (Dillard et al. 2012). Besides, Weinstein et al. (2007) recently proposed a feeling-of-risk measure (e.g., a person's feeling to get a serious disease), which is measured on a verbal scale (e.g., from disagree to agree), and according to the authors represents a more intuitive and thus accurate measure of the personal risk exposure (see a detailed description of the different perspectives on risk perceptions in Subsection 3.1.4).

However, independent from the utilized risk measures, the risk perceived by a person or group of persons often considerably deviates from the actual risk of a negative event. In particular, previous research has found people generally to be very incongruent in their risk perceptions in a wide range of areas (e.g., Slovic 1987). The early studies on perceived risk even arose from the observation that experts and people often disagreed about the risks of various technologies and natural hazards (Starr 1969). In this regard, several theories have been proposed to understand why different people make different estimations of the characteristics and severity of any one risk. While early approaches assumed that individuals always behave rationally, weighing information before making a decision, and that misjudgments of risks were caused by inadequate or incorrect information (e.g., Douglas 1985), several studies have rejected the belief that additional information can shift perceptions (e.g., Freudenburg 1993). Furthermore, previous studies have demonstrated that even people with the same information

regarding a risk are likely to have different risk perceptions (e.g., Sjöberg 1998). Psychological research later revealed that numerous social and cognitive factors in risk information processing substantially predict a person's risk perceptions. In this context, especially the use of cognitive heuristics – such as representativeness heuristic, availability heuristic, and anchoring heuristic – in sorting and simplifying information was found to lead to biases in comprehension and risk perception (Tversky and Kahneman 1973). Later works identified various additional factors responsible for influencing a person's risk perception; these include dread, susceptibility, newness, stigma, and other factors (e.g., Lerner and Keltner 2000). Furthermore, numerous cognitive biases are likely to distort an individual's risk assessment. For example, since people tend to attribute to themselves various desirable characteristics that they do not necessarily possess, and interpret ambiguous information or unknown situations in self-favoring ways, they are likely to perceive themselves as having more control (i.e., illusion of control bias) and to be less vulnerable than others (i.e., unrealistic optimism; see Subsection 3.1.5) (e.g., Thompson 1999; Weinstein 1980). Errors in people's risk perceptions are likely to inevitably reduce their motivation to take necessary precautionary actions (Weinstein 1989).

In psychological research, a rich stream of literature showed that the risk perception largely predicts people's protection behavior (e.g., Rogers 1975; Weinstein 1993; Witte 1992). In particular, it has been conclusively shown that people's protection behavior usually results from a rational weighing up of the expected costs and perceived benefits of the precautionary actions (see also protection motivation theory and models in Subsection 3.1.2). In this context, psychological studies have revealed that people's beliefs about the benefits of protection behavior are predominantly predicted by their assessment of their personal threat situation, or their vulnerability to a threat (e.g., Milne et al. 2000; Weinstein et al. 1998; Witte and Allen 2000). Therefore, decision-oriented theories assume that peoples' perception of their susceptibility to threats and the perceived severity of the threats increase their perceived benefits of protection behavior (Milne et al. 2000). Accordingly, people who perceive themselves to be particularly at risk are in general more likely to assess protection behaviors as beneficial, and subsequently to take precautionary actions, such as having medical checkups, eating low-fat food, or using condoms, than other people (Breakwell 2000; Goodman et al. 1995; Weinstein and Nicolich 1993).

In economic research, perceived risk is commonly understood as the feeling of uncertainty regarding the possible negative consequences of adopting a product or service (e.g., the expectation of loss associated with the purchase of a product) (Cunningham 1967). In particular, previous studies frequently demonstrated that perceived risks are an important inhibitor of purchase behavior in various contexts, ranging from automobile to food purchases (e.g., Hammitt 1990; Peter and Ryan 1976; Srinivasan and Ratchford 1991). In particular, many

studies have demonstrated that the perceptions of risks are especially relevant for purchasing decisions when the decision-making is associated with uncertainty, discomfort, and/or fear, or creates conflicts with other people (e.g., Peter and Ryan 1976). Therefore, it has been revealed by various studies at the individual level that the perception of risks strongly determines the formation of adoption intentions and downstream beliefs (Ajzen and Fishbein 1980; Mitchell 1999; Murray and Schlacter 1990). At the organizational level (March and Shapira 1987), previous studies similarly revealed that the behavior of organizations is strongly influenced by its decision makers' perception of the risks related to the behavior, such as decisions to out-source business processes or to use banking services (e.g., Gewald and Dibbern 2009; Rotchanakitumnuai and Speece 2003).

In our discipline, drawing on economic and psychological literature, previous research demonstrated that people's perception of IT security risks strongly inhibits their intention to adopt technologies, such as the Internet, e-banking, or e-commerce systems (e.g., Lee 2009; Liebermann and Stashevsky 2002). Moreover, a growing stream of IS research investigates the effects of users' perception of privacy risks, which are closely related to IT security risks, on their usage behavior of different technologies, such as ubiquitous computing or online social networks (Smith et al. 2011). In this regard, people's privacy concerns were repeatedly found to considerably reduce their intention to use IT systems or disclose specific information (e.g., Dinev and Hart 2006; Krasnova et al. 2009). Based on psychological protection motivation research (see protection motivation and technology threat avoidance theory in Subsection 3.1.2), other IS studies demonstrated that the perception of ITSR strongly facilitates people's IT security behavior, like users' intention to implement certain safeguarding measures or employees' motivation to comply with corporate IT security policies (e.g., Liang and Xue 2010; Vance et al. 2012). In contrary, drawing on fear appeals research, other studies found that the perception of ITSR concurrently arouse fear on parts of the users, which in many cases reduces their intention to adopt safeguards and emotionally downplay the risk (e.g., Johnston and Warkentin 2010; Liang and Xue 2009). However, in contrast to economic research, previous IS studies regarding the perception of ITSR are predominantly focused on the individual user level and largely neglected the organizational level. Moreover, despite their theoretical importance, no IS research has been devoted to understand how the two different perspectives of perceived risks (i.e., absolute and comparative ITSR perception) affect IT security-related behaviors of people (the two different measures of risk perception in the IT security context are introduced in Subsection 3.1.4).

2.2.2 *Perceived IT Security Risks in the Context of the Cloud*

A considerable amount of literature has been published concerning the risks related to ITO (Ackermann 2013). Earl (1996) analyzed the risks in the context of traditional ITO, such as the possibility of hidden costs, business uncertainty, outdated technology skills, loss of inno-

vative capacity, and technology indivisibility. Dibbern et al. (2004) and Willcocks et al. (2007) present a literature review of ITO risks. Lacity et al. (2009) published a review of risks in the context of application service provision (ASP). They revealed that the major risks are contract, security, or privacy breaches by the provider, poor capability of the service, lack of trust in the providers, and vendor lock-in effects due to high switching costs. Ackermann et al. (2011) provides a comprehensive literature review and presents a list of 70 technological risks of ITO, such as theft of intellectual property, unsatisfactory software quality, and network issues.

Previous IS studies also examined different factors of perceived risk in the context of e-services, and studied its effects on users' adoption intention. Featherman and Pavlou (2003) were the first to operationalize perceived risks in the context of e-services based on the general risk perception theory. The authors conceptualize the overall perceived risk of e-services as a multi-dimensional constructs, consisting of performance, financial, time, psychological/social, technological, and privacy risk factors. Additionally, they have proposed and empirically tested an e-service adoption model, which explicates the effects of perceived risk on the use of e-services. The results of their study thereby reveal that the development of e-services has shifted the focus of the considered risk dimensions. While for traditional IT services, mainly strategic and financial risks were relevant to decision makers, the emergence of e-services substantially increased the importance of technology-related risks (Featherman and Pavlou 2003). Benlian and Hess (2011) have studied the opportunities and risks of Cloud adoption, perceived by IS executives at adopter and non-adopter firms. In the context of the Cloud, they were able to demonstrate that the ITSR is the most important factor predicting user adoption intentions.

Ackermann et al. (2012) defined perceived ITSR in the context of CC as decision-makers' perceived risk related to the IT security of a company's systems and data if the Cloud is utilized as a delivery model. The authors proposed a framework with a set of 31 risk items, which cover the identified ITSRs of the Cloud mutual exclusively and exhaustively (see Table 2-1). The risk items are grouped into six distinct risk dimensions (see also): confidentiality, integrity, availability, performance, accountability, and maintainability. The risk dimension availability means that users are able to access a service and the data whenever they wish. Confidentiality means that data can be read only by authorized users. Integrity relates to risks concerning data modification by unauthorized persons. Performance denotes that service and data usage speeds meets customers' requirements. Maintainability remains intact when it is possible to adapt a service to individual requirements and when the provider ensures maintenance and support. Accountability risks arise if authentication mechanisms can be eluded and if actions cannot be attributed clearly to one user (Ackermann et al. 2012).

IT Security Risk Management in the Context of Cloud
Computing
Towards an Understanding of the Key Role of Providers'
IT Security Risk Perceptions
Loske, A.
2015, XXII, 167 p. 11 illus., Softcover
ISBN: 978-3-658-11339-1