ESCUELA DE DOCTORADO
INTERNACIONAL (EDIUS)

TESIS DE DOCTORADO

A FRAMEWORK FOR EVALUATING SECURITY, TRUST, AND EFFICIENCY OF SUSTAINABLE CLOUD COMPUTING FOR BIG DATA PROCESSING

Presentada por:

Mohammad Naser Aladwan

Dirigida por:

Tomás Fernández Pena
José Carlos Cabaleiro

ESCUELA DE DOCTORADO INTERNACIONAL PROGRAMA DE DOCTORADO EN INVESTIGACIÓN EN TECNOLOXÍAS DA INFORMACIÓN

Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS)
SANTIAGO DE COMPOSTELA
10 de junio de 2020

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DEL AUTOR DE LA TESIS

A framework for evaluating security, trust, and efficiency of sustainable Cloud Computing for Big Data processing

Don Mohammad Naser Aladwan

Presento mi tesis, siguiendo el procedimiento adecuado al Reglamento, y declaro que:

1. *La tesis abarca los resultados de la elaboración de mi trabajo.*
2. *En su caso, en la tesis se hace referencia a las colaboraciones que tuvo este trabajo.*
3. *La tesis es la versión definitiva presentada para su defensa y coincide con la versión enviada en formato electrónico.*
4. *Confirmo que la tesis no incurre en ningún tipo de plagio de otros autores ni de trabajos presentados por mí para la obtención de otros títulos.*

En Santiago de Compostela, 10 de junio de 2020

Fdo. Mohammad Naser Aladwan

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DEL DIRECTOR/TUTOR DE LA TESIS

A framework for evaluating security, trust, and efficiency of sustainable Cloud Computing for Big Data processing

Don Tomás Fernández Pena, Profesor Titular da Área de Arquitectura e Tecnoloxía de Computadores da Universidade de Santiago de Compostela

Don José Carlos Cabaleiro, Profesor Titular da Área de Arquitectura e Tecnoloxía de Computadores da Universidade de Santiago de Compostela

INFORMAN:

*Que la presente tesis, corresponde con el trabajo realizado por **Don Mohammad Naser Al-adwan** bajo nuestra dirección, y autorizamos su presentación, considerando que reúne los requisitos exigidos en el Reglamento de Estudios de Doctorado de la USC, y que como directores de ésta no incurre en las causas de abstención establecidas en Ley 40/2015.*

En Santiago de Compostela, 10 de junio de 2020

Fdo. Tomás Fernández Pena
Director/a tesis

Fdo. José Carlos Cabaleiro
Director/a tesis

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Dedication

I am dedicating this thesis to my beloved parents, who have meant and continue to means so much to me.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Are those who have knowledge equal to those who do not have knowledge

Holy Qura'n, Surah Az-Zumar 39:9

And Allah taught Adam all the names…

Holy Qura'n, Surah Al-baqarah 2:31

When a man dies, his deeds come to an end except for three things: Sadaqah Jariyah (ceaseless charity); a knowledge which is beneficial, or a virtuous descendant who prays for him (for the deceased)

Prophet Mohammad

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Acknowledgments

All praise and thanks are due to the Almighty Allah, who always guides me to the right path and has helped me to complete this chapter of my life.

Undertaking this Ph.D. has been a truly life-changing experience for me, and it would not have been possible to do without the support and guidance that I received from many people.

I would like to express my special appreciation to everyone who supported me on the journey of completing my Ph.D. thesis. Special thanks to my parents (Abo Bashar and Om Bashar) for their support, sacrifices, and advice. My words and expressions cannot describe my feelings and the amount of my gratitude, my thanks, and my pride. You are Illuminates my way, and Without your prayers, I would not have reached this stage.

I want to extend my sincere thanks and gratitude to my supervisors, Assoc. Prof. Tomás Fernandez Pena and Assoc. Prof. José Carlos Cabaleiro, for their patience, guidance, continuous advice, and dedication during my Ph.D., without your directives, instructions, and valuable comments that enriched my research, I could not achieve this success.

My deepest gratitude to my brothers Bashar (Abo Nasser), Mizyed (Abo Mohammad), Yahia, Meghem (Abo Moaath), Emad (Abo Own), Murad (Abo Ahmad), and my lovely sisters for their help and support, also will not forget to thank all of my uncles, aunts, cousins, nephews, and nieces for their support, without you what would be a dream come true.

I am also profound gratitude to all those who support and help without bothering, especially Dr. Firas Awaysheh, Dr. Saadi Abadi, and Dr. Ahmad Abu Roman. I am indebted to thank all my friends, family, all my Jordanian colleagues in Santiago De Compostela. I also want to thank all my dear friends in the USA, Canada, Australia, and Europe.

Finally, to everyone who believes in me and my success, it is an honor to pour out my heart on this stage for you.

10 de junio de 2020

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Summary

Many organizations seek to employ Big Data (BD) to improve their decision-making and business sustainability. Cloud computing becomes the preferable deployment model for many industries and academia alike. The cloud paradigm is enabling them to take advantage of the provided efficiency, scalability, cost savings, and flexibility for large-scale data processing, management, and storage. The success of the cloud for BD is no longer dominated only by performance and cost. Rather, there are several other elements, areas, and criteria related to trust that need to be taken into consideration. These considerations are mainly due to the fact that data is no longer physically maintained under the organization's direct control, which raises new security concerns. Lay the groundwork for designing, improving, and implementing strategies to ensure a realistic BD over Cloud (BigCloud) model security is now critical more than ever. However, addressing BigCloud vulnerabilities and the research in this critical domain was limited and considered a cloud computing security issue without remedying BD specifications.The need for an in-depth investigation to develop a technology-independent reference architecture and evaluation of trust for BigCloud systems is very limited. Such a study can be focused on addressing the security concerns of the BigCloud and associated BD technologies, as for IoT-to-the cloud paradigm.

This thesis establishes a framework relying on the security of the BigCloud architecture to improve security issues. Specifically, this thesis research aims to examine the common features and the security challenges of this integration to provide an architecture relying on the security analysis, evaluation theory, and security by design of the cloud deployment architecture to improve the large-scale data processing security issues. Also, it aims at enhancing the cloud-based BD frameworks security in storage, motion, and process. Implementing best guidelines and practices for managing security related to BD operations over cloud computing technology and updating industry security guidelines, frameworks, and standards are of this

thesis concerns. Moreover, this thesis introduces an analytical model for data-intensive use case (i.e., Iot-to-cloud data streaming) security measurements, within any cloud-based framework. The thesis aims to fill the existing gap between the statistical representation of quality approaches of software engineering and the analysis of securing BD applications. Also, recommend best practices and measurements when constructing BD systems in both centralized and decentralized clouds. The proposed reference model and framework provides a comprehensive and fundamental basis to optimize the design of BigClouda frameworks regarding security ultimately.

Key Words Cloud Computing, Vehicular Cloud, IoT, Privacy, Evaluation Framework, Security by Design.

Background

This thesis's focus is to enhance the security modeling, implementation, and testing of cloud computing to cope with the BD environment without redesigning or modifying its structures. Also, toward improving BigCloud frameworks security in storage, motion, and processing. Our investigation of BigCloud security challenges has identified a research gap in the area of security by design of BD systems and how to evaluate the system security components. Figure 1.1 shows a diagram that illustrates our vision for the BigCloud software platform as a "Big Data as a Service" for BD frameworks deployments, whether centralized or decentralized architectures, in the public cloud.

This thesis researches two main problems related to modern cloud security: BD operational management in centralized and decentralized clouds. The thesis combines these domains as well as the associated requirements and features.

The first problem is related to meeting centralized cloud security at the design level. Security by design is increasingly becoming the mainstream development approach to ensure the security and privacy of different deployment systems. It indicates that the system has been designed from the foundation to be secure. In such an approach, the alternate security tactics and patterns are first thought, and, among them, the best is selected and enforced by the architecture design. It is then used as guiding principles for developers.

The second domain discussed in this thesis is related to developing a security evaluation framework for one of the most trending challenges in the modern decentralized cloud deployments architectures, in particular, the vehicular clouds that represent a special case of the Intelligent Transportation Systems (ITS). This use-case represents a data streaming approach

towards the cloud. We also discuss the critical realization of security by design, security control elements, and an evaluation approach in modern computer science.

Cloud computing

Cloud computing is changing the way IT industry is delivered in enterprises around the world. Its features include improving cost efficiencies, availability, accelerated innovation, faster time-to-market, ability to scale applications on-demand, and better resources customized for on-demand use-cases. Cloud computing is very well known as a critical IT resource for organizations that seek to harness the power of big data analysts (BDA) to improve their decision-making and business sustainability. Its importance lies in allowing organizations to scale up or scale down IT infrastructure properties (such as memory storage, CPU) according to their demands. Additionally, the cloud's adoption could help organizations save unnecessary expenditure on buying, managing, and upgrading IT resources for BD and BDA processing and handling. The use of cloud computing plays a vital role in shifting the responsibilities of buying, controlling, and maintaining the infrastructure and software, which are shifted from organizations to cloud providers. This shift will help organizations to focus more on their core business and leave many IT-related activities to be handled by cloud providers. In the cloud computing context, privacy is a crucial dimension of trust and many service providers have stressed it.

Vehicular Cloud

This thesis sheds light on the Vehicular Cloud (VC) security as a case of study for BD streaming to the cloud. VC is a group of broadly autonomous vehicles whose corporate computing, sensing, communication, and physical resources can be coordinated and dynamically allocated to share Internet access and therefore, exchanging data and information with other devices both inside and outside the vehicle. The VC can be formed by autonomous vehicles and provides a vast number of applications and services that can benefit the entire transportation system and drivers, passengers, and pedestrians. The intent to utilize the excessive onboard resources in the transportation system, along with the latest computing resource management technology in conventional clouds, has cultivated VC's concept. In general, it is composed of Vehicular Ad Hoc Networks (VANET), which is a wireless ad hoc network of mobile vehicles. Also, the communication between the vehicles and other objects (vehicle-to-everything, V2X) as for infrastructure and other vehicles.

Big Data and Hadoop

Data can be structured (e.g., financial, electronic medical records, government statistics),

semi-structured (e.g., text, tweets, emails), unstructured (e.g., audio and video), and real-time (e.g., network traces, generic monitoring logs). All of these applications share the potential for providing invaluable insights if organized and analyzed appropriately. Big Data Analytics allows getting enormous benefits by dealing with any massive volume of unstructured, structured, and semi-structured content. However, Big Data processing represents a magnificent risk for many users due to the various data gathered from all available sources. On the other hand, Hadoop and its associated technologies, i.e., YARN for resource management and HDFS for distributed storage, had been widely used in the last years as a defacto platform for all BD workloads. These workloads are spanning from batch-processing, e.g., MapReduce, and micro-batch processing (Apache Spark) to real-time processing frameworks (Apache Flink or Apache Storm). As a large-scale system, Hadoop reliability and availability are significant issues not solved in the standard implementation.

Hadoop abstracts computing resource management, task scheduling, and data management while maintaining a satisfactory security and isolation level. The Hadoop stack is typically deployed as part of a large-scale BD platform to support commodity hardware and accommodate different processing frameworks. It is utilized to handle data management and access to the Hadoop ecosystem using a master/slave architecture. The large community behind Hadoop has been working to improve its stack to meet BD's increasing demands and requirements. The new features shift proves the maturity and applicability of Hadoop 3.x to serve different markets, thus, making it a prominent paradigm that combines large-scale computing nodes to build an analytical environment, e.g., BD processing and storage.

Research Objectives

There is a critical need to systematically and comprehensively evaluate the security mechanisms implemented to meet the security requirements of BD applications and ensure that computer security is fail-safe and provided on IaaS/PaaS cloud environments. The developed model when deployed assesses the compliance, capabilities, and limitations of the security mechanism implemented on the cloud environment. Also, this model attempts to test the effectiveness of the BD frameworks in capturing customer security requirements in the evaluation and assessment process. In this thesis, we evaluate the compliance, abilities, and restrictions of the security mechanism when a cloud-based BD solution is deployed. Also, we attempt to test the effectiveness of the framework in capturing customer security requirements in the evaluation and assessment process.

BigCloud presents a complex architecture that needs to be broken down and understood to evaluate security requirements that are expected to be met by security methods and controls. Mainly, this objective aims to evaluate cloud architecture's ability to deliver a system that fulfills the users' security requirements and to identify potential risks on each component of the BD frameworks. Also, it is required a developed reference model that can be adapted to segregate IaaS/PaaS cloud architectures into layers in an attempt to identify its components, and how these components integrate to provide cloud services and security mechanisms implemented in the cloud.

The primary goals of this thesis are:

- **Objective 1:** A critical evaluation of BD platforms over cloud architectures.
- **Objective 2:** A thorough review of current security and privacy issues in cloud-based Big Data applications/frameworks and their security requirements mapping and classification.
- **Objective 3** A comprehensive evaluation of different techniques regarding BD service security, concerning infrastructure, data, and applications.
- **Objective 4** To establish an architecture relying on the security of the network and identify the framework components.
- **Objective 5** To propose a security analysis model for Big Data platforms, which integrates software quality concepts and best practises.
- **Objective 5:** To formalize security abstractions, besides reliability and trust measurements of cloud-based Big Data solutions.
- **Objective 6** To propose and intensively evaluate a security evaluation framework for vehicular cloud applications.

Research Questions

The critical questions that this thesis discusses on BigCloud security are:

i. How to optimize BD framework security at the design phase of cloud deployment architectures.

ii. How to analyze and evaluate the security level of a cloud-based BD solution.

iii. What are the common security criteria of all BigCloud models, and how can we optimize them toward large-scale data streaming models.

Thesis Methodology

The primary objective of our research plan is to focus on the integration of both BD processing frameworks (e.g., Hadoop) and BD streaming (e.g., IoT-based applications) to the cloud deployment model, analyzing the characteristic features, security gaps, and providing a security model to improve the large-scale data processing over the cloud.

This thesis achieved its goals by survey all the current security and privacy challenges imposed by processing BD workloads in the cloud. We start assessing all require tools, applications, and data dependencies in deploying the proposed solutions. Next, we define common criteria that unified the security evaluation of different BD implementations and use cases. Afterward, we enrolled in studying the security in design impact on the overall BD security. Finally, we draft our novel evaluation framework.

For formalizing the thesis background, we start with survey the security challenges of the integration of BD deployments and Cloud Computing. Our objective is to provide an architecture relying on the security of the network in order to improve the security issues. Next, we search to define and examine the needed software tools for integrating the model-driven engineering concepts in the form of a meta-model, which a cloud-based big data application can exploit to become security-aware. With this approach, we seek to capture security requirements and capabilities to drive application deployment as well as security-oriented scalability rules to guide application re-configuration.

In a third phase, we study the underlying cloud-based system vulnerabilities aiming to addresses its challenges. Specifically, we focus on improving the large-scale data processing security issues in both centralized and decentralized clouds, and the Cloud-based BD frameworks security in storage, motion, and processing. We achieve these goals by implementing the best guidelines and practices for managing security related to BD operations over cloud computing technologies and updating industry security guidelines, frameworks, and standards. However, to establish an evaluation framework, we relied on defining standard security criteria within any data stream solution to the cloud. These criteria can be shared within any BD platforms over the cloud deployment architectures.

A fourth phase includes designing and implementing the proposed model architecture to improve the security issues based on the security of the BD platform/user. Besides, a method for creating re-usable security elements facilitating rapid security model specification conforming to the meta-model is also suggested, to reduce the designer's modeling effort and automate the cloud-based Big Data applications security management. While at the fourth phase, the evaluating experiments start, to conceptually evaluate the model performance, the results checking will take place during the fifth phase, to modify the selected modes and edit the architecture design. During the sixth phase, we will conduct an extensive evaluation of the proposed model, and, finally, in the last phase, we submit the final architecture design and we write up the thesis.

Results and Publications

This thesis aims to explore the BigCloud security criteria that can influence the overall security degree for BD processing over the cloud environment. Converging cloud-security concerns to the software engineering technical expertise can sustain this tendency, by drafting security models that address these concerns.

This research contributes to the deployment of the VC security body of knowledge by first examining in detail the building blocks of the vehicular cloud security stack architecture. Second, it analyses and classifies state-of-the-art security frameworks, which are mainly available today as open-source platforms for sophisticated criteria selection. Third, a rigorous and robust evaluation framework based on evaluation theory, which guides VC service providers to identify security gaps, is proposed. Finally, we highlight some open challenges and recommendations for both service providers and customers for a comprehensive discussion toward achieving the vision of providing IoV-cloud secure services.

Our thesis led to the exploration, identification, and understanding of six common security criteria (CSC's) that generally influence BigCloud security, namely:

- Physical and Environmental Protection
- Logical Access Control
- Communication Confidentiality
- Communication Integrity

- Data and Service Availability
- Data Privacy and Governance

The findings of these CSCs are presented according to the directed content analysis approach of the BigCloud environment. These CSCs focuses on identifying and extracting the security criteria that affect the deployment of BD applications in the cloud, including IoT-to-cloud applications. After comparing the similarity and differences of those criteria, they were grouped and labeled into sub-categories according to their relationships. All categories were combined with the exact main category.

However, evaluating the trustworthiness of the security design and implementation is vital for the sustainable integration of such a paradigm. This can be achieved by regularly checking the security components of the VC to devise an adequate plan for improvement. In this regard, we present a rigorous and robust evaluation model called trustworthiness evaluation of vehicular cloud (TrustE-VC) [10]. The proposed framework formalizes the main ideas discussed in the literature for security criteria evaluation and selection. Diagrammatic security levels, multicriteria decision making, and additive weight fuzzy ranking are the framework components. TrustE-VC identifies the standard security criteria, providing security gap analysis according to a multicriteria decision-making algorithm for supporting VC applications as a commodity service in the cloud. The results show that the average performance rate of TrustE-VC and the ideal point that describes the security criteria associated with each VC component need urgent improvement in several domains. Moreover, based on the framework results, several recommendations have been made to fill these gaps.

On the other hand, this thesis introduces an analytical model for BD security measurements within any cloud-based framework in the design phase. It is aiming to fill the existing gap between the statistical representation of quality approaches of software engineering and the analysis of securing BD applications. Our thesis recommends best practices and measurements when constructing BD systems in both centralized and decentralized clouds. The proposed model, thus, provides a comprehensive and fundamental basis to optimize the design of BD frameworks regarding security ultimately.

This thesis also summarizes the relationship between the security service and other cloud services as well as their functions. We offer a solution to the main research question, which deals with the security elements associated with BD deployment over the cloud. Our solution consists of main actors that emphasize the separation of concerns regarding the service func-

tionality (data service security, IaaS security, etc.) and non-functional security requirements at the beginning of the design, through a reuse-based approach and specifications. Moreover, we propose a security analysis pattern that refines the cloud context-pattern in synergy to an extended CIA triad (CIA refers to confidentiality, integrity, and availability of data), providing a set of guidelines for the structuring of BD specifications, which relates a cloud design to its security environment. Finally, we suggest a structured election method for BigCloud-specific security selection that delivers many insights regarding the latest ongoing developments and cutting-edge frameworks by mapping each security domain to its solution knowledge.

Publications

Next, we present a list of publications derived from the work developed in this thesis in peer reviewed journals and high-impact conferences.

Articles in international conferences:

- Mohammad N. Aladwan, Feras M. Awaysheh, Mamoun Alazab, J. C. Cabaleiro, T. F. Pena, and M. Alazab, "Common security criteria for vehicular clouds and internet of vehicles evaluation and selection," in 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2019, pp. 807–813 **GGS Rating: B-**

Published Journal articles:

- Mohammad N. Aladwan, Feras M. Awaysheh, Sadi Alawadi, Mamoun Alazab, Tomás F. Pena, José C. Cabaleiro "TrustE-VC: Trustworthy Evaluation Framework for Industrial Connected Vehicles in the Cloud" *IEEE Transactions on Industrial Informatics*, 16(9):6203-6213, 2020. Both first authors contributed equally to this publication. **Impact factor (JCR 2018): 7.377 (Q1)**. Rank **1/46** in the category ENGINEERING, INDUSTRIAL.

Under Review Journal articles:

- Feras M. Awaysheh, Mohammad Aladwan, Tomás F. Pena, José C. Cabaleiro "Security by Design for Big Data Frameworks over Cloud Computing" *IEEE Transactions on*

Engineering Management. **Impact factor (JCR 2018): 1.867 (Q3)**. Rank **24/46** in the category ENGINEERING, INDUSTRIAL.

Resumen

Moitas organizacións buscan empregar o Big Data (BD) para mellorar a toma de decisións e a sustentabilidade empresarial. A computación na nube vense convertendo, deste xeito, no modelo de despregamento preferible non so para a industria senón tamén para a comunidade investigadora. O paradigma na nube permítelles aproveitar a eficiencia, a escalabilidade, o aforro de custos e a flexibilidade proporcionados para o procesamento, xestión e almacenamento de datos a grande escala. O éxito da nube para BD xa non está dominado só polo rendemento e o custo, senón que hai que considerar tamén outros elementos, áreas e criterios relacionados coa confianza. Estas consideracións débense principalmente a que os datos xa non se manteñen fisicamente baixo o control directo da organización, o que suscita novos problemas de seguridade. Debido a isto, sentar as bases para deseñar, mellorar e implementar estratexias para garantir a seguridade dun modelo BD sobre Cloud (BigCloud) é agora máis importante que nunca. Os aspectos de seguridade no dominio BigCloud céntrase na actualidade na seguridade xeral da nube, sen ter en consideración as especificidades da natureza BD. A necesidade dunha investigación en profundidade para desenvolver unha arquitectura de referencia independente da tecnoloxía e a avaliación da confianza dos sistemas BigCloud é clara. Un estudo como este pode centrarse en atender as cuestións de seguridade do BigCloud e as tecnoloxías BD asociadas, como por exemplo no paradigma IoT-cloud.

Esta tese establece un marco baseado na seguridade da arquitectura BigCloud para mellorar os problemas de seguridade. En concreto, esta investigación busca examinar as características comúns e os retos de seguridade desta integración para proporcionar unha arquitectura baseada na análise de seguridade, a teoría de avaliación e a seguridade mediante o deseño da arquitectura de despregamento na nube co obxectivo de mellorar os problemas de seguridade no procesamento de datos a gran escala. Ademais, búscase mellorar a seguridade nas contornas Big Data baseados na nube tanto para o almacenamento, o movemento e o

procesamento da información. Outro interese da tese é a aplicación de directrices e prácticas óptimas para xestionar a seguridade relacionada coas operacións de BD a través da tecnoloxía de computación na nube, así como actualizar as directrices, marcos e estándares de seguridade da industria. Ademais, esta tese introduce un modelo analítico para as medidas de seguridade para casos de uso intensivo de datos (por exemplo, transmisión de datos de IoT á nube), dentro de calquera marco baseado na nube. A tese ten como obxectivo cubrir a brecha existente entre a representación estatística dos enfoques de calidade da enxeñería de software e a análise de aplicacións BD. Ademais, recomenda mellores prácticas e medidas á hora de construír sistemas de BD tanto en nubes descentralizadas como centralizadas. O modelo e marco de referencia proposto proporciona unha base completa e fundamental para optimizar o deseño de sistemas BigCloud tendo, en última instancia, a seguridade no seu foco.

Palabras chave Computación na nube, nubes de vehículos, IoT, privacidade, contorna de avaliación, seguridade por deseño.

Contexto

O obxectivo desta tese é mellorar o modelado, implementación e probas de seguridade da computación na nube para facer fronte as contornas BD sen redeseñar nin modificar as súas estruturas, buscando, ademais, incrementar a seguridade das contornas BigCloud tanto no almacenamento, como no movemento e o procesamento dos datos. O noso traballo sobre os retos de seguridade no BigCloud identificou unha falta de investigacións na área de seguridade no deseño de sistemas BD e na avaliación dos compoñentes de seguridade do sistema. A figura 1.1 mostra un diagrama que ilustra a nosa visión para unha plataforma de software BigCloud, que se presenta como unha solución tipo ”Big Data como servizo”para os despregamentos de contornas BD, tanto arquitecturas centralizadas ou descentralizadas, na nube pública.

Esta tese investiga dous dos principais problemas relacionados coa seguridade na nube moderna, que son, por unha banda, a xestión operativa de BD en nubes centralizadas e, pola outra, a correspondente a nubes descentralizadas. A tese combina estes dominios, así como os requisitos e as características asociadas.

O primeiro problema está relacionado co cumprimento a nivel de deseño da seguridade nas nubes centralizadas. A seguridade por deseño (*security by design*) estase a converterse cada vez máis no enfoque principal de desenvolvemento, xa que busca garantir a seguridade e a privacidade dos distintos modelos de despregamento, garantindo que o sistema foi deseñado

dende os seus inicios para ser seguro. Este enfoque comeza buscando diferentes tácticas e patróns de seguridade, e o mellor entre eles seleccionase e aplícase no deseño da arquitectura, actuando como principio director para os desenvolvedores.

O segundo aspecto analizado nesta tese está relacionado co desenvolvemento dun marco de avaliación de seguridade para un dos retos máis candentes nas arquitecturas modernas de despregamento en nubes descentralizadas, en particular, as nubes de vehículos, que representan un caso especial dos Sistemas Intelixentes de Transporte (*Intelligent Transportation Systems*, ITS). Este caso de uso é un exemplo de transmisión de datos (*streaming*) cara á nube. Por último, discutiremos a realización crítica da seguridade por deseño, os elementos de control de seguridade e as aproximacións para avaliación dos sistemas de computación actuais.

Computación na nube

A computación en nube está a cambiar o xeito no que a industria de TI se ofrece en empresas de todo o mundo. As súas características inclúen a mellora da relación custe/eficiencia, a mellora da dispoñibilidade, a aceleración da innovación, a redución do *time-to-market*, a capacidade de escalar aplicacións baixo demanda e a dispoñibilidade baixo demanda de mellores recursos personalizados para os casos de uso. A computación en nube é considerada un recurso crítico en TI para as organizacións que buscan aproveitar o poder dos analistas de grandes datos (BDA) para mellorar a toma de decisións e a sustentabilidade empresarial. A súa importancia reside en permitir ás organizacións escalar ou ampliar as propiedades da infraestrutura de TI (como, por exemplo, o almacenamento de memoria ou as necesidades de CPU) segundo as súas demandas. Ademais, a adopción da nube podería axudar ás organizacións a aforrar gastos innecesarios en compra, xestión e actualización de recursos de TI para o procesamento e manipulación do BD así como para o seu análise. O uso da computación en nube xoga un papel fundamental para cambiar as responsabilidades de compra, control e mantemento da infraestrutura e do software, que pasa de ser algo interno ás organizacións a o seu manexo por parte de provedores de infraestruturas na nube. Este cambio ven axudando ás organizacións a concentrarse máis no seu negocio principal, permitindo que actividades relacionadas coas TI sexan manexadas por provedores de nube. No contexto da computación na nube, a privacidade é unha dimensión crucial para a confianza dos clientes, tal e como xa veñen sinalando moitos provedores de servizos.

Nubes de vehículos

Esta tese arroxa luz sobre a seguridade da nube de vehículos (*Vehicular Cloud*, VC) co-

mo caso de estudo para a transmisión (*streaming*) de BD á nube. Unha VC é un grupo de vehículos amplamente autónomos cuxos recursos de computación, detección, comunicación e físicos pódense coordinar e asignar dinámicamente para compartir acceso a Internet e, polo tanto, intercambiar datos e información con outros dispositivos dentro e fóra do vehículo. Unha VC pode estar formado por vehículos autónomos e proporciona un gran número de aplicacións e servizos que poden beneficiar a todo o sistema de transporte, incluíndo condutores, pasaxeiros e peóns. O obxectivo de aproveitar o gran número de recursos a bordo do sistemas de transporte, xunto coas últimas tecnoloxías de xestión de recursos informáticos nas nubes convencionais, axudou á aparición do concepto de VC. Unha nube de vehículos baséase, polo xeral, en redes ad hoc (*Vehicular Ad Hoc Networks*, VANET), é dicir, redes sen fíos deseñadas especificamente para vehículos. Mediante estas redes se realiza a comunicación entre os vehículos e outros obxectos (*vehicle-to-everything*, V2X), como son as infraestruturas e outros vehículos.

Big Data e Hadoop

Os datos poden ser estruturados (por exemplo, rexistros médicos financeiros, electrónicos, estatísticas do goberno), semiestruturados (por exemplo, texto, tweets, correos electrónicos), non estruturados (por exemplo, audio e vídeo) e en tempo real (por exemplo, conexións de rede, logs de seguimento xenéricos). Todas estas aplicacións comparten o potencial de ofrecer coñecementos inestimables se se organizan e analizan adecuadamente. A analítica do Big Data (BDA) permite obter enormes beneficios ao poder manexar enormes cantidades de contido non estruturado, estruturado e semiestruturado. Non obstante, o procesamento do Big Data pode supoñer un gran risco para moitos usuarios, debido á diversidade dos datos recollidos de todas as fontes dispoñibles. Por outra banda, Hadoop e as súas tecnoloxías asociadas, é dicir, YARN para a xestión de recursos e HDFS para almacenamento distribuído, veñen sendo amplamente utilizadas nos últimos anos como plataforma defacto para todas as cargas de traballo BD. Estas cargas de traballo abarcan desde o procesamento por lotes, por exemplo, MapReduce e o procesamento de micro-lotes (Apache Spark), ata as contornas de procesamento en tempo real (Apache Flink ou Apache Storm). Como sistema a grande escala, a fiabilidade e a dispoñibilidade de Hadoop supón un problema significativo, que non ven totalmente resolto na súa implementación estándar.

Hadoop abstrae a xestión de recursos, a planificación de tarefas e a xestión de datos, mantendo un nivel de illamento e seguridade satisfactorio. A pila Hadoop desprégase normalmente como parte dunha plataforma BD a grande escala baseada en hardware de baixo custe, sendo

capaz de manexar diferentes contornas de procesamento. Baséase nunha arquitectura mestre/escravo que permite controlar a xestión de datos e o acceso ao ecosistema Hadoop. A gran comunidade detrás de Hadoop está a traballar para mellorar a súa arquitectura con vistas a atender ás demandas e necesidades crecentes de BD. A incorporación de novas funcionalidades demostra a madurez e a aplicabilidade de Hadoop 3.x para servir a diferentes mercados, o que o ten convertido nun paradigma destacable que combina nodos de computación a gran escala para construír unha contorna de análise de datos, incluíndo o almacenamento e o procesamento de datos masivos.

Obxectivos da investigación

Hai unha necesidade crítica de avaliar sistemática e exhaustivamente os mecanismos implementados para cumprir os requisitos de seguridade das aplicacións BD, así como de asegurarse de eses mecanismos son a proba de fallos e se atopan dispoñibles nas contornos IaaS/PaaS. O modelo desenvolvido nesta tese avalía o cumprimento, as capacidades e as limitacións dos mecanismos de seguridade implementados nas contornas na nube. Ademais, este modelo intenta comprobar a efectividade das contornas BD á hora de capturar os requirimentos de seguridade dos clientes nos proceso de avaliación e valoración. Nesta tese, avaliamos o cumprimento, as habilidades e as restricións dos mecanismos de seguridade cando se despreg a unha solución BD baseada na nube.

BigCloud presenta unha arquitectura complexa que debe ser desglosada e entendida para avaliar os requisitos de seguridade que se espera que cumpran os métodos e controis de seguridade. O obxectivo é, principalmente, avaliar a capacidade da arquitectura na nube para ofrecer un sistema que cumpra os requisitos de seguridade dos usuarios e identificar riscos potenciais en cada compoñente das contornas BD. Ademais, é necesario desenvolver un modelo de referencia que poida adaptarse para segregar arquitecturas dIaaS/PaaS en capas, de forma que sexa posible identificar os seus compoñentes e a súa integración para proporcionar servizos e mecanismos de seguridade na nube.

Os obxectivos principais desta tese son:

- **Obxectivo 1:** Facer unha análise crítica das plataformas BD sobre arquitecturas na nube.

- **Obxectivo 2:** Levar a cabo unha revisión profunda dos problemas de seguridade e privacidade actuais en aplicacións/contornas Big Data baseados na nube e a clasificación

dos seus requisitos de seguridade.

- **Obxectivo 3:** Acometer unha avaliación completa de diferentes técnicas sobre a seguridade do servizo BD, relativa a infraestruturas, datos e aplicacións.
- **Obxectivo 4:** Establecer unha arquitectura baseada da seguridade da rede e identificar os compoñentes da mesma.
- **Obxectivo 5:** Propoñer un modelo de análise de seguridade para plataformas Big Data, que integre conceptos e mellores prácticas de calidade do software.
- **Obxectivo 5:** Formalizar abstraccións de seguridade, ademais de medidas de fiabilidade e confianza de solucións Big Data baseadas na nube.
- **Obxectivo 6:** Propoñer e analizar de xeito intensivo unha contorna de avaliación da seguridade para aplicacións de nubes de vehículos.

Preguntas respondidas

As preguntas críticas que trata esta tese sobre a seguridade en BigCloud son:

i. Como optimizar a seguridade das contornas BD na fase de deseño das arquitecturas de despregamento na nube.

ii. Como analizar e avaliar o nivel de seguridade dunha solución BD baseada na nube.

iii. Cales son os criterios comúns de seguridade de todos os modelos BigCloud e como podemos optimizalos cara a modelos de transmisión de datos a gran escala.

Metodoloxía

O obxectivo principal do noso plan de investigación centrase na integración tanto de contornas de procesamento de BD (por exemplo, Hadoop) como de procesamento de fluxos de datos masivos (por exemplo, aplicacións baseadas en IoT) ao modelo de despregamento na nube, analizando as características principais, carencias de seguridade e proporcionando un modelo de seguridade para mellorar o procesamento de datos a gran escala na nube.

Esta tese alcanzou os seus obxectivos estudando todos os retos actuais de seguridade e privacidade impostos ao procesar as cargas de traballo asociadas a datos masivos na nube.

Comezamos avaliando todos os requirimentos de ferramentas, aplicacións e datos para implementar as solucións propostas. A continuación, definimos criterios comúns que unifican a avaliación de seguridade de diferentes implementacións de BD e casos de uso. Despois, nos centramos en estudar o impacto da seguridade no deseño xeral das solucións BD. Finalmente, redactamos o noso novo marco de avaliación.

Para formalizar o contexto da tese, comezamos co estudo dos retos de seguridade na integración dos despregamentos de BD sobre infraestruturas de computación na nube. O noso obxectivo é proporcionar unha arquitectura dependente da seguridade da rede para mellorar os problemas de seguridade. A continuación, buscamos definir e examinar as ferramentas software necesarias para integrar os conceptos de enxeñaría baseada no modelo (*model-driven engineering*) en forma dun meta-modelo, que poida ser explotado por unha aplicación de datos masivos baseada na nube para ser máis consciente dos problemas de seguridade. Con este enfoque, buscamos captar os requisitos e as capacidades de seguridade para dirixir o despregamento de aplicacións, así como regras de escalabilidade orientadas á seguridade para guiar a configuración das mesmas.

Nunha terceira fase, estudamos as vulnerabilidades subxacentes de sistemas baseados na nube co obxectivo de afrontar os seus retos. Concretamente, centrámonos en mellorar as cuestións de seguridade no procesamento de datos a gran escala tanto nas nubes centralizadas como nas descentralizadas, e na seguridade das contornas BD baseadas en Cloud, tanto no tocante ao almacenamento, ao movemento e ao procesamento. Conseguimos estes obxectivos implementando as mellores directrices e prácticas para xestionar a seguridade relacionada coas operacións de BD a través de tecnoloxías de computación na nube e actualizando directrices, marcos e estándares de seguridade da industria. Non obstante, para establecer un marco de avaliación, buscamos definir criterios estándar de seguridade dentro de calquera solución de fluxo de datos para a nube. Estes criterios pódense compartir con calquera plataformas BD a través das arquitecturas de despregamento na nube.

Unha cuarta fase inclúe o deseño e implementación da arquitectura modelo proposta para mellorar os problemas de seguridade baseados na seguridade da plataforma/usuario BD. Ademais, tamén se suxire un método para crear elementos de seguridade reutilizables que faciliten unha especificación rápida do modelo de seguridade conforme ao meta-modelo, para reducir o esforzo de modelado do deseñador e automatizar a xestión de seguridade das aplicacións Big Data baseadas na nube. Mentres que na cuarta fase comezamos os experimentos de avaliación, para avaliar conceptualmente o rendemento do modelo, a comprobación de resultados

tivo lugar durante a quinta fase, para modificar os modos seleccionados e editar o deseño de arquitectura. Durante a sexta fase, realizamos unha ampla avaliación do modelo proposto e, finalmente, na última fase, establecemos o deseño da arquitectura final e redactamos a tese.

Resultados e publicacións

Esta tese pretende explorar os criterios de seguridade no BigCloud que poden influír no grao de seguridade global para o procesamento de BD sobre a nube. Enfocar os problemas de seguridade na nube desde o punto de vista da experiencia técnica da enxeñería de software resulta de interés, ao permitir a definición de modelos de seguridade que respondan a estes problemas.

Esta investigación contribúe á mellora do corpo de coñecemento relacionado coa seguridade nas nubes de vehículos, en primeiro lugar, examinando en detalle os bloques da arquitectura da pila de seguridade na nube vehicular. En segundo lugar, analiza e clasifica as contornas de seguridade de última xeración, hoxe dispoñibles principalmente como plataformas de código aberto, para a selección de criterios sofisticados. En terceiro lugar, propón un marco de avaliación rigoroso e robusto baseado na teoría da avaliación, que guía aos provedores de servizos de VC a identificar as carencias de seguridade. Finalmente, resaltamos algúns desafíos e recomendacións abertos tanto para provedores de servizos como para clientes para unha discusión integral con vistas ofrecer servizos seguros IoV-cloud.

A tese levou a cabo á exploración, identificación e comprensión de seis criterios comúns de seguridade (*Common Security Criteria*, CSC) que habitualmente inflúen na seguridade do BigCloud, a saber:

- Protección física e ambiental.
- Control de acceso lóxico.
- Confidencialidade das comunicacións.
- Integridade das comunicacións.
- Dispoñibilidade de datos e servizos.
- Privacidade e gobernanza de datos.

A definición destes CSCs preséntanse segundo a aproximación baseada en análise de contido dirixido das contornas de BigCloud. Estes CSCs céntranse na identificación e extracción de criterios de seguridade que afectan ao despregamento de aplicacións BD na nube, incluídas as aplicacións IoT-to-cloud. Despois de comparar a semellanza e diferenzas deses criterios, agrupáronse e etiquetáronse en subcategorías segundo as súas relacións. Todas as categorías combináronse coa categoría principal concreta.

Non obstante, avaliar a confiabilidade do deseño e a implementación da seguridade é fundamental para a integración sostible deste paradigma. Isto pódese conseguir comprobando regularmente os compoñentes de seguridade do VC para elaborar un plan adecuado de mellora. Neste sentido, presentamos un rigoroso e robusto modelo de avaliación, denominado avaliación de confianza da nube vehicular (*Trustworthiness Evaluation of Vehicular Cloud*, TrustE-VC) [10]. O marco proposto formaliza as principais ideas discutidas na literatura para a avaliación e selección de criterios de seguridade. Os niveis esquemáticos de seguridade, a toma de decisións multicriterios e a clasificación difusa do peso aditivo son os compoñentes deste marco. TrustE-VC identifica os criterios estándar de seguridade, proporcionando análises das fendas de seguridade segundo un algoritmo de toma de decisións multicriterio para apoiar as aplicacións VC como servizo de mercadorías na nube. Os resultados mostran que a taxa de rendemento media de TrustE-VC e o punto ideal que describe os criterios de seguridade asociados a cada compoñente de VC precisan dunha mellora urxente en varios dominios. Ademais, en función dos resultados marco, fixéronse varias recomendacións para cubrir estas lagoas.

Por outra banda, esta tese introduce un modelo analítico para as medicións de seguridade, na fase de deseño, dentro de calquera contorna BD baseada na nube. Pretende cubrir o oco existente entre a representación estatística dos enfoques de calidade da enxeñería de software e a análise da seguridade de aplicacións BD. A nosa tese recomenda as mellores prácticas e medidas á hora de construír sistemas de BD en nubes descentralizadas e centralizadas. O modelo proposto, polo tanto, proporciona, en última instancia, unha base completa e fundamental para optimizar o deseño en materia de seguridade de contornas BD.

Esta tese tamén resume a relación entre o servizo de seguridade e outros servizos na nube, así como as súas funcións. Ofrecemos unha solución á pregunta principal de investigación, que trata dos elementos de seguridade asociados ao despregamento de aplicacións BD na nube. A nosa solución está formada por actores principais que destacan a separación de intereses con respecto á funcionalidade do servizo (seguridade do servizo de datos, seguridade IaaS, etc.) e

os requisitos de seguridade non funcionais ao comezo do deseño, a través dun enfoque e especificacións baseadas na reutilización. Ademais, propoñemos un patrón de análise de seguridade que refina o patrón de deseño de contexto da nube en sinerxia cunha tríada CIA estendida (CIA refírese á confidencialidade, integridade e dispoñibilidade dos datos), proporcionando un conxunto de directrices para a estruturación das especificacións dos BD, que relaciona un deseño na nube co seu contorno de seguridade. Finalmente, suxerimos un método de elección estruturado para a selección de aspectos de seguridade específicos do BigCloud que proporciona achegas valiosas sobre as contornas de vangarda, mediante a asignación á cada dominio de seguridade da súa solución coñecida.

Publicacións

A continuación, presentamos unha lista de publicacións en revistas revisadas por pares e conferencias de alto impacto derivadas do traballo desenvolvido nesta tese.

Articles in international conferences:

- Mohammad N. Aladwan, Feras M. Awaysheh, Mamoun Alazab, J. C. Cabaleiro, T. F. Pena, and M. Alazab, "Common security criteria for vehicular clouds and internet of vehicles evaluation and selection," in 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 2019, pp. 807–813 **GGS Rating: B-**

Artigos en conferencias internacionais:

- Mohammad N. Aladwan, Feras M. Awaysheh, Sadi Alawadi, Mamoun Alazab, Tomás F. Pena, José C. Cabaleiro "TrustE-VC: Trustworthy Evaluation Framework for Industrial Connected Vehicles in the Cloud" *IEEE Transactions on Industrial Informatics*, 16(9):6203-6213, 2020. Os dous primeiros autores contribuíron en igual medida nesta publicación. **Impact factor (JCR 2018): 7.377 (Q1)**. Ranking **1/46** na categoría ENGINEERING, INDUSTRIAL.

Artigos en revisión:

- Feras M. Awaysheh, Mohammad Aladwan, Tomás F. Pena, José C. Cabaleiro "Security by Design for Big Data Frameworks over Cloud Computing" *IEEE Transactions on*

Engineering Management. **Impact factor (JCR 2018): 1.867 (Q3)**. Ranking **24/46** na categoría ENGINEERING, INDUSTRIAL.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

CHAPTER 1

INTRODUCTION

1.1 Context Background

Cloud computing has recently emerged as a reliable, trustworthy, and cost-benefit model in the large-scale data analytics industry. Numerous IT vendors are offering computation, storage, and application hosting services in response to the rapidly increased amounts of available data to be known as Big Data (BD). Processing BD over cloud configurations is a recently investigated technology, and it has many gaps in security and privacy. Cloud clients are not sure of the security techniques implemented and how they are integrated to provide sufficient security for their data and applications. The research community, therefore, must understand and evaluate security mechanisms and controls implemented to maintain the availability, integrity, and confidentiality of data processed, stored, and accessed in the cloud. Also, there is a need to ensure that these mechanisms meet security standards and requirements to mitigate any security risks. The research community must consider these issues by proposing robust protection techniques that enable getting benefits from BD without risking privacy.

In general, cloud computing paradigm is constructed based on several technologies such as virtualization, container technology, Service-Oriented Architecture (SOA), and utility computing, involving three main service layers, namely, Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). Security aspects are essential in motivating clients to adopt cloud computing services. A survey on cloud challenges illustrates that security is the main barrier and obstacle to widespread adoption of cloud computing for BD services [102]

The focus of this research is to enhance the security modeling, implementation, and test-

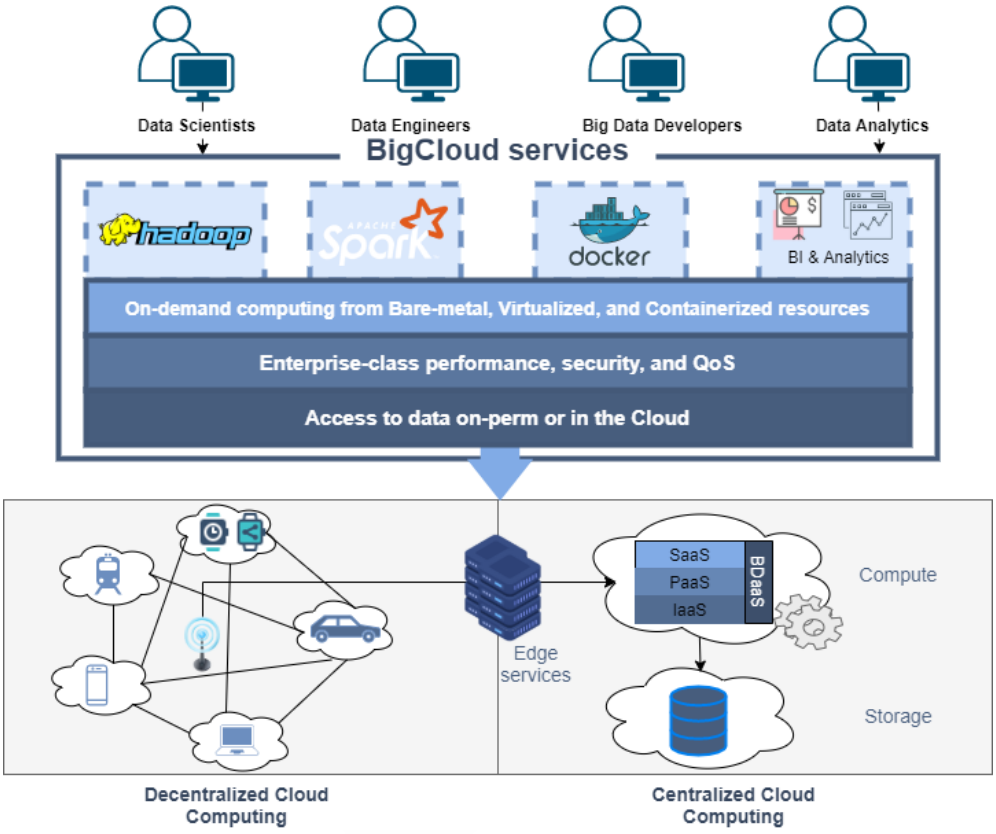


Figure 1.1: BigCloud abstracted layers and architecture.

ing of cloud computing to cope with the BD environment without redesigning or modifying its IaaS structures. Also, toward improving Cloud-based Big Data (simply BigCloud) frameworks security in storage, motion, and process. The investigation of BigCloud security challenges, detailed in the next sections, identifies a research gap in the area of security by design of BD systems and how to evaluate the system security components. Figure 1.1 shows a diagram that illustrates our vision for the BigCloud software platform as a "Big Data as a Service" for Big Data frameworks deployments, whether centralized and decentralized architectures in the public cloud.

This thesis researches two main problems related to modern cloud security, as illustrated in the tree diagram in Figure 1.2 that illustrates the scope of this research. It also illustrates the

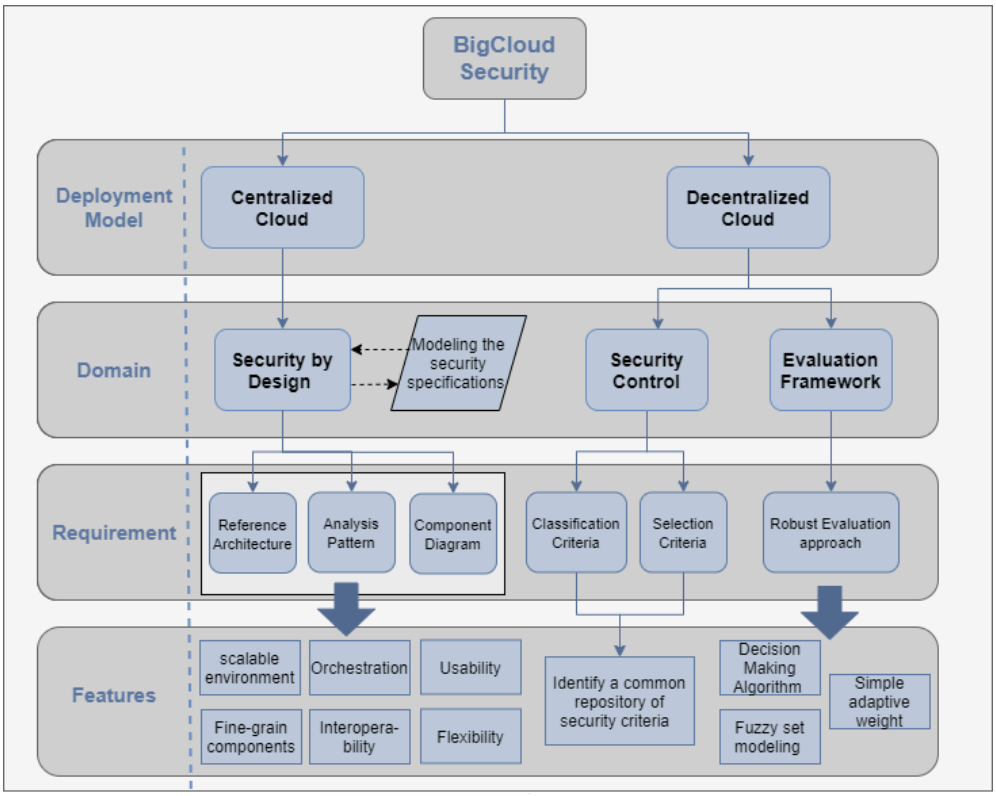


Figure 1.2: Modelling the thesis structure and components.

research flow between the thesis domains as well as the associated requirements and features. The first problem is related to meeting centralized cloud security at the design level. Security by Design [66] is increasingly becoming the mainstream development approach to ensure the security and privacy of different deployment systems. It indicates that the system has been designed from the foundation to be secure. In such an approach, the alternate security tactics and patterns are first thought; among them, the best is selected and enforced by the architecture design, and then used as guiding principles for developers.

The second problem discussed in this thesis is related to developing a security evaluation framework for one of the most trending challenges in the modern decentralized cloud deployments architectures, in particular, the vehicular clouds that represent a special case of the Intelligent Transportation Systems (ITS). This use-case represents a data streaming approach

towards the cloud. Also, we discuss the critical realization of security by design, security control elements, and an evaluation approach in the modern computer science.

1.2 Overview: Supporting Concepts

Cloud-based services are gaining more recognition in the IoT applications, thanks to the evergrowing cellular network technology. Several research works aim at integrating cloud computing with different industries. Herein, we provide an overview of the supporting concepts, and we define the thesis use-cases.

- **Vehicular Cloud (VC)**.

  VC refers to a group of broadly autonomous vehicles whose corporate computing, sensing, communication, and physical resources can be coordinated and dynamically allocated to share internet access, and hence data, with other devices both inside and outside the vehicle. The VC can be formed by autonomous vehicles and provides a vast number of applications and services that can benefit the entire transportation system, as well as drivers, passengers, and pedestrians. The intent to utilize the excessive onboard resources in the transportation system, along with the latest computing resource management technology in conventional clouds, has cultivated the concept of the VC. In general, it is composed of (i) Vehicular Ad Hoc Networks (VANET), which is a wireless ad hoc network of mobile vehicles, in which communication can be between vehicle and vehicle (V2V) or vehicle and roadside infrastructure (V2I). (ii) Connected vehicles that interact with each other (V2V), the roadside infrastructure (V2I), and beyond Vehicle-to-everything (V2X) via wireless communications. (iii) Cloud computing service provider to support many novel applications.

- **IoT-to-Cloud paradigm**.

  As the Internet of Things (IoT) has grown in popularity, modern cloud computing providers have begun offering IoT-specific features. These features could provide IoT developers with much aid, make it an essential part of the modern IoT applications. Nowadays, we can see more and more emerging IoT applications on the market that have the cloud as back-end. The reason for this trend is the enormous benefit that the cloud computing paradigm provides to IoT. These benefits include but not limited to scalability (computing and storage), data mobility (and accessibility from anywhere),

cost-effectiveness (pay as you use), and many more features. Thus, IoT and cloud computing was a natural pairing, enabling advanced predictive analytics, historical data analysis, and automation, the likes of which the world has never seen before. When paired with edge computing, the cloud offers the most significant business benefit and facilitates data integration.

1.3 Motivation

In this thesis, we establish a security framework relying on the common security criteria within the BD platforms over the cloud deployment architectures. This framework aims at improving cloud vulnerabilities and hence addresses its challenges. Specifically, we examine the common features and the security challenges of this integration to provide an architecture relying on the security of the network. We focus on improving the large-scale data processing security issues in both the centralized and decentralized clouds, and the Cloud-based Big Data frameworks security in storage, motion, and process. We achieve these goals by implementing the best guidelines and practices for managing security related to BD operations over cloud computing technology and updating industry security guidelines, frameworks, and standards.

1.4 Problem statement

Cloud computing is changing the way IT industry is delivered in enterprises around the world; it is consumed and managed, improving cost efficiencies, availability, accelerated innovation, faster time-to-market, and the ability to scale applications on demand. On the other hand, BD analytics allows getting enormous benefits by dealing with any massive volume of unstructured, structured, and semi-structured content. However, BD processing represents a high risk for many users due to the various data gathered from all available sources. The Apache Hadoop ecosystem had been widely used in the last years as a defacto platform for all BD workloads. These workloads are spanning from batch-processing, e.g., Hadoop MapReduce[1] and Apache Spark[2] to real-time processing frameworks, e.g., Apache Flink[3] or Apache Storm[4]. As a large-scale system, Hadoop reliability and availability are significant issues not solved in the standard implementation.

[1] https://hadoop.apache.org/
[2] https://spark.apache.org/
[3] https://flink.apache.org/
[4] http://storm.apache.org/

Nowadays, cloud deployment architectures have become a preferable computation model of BD operations. Their scalability, flexibility, and cost-effectiveness motivated this trend. In this context, the data is no longer physically maintained under the user's direct control, which raises new security concerns. Security management plays a decisive role as either compliance or promotes widespread adoption and acceptance. However, it is challenging to develop a comprehensive security plan and design unless it is based on the results of a preliminary analysis that ensures a realistic secure assembly and addresses domain vulnerabilities. Employing software engineering technical expertise to address cloud security concerns can sustain this tendency by drafting security models that address these concerns. Recent literature lacks a technology-independent reference architecture for BD systems that sheds light on security by design. Furthermore, a proper definition of the security components that is responsible for delivering reference modules, basic requirements, and central system characteristics of a BD-cloud environment are required. Doing so will demonstrate how to use this model in the development of effective BD security-solutions and evaluation frameworks.

Internet of Things (IoT) is becoming increasingly crucial to intelligent transportation system stakeholders, including cloud-based vehicular cloud (VC) and internet of vehicles (IoV) paradigms. This new trend involves communication and data exchange between several objects within different layers of control. Security in such a deployment is pivotal to realize the general IoT-based smart city. However, the evaluation of the degree of security regarding these paradigms remains a challenge. This study aims to discover and identify common security criteria (CSC) from a context-based analysis pattern and, later, to discuss, compare, and aggregate a conceptual model of CSC impartially. A privacy granularity classification that maintains data confidentiality is proposed alongside the security selection criteria.

The integration between cloud computing and vehicular ad hoc networks (VANETs), namely, vehicular clouds (VCs), has become a significant research area. The trustworthiness in VCs is expected to carry more computing capabilities that manage large-scale collected data. This trend requires a need for a security evaluation framework that ensures data privacy protection, the integrity of information, and the availability of resources. Evaluating the trustworthiness of the security design and implementation is vital for the sustainable integration of such a paradigm. This evaluation can be achieved by regularly checking the security components of the VC to devise an adequate plan for improvement. Any development in this sector will require a rigorous and robust evaluation model. This problem demands to identify the standard security criteria and providing security gap analysis of such deployment

architecture.

The scope of this thesis is restricted to big data frameworks and applications over Cloud environments with its applications. The work, however, is considered suitable for the evaluation and analysis of the cloud environment and the security controls implemented on the cloud service in general. The research scope is in line with the shared security and management responsibilities between cloud service providers and their customers. This thesis is focused on ensuring that BigCloud customer security requirements are maintained by the environment surrounding BD application as in IoT-to-Cloud systems.

1.5 Aims and objectives

There is a critical need to systematically and comprehensively evaluate the security mechanisms implemented to meet the security requirements of BD applications over cloud environments. In this thesis we assesses the compliance, capabilities, and limitations of the security mechanism implemented in the cloud environment. Also, we attempts to test the effectiveness of the framework in capturing customer security requirements in the evaluation and assessment process.

The Cloud presents a complex architecture that needs to be broken down and understood to evaluate security requirements that are expected to be met by security methods and controls. Mainly, this objective aims to evaluate cloud architecture's ability to deliver a system that fulfills the users' security requirements and to identify potential risks on each component of the BD frameworks. Also, it is required a developed reference model that can be adapted to segregate IaaS/PaaS cloud architectures into layers in an attempt to identify its components, and how these components integrate to provide cloud services and security mechanisms implemented in the cloud.

The primary goals of this thesis are:

- **Objective 1** A critical evaluation of BD platforms over cloud architectures.
- **Objective 2** A thorough review of current security and privacy issues in cloud-based Big Data applications/frameworks and their security requirements mapping and classification.
- **Objective 3** A comprehensive evaluation of different techniques regarding BD service security, concerning infrastructure, data, and application.

- **Objective 4** To establish an architecture relying on the security of the network and identify the framework components.
- **Objective 5** To formalize security abstraction, besides reliability and trust measurements of cloud-based Big Data solutions.
- **Objective 6** To propose and intensively evaluate a security evaluation framework for vehicular cloud applications.

1.6 Thesis Contribution

This thesis has successfully achieved the targeted objectives listed in section 1.5 and has achieved the following contributions:

1. The first contribution of this thesis is composing a reference architecture that summarizes the relationship between the security service and other cloud services as well as their functions. Also, we offer a solution to the main research question, which deals with the security elements associated with BD deployment over the IaaS cloud model. Our solution consists of main actors that emphasize the separation of concerns regarding the service functionality (data service security, IaaS security, etc.) and non-functional security requirements at the beginning of the design, through a reuse-based approach and specifications. Moreover, we propose a security analysis pattern that refines the cloud context-pattern in synergy to an extended CIA triad[5], providing a set of guidelines for the structuring of BD specifications, which relates a cloud design to its security environment. Finally, we suggest a structured election method for BigCloud-specific security selection that delivers many insights regarding the latest ongoing developments and cutting-edge frameworks by mapping each security domain to its solution knowledge.

2. The second contribution of this thesis is to review and further discuss the various criteria that influence the degree of security in the Vehicular Clouds (VC) context. The thesis first outlines a conceptual model that provides an overview of the criteria influencing both the client and provider security formation. Next, it creates a security analysis pattern that envisions VC to have both back-end data aggregation channels (e.g., Message

[5]The CIA triad refers to the *confidentiality, integrity, and availability* of data

Queuing Telemetry Transport (MQTT) broker) and edge/fog data services to supplement the sensor mesh network (e.g., VANET) and the cloud architecture layers. It also represents VC privacy as a multi-criteria construct affecting the data life cycle.

This research concludes that the majority of current VC systems fail to provide the security level for their system. We also advocate that, when evaluation and selection are realized, VCs can lead to a significant enhancement in the system security that sustains its adoption by the clients. Thus, we explore the various criteria that influence the security degree in the Internet-of-Vehicles (IoV) to the cloud context.

3. The third contribution of this thesis is providing practical analysis for security by design of cloud-based IoT and VC systems. Ensuring sustainable security integration of industrial VCs in the cloud environment with systematic security evaluation and selection has been limited in this context. This thesis investigates a VC evaluation to offer assurances of the functional security properties of VC deployment architectures. The proposed framework expresses imprecise trust evaluation information to facilitate decision making within industrial VC environments.

   This goal provides a theoretical contribution by categorizing the security evaluation criteria in VC based on the evaluation theory. It is also proposing a better security criteria selection with a fuzzy evaluation and ranking technique to evaluate and classify the unimproved security vulnerabilities in IoV-to-cloud deployment architectures. Therefore, it ensures the trustworthiness of the VC environment.

1.7 Outlines

This thesis is dedicated to solving the centralized cloud security challenges as well as the IoT-to-Cloud data streaming security challenges. Figure 1.3 represent the thesis structure and its components. The first problem is related to the centralized deployment models, where we propose modeling the security specifications in the security by design in Chapter 2. A reference architecture (RA), an analysis pattern (AP), and a component diagram (CD) as security elements of the solution requirements are the main components of this chapter.

The decentralized architectures, i.e., large-scale distributed systems as in a typical IoT-to-cloud security criteria, are discussed in Chapter 3. The specification domain includes both, security control elements for evaluation and election, and a robust evaluation framework of these security control elements within any deployment system. The security control requires

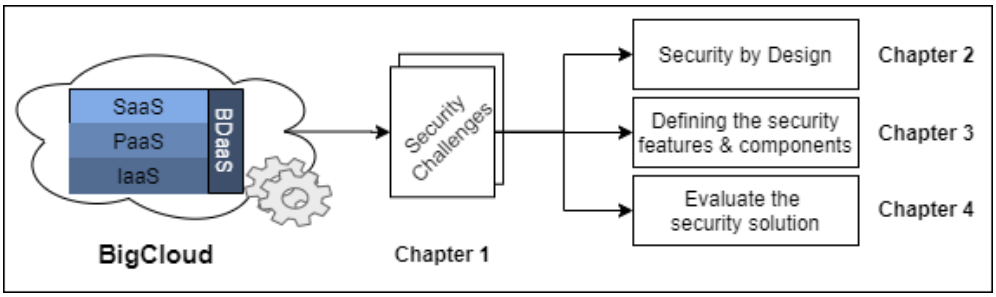


Figure 1.3: Mapping the thesis research components to the thesis outlines.

a comprehensive classification criteria for all of the security components and features, as well as a selection criteria among a set of these security features. An evaluation framework that employs a multi-criteria decision-making algorithm, a fuzzy set modeling approach, and a simple adaptive weight mechanism is presented in Chapter 4. This framework aims at identifying a unified repository of standard security criteria and an evaluation approach of them.

This research introduces an analytical model for BD security measurements within any cloud-based framework. It is aiming to fill the existing gap between the statistical representation of quality approaches of software engineering and the analysis of securing BD applications. Also, this work recommends best practices and measurements when constructing BD systems in both centralized and decentralized clouds. The proposed model, thus, provides a comprehensive and fundamental basis to optimize the design of BD frameworks regarding security ultimately.

CHAPTER 2

SECURITY BY DESIGN: CONCEPTS AND IMPLEMENTATIONS

2.1 Introduction

In this new digital era, many companies use cloud technology to store, process, and analyze petabytes of both structured and non-structured data relating to their business and customers [54]. Advancements in cloud computing (or simply cloud) technology have shaped the modern application delivery model [113]. The advantages of adopting cloud computing are inarguable due to its great potential to provide cheap and straightforward access to substantial computing power. This paradigm shifts the location of the datacenter to an off-premise location, with the potential of higher substantial scalability and elasticity than traditional models. Big data (BD) frameworks over cloud computing (BigCloud) promote the transference of data to off-premise datacenters. As the outsourced data usually contain confidential information, such as sensitive financial records, proprietary research data, healthcare data, or sensitive government information, information classification and security become even more critical.

Fulfilling the highest security and data protection requirements among all cloud delivery layers is a common objective. Nevertheless, ensuring the security properties of computation outsourcing to the cloud can be challenging. For this reason, security and privacy of data management are among the cloud's leading next-decade research directions [30]. This research direction aims to maintain the efficiency of sustainable BD operations over cloud systems.

The main security challenge pertains to the client's trust in data transfer in and out of

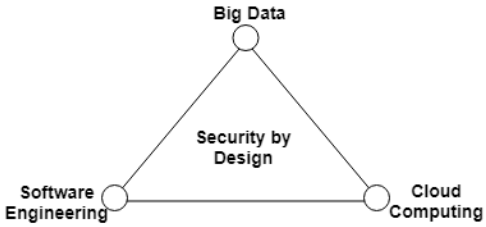


Figure 2.1: BigCloud Security by design domain

the cloud environment as well as the storing and processing of critical data within an off-premise datacenter. Some fundamental characteristics of the cloud (such as multitenancy and virtualization) ensure better utilization of resources, but make it challenging to deliver secure computation. Other security threats associated with this adoption include privacy, integrity, confidentiality, and the availability of stored data, which are magnified by the properties of BD systems (i.e., volume, velocity, and variety) [54].

Numerous public cloud service providers deliver their service within different layers, such as infrastructure, platform, and software-as-a-service (IaaS, PaaS, SaaS). This study focuses on the IaaS cloud, where both the user and the service provider share the responsibility for their BD stages —data-at-rest, data-in-transit, and data-in-process security. Our research domain is distributed over three main computer science fields, namely, big data, cloud computing, and software engineering as represented in Figure 2.1. In this regard, a reference architecture (RA) that illustrates security components in the context of the BigCloud conceptual model will ensure the representation of the problem domain (i.e., security, reliability, and privacy). This representation serves as a mechanism to transfer knowledge of security modeling and software engineering tools to that problem domain. It also serves as a knowledge capture, which contains both domain knowledge (e.g., using cases and scenarios) and solution knowledge (e.g., mapping current technologies) that present the domain under study.

2.1.1 Contribution

This work contributes to the adoption of secure IaaS cloud models as scalable BD deployment architectures by taking the following measures:

- Foregrounding the security components and essential qualities of a BigCloud framework;

- Mapping and categorizing current security technologies onto the concerns;
- Defining domain concepts based on the grouping of relevant concerns into an architectural reference module that facilitates the design of security systems;
- Demonstrating how to utilize the reference architecture as an effective medium to create and evaluate secure BigCloud systems with a sufficient trust level.

2.1.2 Organization

The structure of this chapter is as follows: Section 2.2 presents a general review of the chapter's scope and background alongside the methodology and related work. Section 2.3 outlines the reference architecture as part of the cloud security management. Section 2.4 provides a component diagram of BigCloud security processes and attributes. Further, Section 2.5 discusses security-related considerations, including service-delivery, security and data service security. Section 2.7.1 discusses converging this study's results in a security evaluation framework. Finally, we conclude in Section 2.7.

2.2 Background

Big data platforms use an architectural pattern that guides data-intensive solutions to create, organize, and reuse their computing components. Meanwhile, cloud computing is a set of enabling technologies that provide broader services and more flexible solutions for enterprises to deploy their frameworks. This section discusses the relationship between BD platforms and cloud computing service providers. Further, it introduces the methodology and motivation behind this work. Finally, it discusses related work.

2.2.1 BigCloud

Historically, big data deployment architectures have been designed as shared-nothing architecture with enough capacity to meet peak demands. However, this architecture could result in the system underutilizing its capacity that organizations must still pay for. On the other hand, once the system's capacity has been reached, a significant investment in time, resources, and money to expand it is expected. Modern industry and academia require "utility" services, through which they can scale capacity vertically and horizontally on demand and pay only for what they use.

The advent of cloud-based clusters promotes implementing a cloud solution to support BD operations. This approach grants a practical solution that not only tackles this challenge but also enhances the system's scalability, reduces maintenance cost, and increases the efficiency of resource management. Over the years, cloud service providers have offered a wide range of BD-supporting services spanning from storage to processing and analyzing vast amounts of datasets. Examples include public-service providers (e.g., Amazon EMR [12], Microsoft Azure HDInsight [82], and Google Cloud Dataproc [49]) and private BD vendors (e.g., Cloudera [34] and MapR [78]).

2.2.2 Motivation and Methodology

This work aims at facilitating the realization of secure BD systems in the IaaS cloud model. When a BigCloud system is realized, important security considerations arise. These security factors include the architectural design of the system and the underlying security technologies and policies/services. IaaS continues to be the fastest growing model [77] and the most preferred by many BD implementers. The main goal of this thesis chapter is to analyze the security services used in Iaas cloud environments and describes BD security items and relationships amongst them. It discusses security systems oriented to BigCloud design in order to present their glossary and landscape techniques and to define research gaps and best practices. Figure 2.2 describes the methodology employed in this study to deliver a generic BigCloud security reference model.

In detail, this study contributes to the BigCloud security deployment body of knowledge by, first, extensively examining the building blocks of the cloud security stack for supporting BD science. In addition, it classifies the different layers of security based on their supported service models into a reference architecture. Second, it examines the vulnerabilities associated with BigCloud adoption by providing the security components of a secure design pattern and its attributes. Third, it provides various insights into BigCloud security specifications by refining the cloud context-pattern into a novel security analysis pattern. This pattern maps the current technologies to the solution domain by extending the CIA (Confidentiality, Integrity, and Availability) triad. Next, it analyzes and classifies the state-of-the-art security frameworks available today mostly as open-source for a detailed criteria election. Finally, it highlights some open challenges and recommendations for both service providers and customers, for a comprehensive discussion towards achieving the vision of providing a secure BigCloud service. To facilitate using the systematic research methodology, we summarize

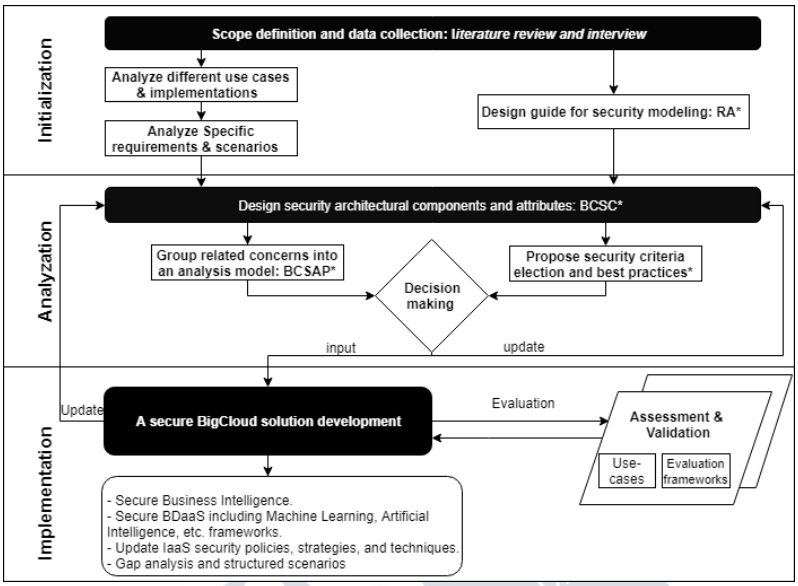


Figure 2.2: A graphical representation of systematic research methodology for modeling and implementing a security solution for BigCloud adoption. The abbreviations and (*) signs represent the chapter scope, and they are further described in Table 2.1

the proposed models and patterns in Table 2.1, which consists of the knowledge domain, its model description, and its designated section within this study.

2.2.3 Related Work

The literature has suggested different Cloud-driven meta-models to support cloud application management. In [52], Hamdaqa et al. propose a service-oriented architecture that captures design elements, configuration rules, and a semantic interpretation of cloud applications in a meta-model. Their work meets the goals of this chapter in that both works aim to standardize cloud-modeling language by drafting reference models. However, the proposed model of this study formalizes cloud security's vocabulary and semantics, which assists in developing a secure BD service-oriented model and suitable cloud runtime security support. In the context of the cloud-security meta-model, the authors in [67] present an integrated domain-specific language coupled with a basic security model that promotes the designers' modeling effort. Furthermore, the work by [119] provides two use-case studies to verify the usability

Table 2.1: Representing a summarization of the research scope mapped to the section architecture.

Abbreviation	Knowledge Domain	Description	Section
BCRA	Preparation	BigCloud Reference Architecture with different service layers and security services that describes the security issues addressed by the study.	Section number 2.3
BCSC	Initialization	BigCloud Security Component that reviews the primary components as of security design pattern with security attributes and constraints affect the security problem.	Section number 2.4
BCSAP	Examination	BigCloud Security Analysis Pattern that reviews the bases for building a secure ecosystem during the analysis phase and describes the basic structure and risks to be considered while applying the solution using a UML diagram.	Section number 2.6.1
SCS	Selection	Security Criteria Selection that reviews the various criteria that influence security in a BigCloud context, and describes different ways a security pattern may be implemented and deployed.	Section number 2.6.2

of their meta-model. Nevertheless, none of the previous studies consider BD-specific security requirements of IaaS cloud deployment architecture as is proposed by this study.

A recently established NIST BD security Sub-Working Group (NBDs-WG) [87] addresses the importance of security and privacy measurements, definitions, requirements, and characteristics of BD systems. The mutual relationship amongst BD technologies and model-driven engineering (represented by software engineering) is investigated in [21]. In [90], Pekka and Pakkala analyze published implementations of BD architectures (e.g., Netflix, Linkedin, and Facebook) to draft a reference architecture. In doing so, they aim to map different BD solutions that facilitate designing BD systems and create a classification of BD technologies, products, and services. In [37], the authors extensively illustrate BD ecosystem components based on the NBD interoperability framework. Their architecture framework consists of BD infrastructure, BD analytics, data structures and models, BD lifecycle management, and BD security. However, they do not investigate cloud-specific security requirements, components, or delivery within an IaaS cloud as this study proposes.

2.3 BigCloud Security Reference Architecture

Many security threats regarding BigCloud platforms can be mitigated using traditional security processes and techniques. However, some security threats require cloud-specific solutions. BD frameworks have different security vulnerabilities and may be exposed to various threats. Thus, in addition to setting BigCloud security service requirements and components of data storage, it is significant to define whose responsibility it is to protect them. Therefore, we specify a vocabulary of design elements associated with BigCloud actors (system components) by presenting the BigCloud Reference Architecture (BCRA), which outlines the main components of cloud applications. BCRA summarizes the relationship between the security service and other cloud services as well as their functions. Further, the BCRA model defines a set of implementational requirements and characteristics that can be used for orchestrating a secure BigCloud ecosystem. Therefore, it relates to companion security requirements and features that are the basis for designing a reliable BigCloud implementation. Figure 2.3 illustrates the five major cloud actors of the BCRA framework, excluding the client itself: service delivery, management, auditing, data, and security services. Section 2.3.2 further presents the characteristics used, as well as standards for describing BigCloud security in detail, whereas Section 2.4 and 2.5 present BigCloud security-specific elements and considerations, respectively.

2.3.1 BigCloud reference architecture components

In general, BD reference architectures within the field of software architecture aims to provide a template solution for an architecture for BD domain. It also provides a common vocabulary with which to discuss implementations, often with the aim to stress commonality. In this chapter we introduce the components of the BigCloud reference architecture.

1. Client Security: An entity (organization or user) that has a formal contract or arrangement to maintain a business relationship with a cloud provider to use IT resources and other services made available by the provider. The cloud client security complements the providers' security and components. The client accesses the service by applying a session that defines interaction security, using service level agreement (SLA) and policies. The session establishes client permissions and log method, and it even configures session timeout values. These sessions have the effect of mirroring services across all layers and system components. Assuring the session's availability regardless of whether

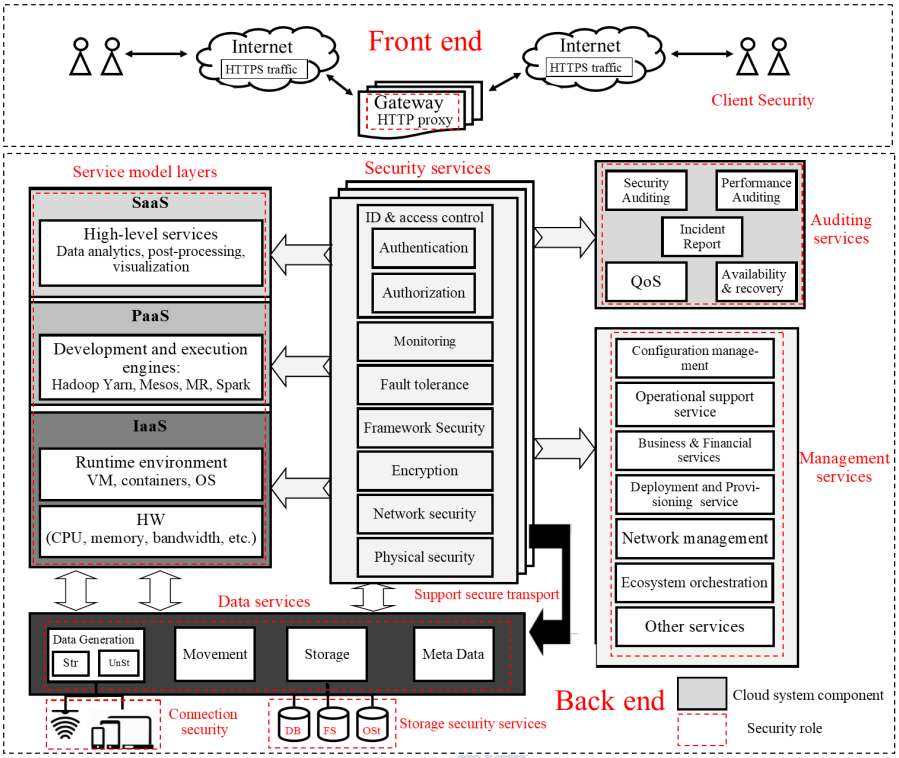


Figure 2.3: The reference architecture as part of the security management of BigCloud specification development.

there is an attack (e.g., denial of service) or a system failure is in the BigCloud service provider's interest, along with securing access, user identification, and authentication. Providing the needed level of training and awareness among users (such as strong passwords) are considered a common interest for both the client and provider.

2. Service delivery: This represents the three types of cloud delivery models in the form of layer abstractions: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). Figure 2.3 shows how each layer defines its role (operation and function) within the BD stack. These roles require input and interaction between customers and the service provider. The service delivery layers serve as a simplified translation of business demand into technology and operational capabilities.

The service delivery model is shaped and presented from a BD viewpoint. It focuses on the BD framework service delivery as cloud-service abstraction layers.

In an IaaS deployment, the capabilities are split into hardware, network, and runtime environment. Fundamental computing resources, such as CPU, RAM, and bandwidth, are at the base of the IaaS cloud. While the runtime environment can deploy and run arbitrary software that includes operating systems (OS), virtual machines (VM), and containers, the client does not control the underlying infrastructure but has limited control to select networking components (e.g., host firewalls and virtual networks). Modern BD schedulers utilize the containers as a runtime environment of their applications [114, 55].

PaaS deployment describes the relationship between the cloud provider and the cloud client. The capability granted to the client is to deploy applications using libraries, programs, services, and tools established by the provider. The development environment consumes the runtime services (VM, etc.) from the previous layer, so the user has administrative rights over deployed applications and configuration settings for the application-hosting environment. For instance, a client can select the BD platform (e.g., Apache Hadoop [35]) and the execution engine (e.g., MapReduce [36] for a batch query, Spark [122] for micro-batch, or Storm [111] for real-time processing).

Finally, the SaaS deployment layer provides high-level capabilities such as post-processing operations and visualization capabilities. However, the client does not manage or control the cloud infrastructure and the development or runtime capabilities. For this reason, only the cloud service provider is responsible for efficiency and security.

3. Management services: The capabilities that enable the management of the service-delivery model. Typically, these are services to which the provider connects rather than the client. They refer to a set of services designed to ensure that other cloud components are working optimally for BD framework operations. Moreover, management services present an entity that manages the operation and interaction between the client and the cloud provider. Thus, it is critical to maintain the same security levels for the service delivery layer as for large-scale security monitoring [79] and continuous system-security auditing [73], while retaining authentication and authorization access control. The cloud provider must assure and maintain overall proactive security governance of management services. Different BD frameworks can be implemented to harness man-

agement services, for instance, Apache Zookeeper[1] for BD ecosystem orchestration or Apache Ambari[2] for cluster deployment and provision service.

4. Auditing services: This includes the assessment of cloud services, operations, performance, and security auditing of cloud implementation [79]. It also assures system availability, quality of service, and recovery plans. Security auditing defines and reports on security policies (e.g., password complexity levels). Furthermore, it evaluates recovery policies and the quality of security services while maintaining the reporting of security incidents. A multi-replica dynamic auditing of public multi-tenant data storage on cloud computing is reported in [74].

5. Data services: The underlying data service provides storage capacities on demand, either within virtual disk drives using a hypervisor and containers or with direct access to physical storage. The tasks associated with these services include all data stages, from data collection (generating structured or unstructured data) to data in rest within file systems (FS), databases (DB), and object storage (OSt) as illustrated in the Figure 2.3. These services also include data movement, also known as data placement, from storage to virtual machines and vice versa and other data operations and processing services, such as storing meta-data. The importance of these services is magnified in data-intensive batch-based systems (e.g., Hadoop MapReduce). Since data must materialize in storage before the process can begin, these services must provide the capability to backup and restore data by establishing data protection policies at the service layer.

6. Security services: They define a broad set of technologies, policies, and controls deployed to protect data, services, applications, and the associated infrastructure resources of cloud computing. By managing the on-going delivery of security, these services represent the capabilities of the security life cycle. The RA verifies that the BigCloud security is a cross-cutting interest that influences all the components in the model.

2.3.2 BigCloud Security Characteristics

Due to its inherently remote operations, resources co-tenancy, distributed management, and administrative control, ensuring the privacy of BD workloads while outsourcing computation

[1] https://zookeeper.apache.org
[2] https://ambari.apache.org

is crucial. Customers do not have direct control over the systems that consume their data because of the cloud's black-box nature. The following are the most pressing challenges in assessing data protection before a move to the public cloud:

- Data residency: This refers to the physical geographic location of the data stored in the cloud. In conventional BD systems, such as on-premise clusters, the geolocation of data is always known and, thus, controlled. When deploying a BigCloud system, the physical location of the data is no longer known or fully trusted. Data residency also includes data flow, file locations, and data input/output.

- Data privacy: This describes the ability to limit data sharing in BigCloud systems, including third parties through an organization or individuals. Maintaining an appropriate data privacy level can be achieved by exploring various technologies and tools, including encryption [71] and virtual mapping [32]. Other solutions include modifying policies and legislation to prevent unauthorized access or use of data. However, defining legal ownership, responsibilities, and privileges of data between owner and data custodian can alleviate privacy threats.

- Data ownership: A serious concern within BigCloud data processing is data ownership. When clients transfer their data to the cloud, the primary processor of that data is then not the physical owner but the provider. Consequently, a new threat parameter is raised regarding trust in that provider. Clients cannot be sure how the cloud system manipulates their data or whether the processing complies with their demands.

2.4 BigCloud Security Component Model

Any model consists of a vocabulary of design elements, a set of configuration rules, and a semantic interpretation. The technology-agnostic BigCloud Security Component (BCSC) model represented in Figure 2.4 is a logical extension of BD application security in cloud computing definition. As highlighted earlier, BCSC is a generic, high-level conceptual model that facilitates the understanding of the successful implementation of trusted BD in a cloud environment. From this perspective, it summarizes operational intricacies and component interaction of BigCloud security. The BCSC does not represent the system architecture of a specific cloud vendor. Instead, it is a framework for describing, evaluating, and developing

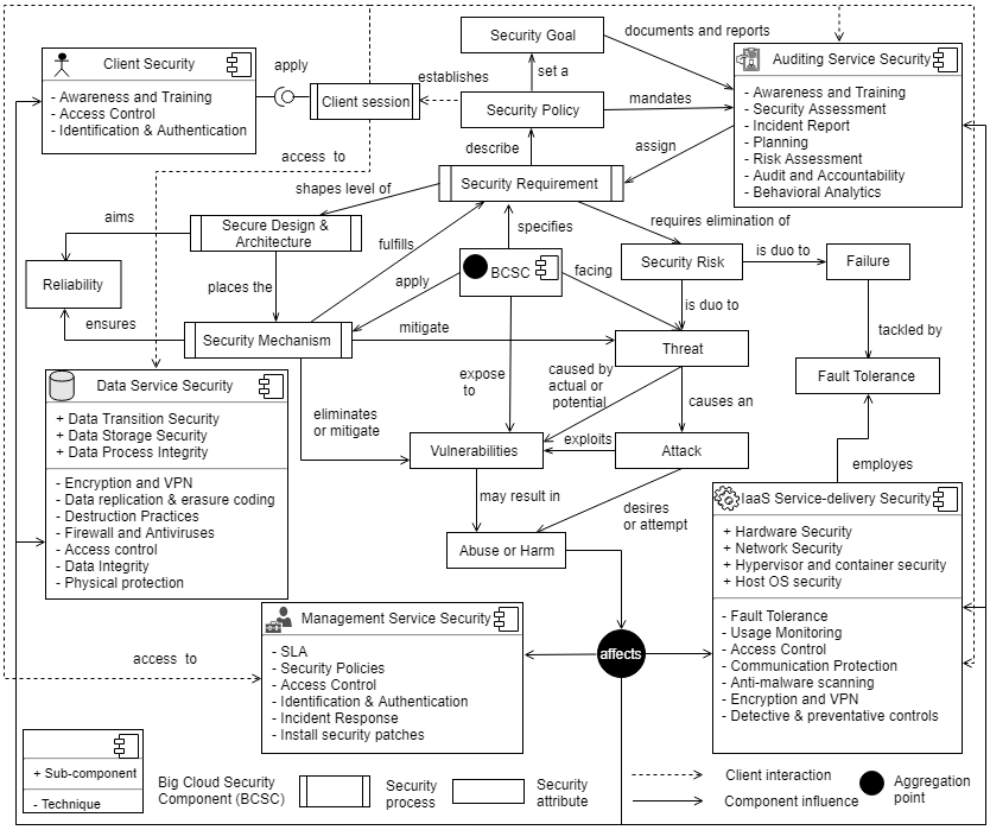


Figure 2.4: BCSC diagram as a structure of software design pattern with associated security attributes.

a system-specific architecture using a shared cloud security component of reference models, along with their activities and functions. This actor-based model is intended to serve designers by representing the overall view of roles and responsibilities for the assessment and management of risk by implementing security components and controls.

The BCSC model specification relies on the predefined BigCloud RA components described in Section 2.3 By analyzing these components, we conceptualized a BCSC model to represent the implementation of security-specific techniques and sub-components. The proposed model demonstrates the security process and attributes (actions) that influence, and are influenced by, the BCSC and describes the structural relationships among them. In

fact, the BCSC model is a logical design constructed with replicable items (i.e., the sub-components and security techniques can be modified according to use-case). This approach ensures reuse and substitutes these components and actions within any BigCloud application security design. It also offers the architects a reference model to verify that the installed security plan/design meets their system security functionality. Moreover, it can be utilized as a communication tool for various development groups, as well as project stakeholders and implementation staff, as it provides a high-level, architectural view of BigCloud security. This model assists considerably in formalizing the implementation roadmap for security integration.

The BCSC model consists of five main components: client, data service, management service, auditing service, and IaaS service-delivery security. Sections 2.5.1 and 2.5.2 further discuss security-related considerations of targeted sub-components. The BCSC model is also composed of three security processes and two aggregation points. These processes (described in Section 2.5.3) represent a set of techniques, tools, and methodologies to achieve their goals. On the other hand, the aggregation points facilitate the diagram with a decent and uncomplex view. The BCSC aggregation point, located in the center of the model, resembles integral components (i.e., it can be implemented among all of them). The relationship between components and other entities is represented by two independent arrows, where the arrowhead connects with the provider. The continuous arrow resembles the entities' interactions (or influence), while the dashed arrows represent the client's interaction.

2.5 BigCloud Security-related considerations

This study offers an extensive analysis of BigCloud IaaS service delivery and data-service security. However, due to the vast research area, an in-depth examination of management, auditing and client service security are not within its research scope. These items offer material for future work or an open research direction. Herein, we examine the implication and remediation of the most relevant security components in the BigCloud paradigm: IaaS service-delivery security and data-service security. This section discusses other security considerations and processes when implementing a BD system within a cloud development model. These security threats mainly originate from issues such as multi-tenancy, loss of control over data, and trust [47].

2.5.1 IaaS Service-delivery Security

The Cloud service delivery model represents the technology stack in which each layer provides services to the layer above. In this context, IaaS provides the base to deliver other services with the cloud model, e.g., PaaS and SaaS. The reference model categorizes security services among the IaaS layer to a runtime environment or hardware and network components. The BCSC diagram in Figure 2.4 specifies these components in detail by charting out the IaaS service-delivery layer as follows:

- Hardware security: Hardware resources (e.g., CPU caches, GPUs, and RAM) deliver their services in a scalable way by sharing infrastructure. The underlying resources that provide this infrastructure were not often designed to offer robust isolation features for multi-tenant architecture. A virtualization hypervisor or container mediates access between guest operating systems and these computing resources are utilized to address this issue. Security measures should still be employed to ensure that individual customers do not impact the processes of other tenants operating on the same cloud provider.

- Network security: Provides the network connection that supports IT activity, which includes network fabric, virtual local area networks (VLAN), connectivity, and segmentation. Network services are responsible for delivering clients' data to storage capacities and linking system components. They also support a secure movement of BD meta-data and pass the workload to the process units among all service-delivery layers. Connection security is a critical factor in securing the delivery of services, as network and communication carriers provide the distribution of any BigCloud services. Due to its significance, the network architecture design should treat client connections with a minimal level of trust. Clients will always access cloud services using a remote network connection. A set of security measures should be employed to mitigate channels (and network services such as DNS) that transmit data to and from cloud structures. Secure sockets layer, transport layer security encryption, and VPN technologies are examples. Firewalls should also deny any attempts to access a BigCloud service from a session that should not be connected to that service.

- Hypervisor and container security: Virtualization technology is a technique that allows multiple OS running concurrently on a host environment. It is also a resource

abstraction component that ensures efficient and reliable usage of underlying physical resources, including computation and storage. This resource abstraction acts as a security component by itself, using proper configuration and permissions. A cloud provider would utilize hypervisors or containers (or both) for resource pooling. Doing so provides and manages secure access to its physical computing resources, among other advantages. The security aspect behind this subcomponent refers to access control (ensuring authorized access to services, data, and other components) and usage monitoring. The security topology should not expose user interface service functionality to non-privileged users. Malware scanning should follow that access control, ensuring comprehensive security monitoring of the whole environment.

- Host OS security: Operating System security commonly involves configuring the host OS that supports the virtualization environment. As with all OS configurations, a fundamental approach is to reduce the attack surface to an acceptable level. For instance, OS images used by a cloud provider can introduce risks to the cloud client when using pre-owned virtual machines. The main threat arises in uploading images with built-in Trojans. Thus, the authentication level to minimize the risk will depend on the overall risk strategy and threat surface model.

2.5.2 Data Service Security

Data service security includes data protection and monitoring in the three stages of the data-security lifecycle, namely data-at-rest, data-in-motion, and data-in-use, as follows:

- Data storage (data-at-rest) security: Data storage's primary capabilities include managing the storage required by BigCloud frameworks. However, modern cloud storage components can provide backups services (including virtual storage). These backups may consider a remediation technique that promotes the utilization of cloud capacities as a storage service. Another remediation technique that the storage component can collaborate with the hypervisor or container is to allow for workload migration and storing metadata among host compute nodes. Data storage security services must cover file systems, databases, and object storage security scanning (data content discovery) to identify and locate sensitive content (e.g., credit card numbers). This method supports data compliance and auditing efforts by providing comprehensive reporting on the effectiveness of data storage protection mechanisms. It also guides decisions on security

measurements for implementing data encryption (disk-level encryption) and masking, removing, or warning the file owner. To avoid this data locality problem and address the fault tolerance issues, Hadoop put forth an efficient data replication scheme in the HDFS [98]. Therefore, keeping up a similar replication instrument for each information record prompts using them in terms of recovery or comparing records for any breach. In general, data-at-rest is considered more vulnerable than data-in-transit [53]. However, Hadoop standard configuration does not provide encryption functionalities at their Distributed File System (HDFS), which leads to the generation of Hadoop security-complementing ecosystems.

- Data transfer (movement) security: Data transfer can be classified, based on the connection domain zone, into internal and external data movement. First, internal data transfer occurs between storage capacities and processing units. This communication usually takes place at the platform layer. The BigCloud should consider the internal network as an untrusted network alongside the Internet. Hence, all data transfers (including the meta-data) should be handled with the same level of minimal trust. However, ensuring a high-level of security requires: a) encapsulating the data workloads; b) sniffing the traffic on the network using proxies (to identify the content); and c) monitoring, reporting, and blocking abnormal bandwidth usage (using central policies) based on the traffic type.

  Second, the external data transfer occurs between the client and the BigCloud provider. Here, the network acts as an intermediary that provides data transport using different communication methods, from dedicated network channels to the open Internet. Using the Internet is still the dominant pattern as it cuts costs. In this case, it is the client's responsibility to recognize the full set of security measurements to secure the data migrated to the cloud, as data can be intercepted in transit. On the other hand, the cloud may require the network provider to provide secure connections between it and its clients to reduce vulnerabilities (e.g., man-in-the-middle attack) in Internet transmission channels to a minimum. The network service provider should maintain security control points, maintain security testing, and prevent suspected tasks.

- Data processing security: This refers to securing the processing environment. High-reliability data execution may be achieved by: a) harnessing robust distributed file systems permissions, and b) enforcing isolation among computing instances, workloads,

Table 2.2: Comparison of key security mechanisms of data stages

Data Security Lifecycle	Access Control	Data Integrity	Data Destruction	Physical Protection	Erasure Coding	Encryption	Firewall and Antimalware
Data storage	√	√	√	√	√	√	√
Data transfer	√	√				√	√
Data process	√	√			√	√	

and applications. Therefore, it is ideal to protect platform/application configuration file(s) with appropriate access control. Doing so will prevent the attacker from modifying these critical settings. According to a classification of malware attacks in IaaS execution environments [96], 71% of these attacks target the hypervisor denial-of-service. In contrast, fault tolerance is the most important aspect when discussing data processing security. To support high reliability and availability of BD operations, data blocks used to be duplicated across multiple nodes. This traditional approach was costly and returned with a moderate performance in massive operation scales [118], which lead to the advancement of modern large-scale distribution storage systems with erasure coding techniques. This storage technique provides the same level of fault tolerance with much less storage space and has been implemented with the HDFS [15]. Also, providing cell-level encryption for HBase on runtime was reported in [25].

2.5.3 Security Processes

Security mechanisms

Table 2.2 maps the data life-cycle stages to the primary implemented security mechanisms. The table illustrates that data-at-rest is considered the most vulnerable, so it requires a larger number of security mechanisms. Both data-replication and erasure-coding techniques consider fault tolerance utilities while an encryption and protection of the integrity of data in the transition stage is expected. The use of an adequate data sensitization technique to deliberately, permanently, and irreversibly remove or delete data after ending the service contract must be set at the service-level agreement. Preserving composable security for high-level abstractions of data analytics and mining is also of security interest. On the other hand, secure computations in distributed data-processing frameworks that are deployed over decentralized

clouds (e.g., edge and fog clouds) should be considered within the security design. In the meantime, these architectures demand real-time security and compliance monitoring.

As mentioned earlier, Hadoop standard security configurations may lead to several vulnerabilities. This issue can be tackled by utilizing the latest Hadoop 3.0 secure model, which consists of a service level of authentication and authorization [16]. The security issues of BD authentication are extensively discussed in [2]. Traditional data encryption may be employed at different layers according to Hadoop 3.0 [17] namely, application-level, database-level, filesystem-level, and disk-level encryption, in which HDFS-level encryption is placed between the database and file-system-level encryption. Accordingly, HDFS continues to provide reliable performance, while BD frameworks run safely over encrypted data. This encryption level limits the runtime level attacks as the OS interacts with encrypted data blocks.

Other BD-specific tools and techniques to improve the security ecosystem may include, but are not limited to the following :

- Apache Knox gateway [18] over HTTP/HTTPs, which provides perimeter security with REST API authentication and control access gateway utility for Hadoop clusters and ecosystems. Apache Knox may also provide end-to-end wire encryption using a keystore to hold the SSL certificate;
- Apache Ranger [19], a security orchestration framework with centralized administration and User Interface (UI) to enable, monitor, and manage data security across the Hadoop Yarn clusters;
- Apache Sentry [20], which provides the ability to establish fine-grained (role-based privileges) authorization on both users' and applications' data and metadata within Hadoop clusters.

Security design and architecture

It is essential to address the BigCloud-specific security demands to completely illustrate the various security components in a conceptual context. These demands may be summarized as follows:

- Continuous vulnerability assessment and remediation;
- Data recovery capability;

- Maintenance, monitoring, and analysis of audit logs;
- Automation data protection.

After identifying the security components and functional requirements for the adoption of BD in the Cloud, it is of research interest to highlight security design guidelines when prototyping a BigCloud system:

- Component-based architecture: Quickly add new behaviours.
- Highly available: Scale to very serious workloads.
- Fault tolerant: Isolated processes avoid cascading failures.
- Recoverable: Failures should be easy to diagnose, debug, and rectify.
- Broad network access.
- Decreased visibility and control by the client.
- Dynamic system boundaries and commingled roles/ responsibilities between client and provider.

Vulnerabilities

Vulnerabilities are exploitable system bugs, and they can be exposed remotely across all cloud-service delivery layers. Attackers mainly target the vulnerabilities within the operating system (system kernel, libraries, and application tools). Hence, all services, components, and data face significant risk. Plenty of remediation mechanisms, spanning from planning a secure design to performing compliance testing to validate the security measurements, may be implemented. Moreover, modeling risk patterns and vulnerability scanning, followed up by installing security patches, can mitigate security gaps, as appropriate risk patterns can capture most vulnerabilities [91]. According to [120], over public clouds, Hadoop suffers from an overloaded authentication key and the lack of fine-grained access control at the data access level.

Table 2.3: Pattern relation BCSAP to the Cloud pattern from [24].

Direction	BigCloud to Cloud Pattern
Relation Type	Refines
Reasoning	The services deployed in a BigCloud can be created or composed based on cloud patterns. Thus, the information in the cloud can be translated as refinements descriptions of the service in our BCSAP model. Specific technologies that form the core of the BigCloud resources pool (e.g., storage) are mainly based and rely on the cloud-architecture. Moreover, the cloud architecture is essential for the mapping of BigCloud security specification design.

2.6 Structured selection of BigCloud Security services

Security election is a control element that shapes policies, practices, procedures, and responsibilities of IaaS cloud provider.

2.6.1 BigCloud Security Analysis Pattern

Figure 2.5 presents the elements and concepts of a secure BigCloud model and the relations among these components. We propose the BCSAP patterns for a structured domain knowledge election using the context-election pattern proposed in the cloud system analysis pattern [24]. Furthermore, the relations between existing context patterns and the BCSAP are defined in Table 2.3. The BCSAP shows a meta-model that forms a uniform basis for the current and future BigCloud security deployment. The generalization of its elements creates the basis of a pattern language that ships other deployment architectures (e.g., fog and mobile edge-cloud security).

2.6.2 BigCloud Security Requirements Election

Hosting data off-premises increases the number of potential security risks. By considering the security as mentioned in the BCSC, we divided the security issues, requirements, and characteristics of BigCloud secure deployment. The following sections give details. Figure 2.6 is a model designed to guide policies for information security life cycle within BigCloud systems.

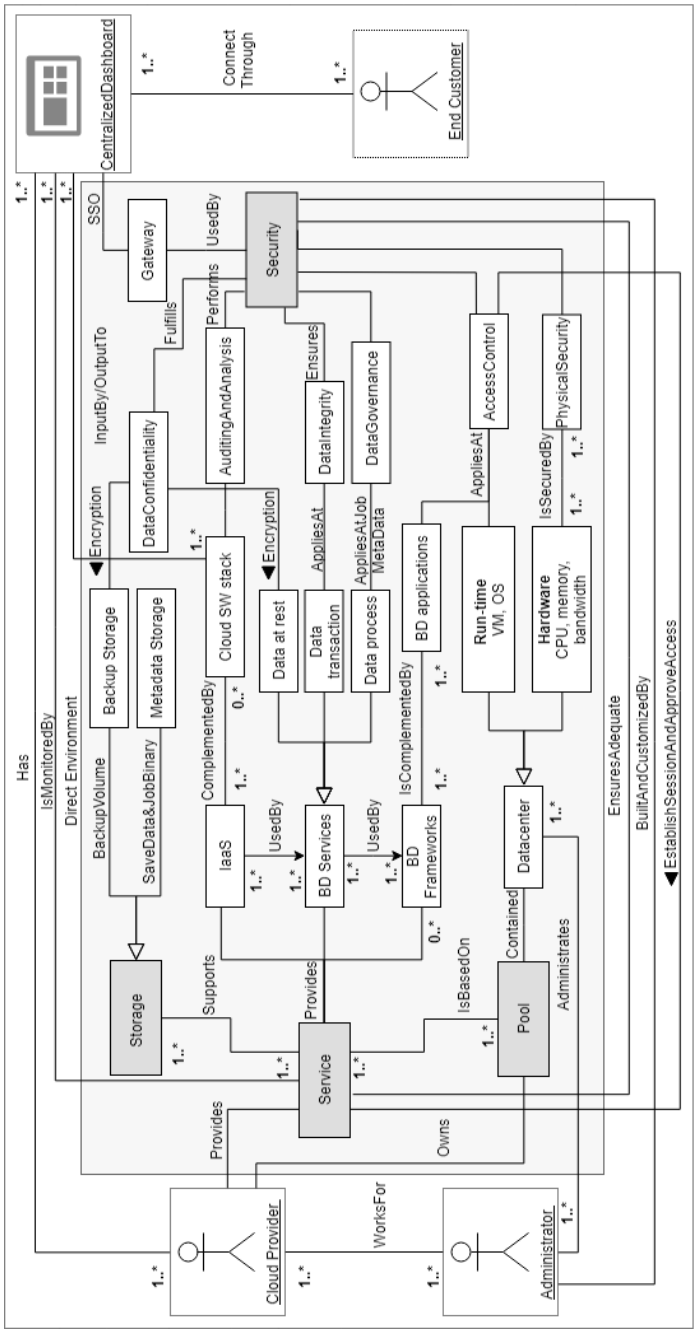


Figure 2.5: BigCloud Security Analysis Pattern (BCSAP).

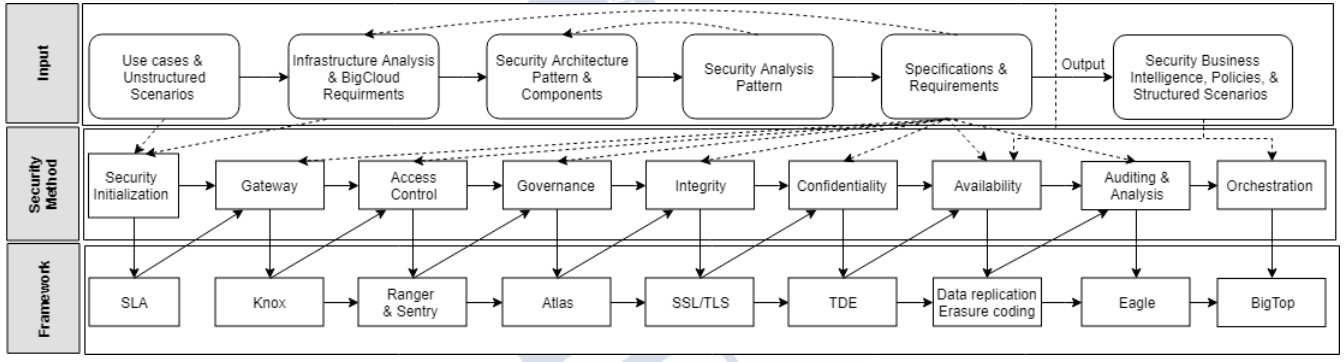


Figure 2.6: Modeling the security requirements election by instantiating the BCSAP and Apache security frameworks with core components highlighted.

Security initialization

This policy establishes the level of security design by describing security requirements through analyzing the security goals and use cases. This process is implemented using a service-level agreement (SLA) with an IaaS client.

Gateway

In a cloud deployment architecture, where a multi-tenant environment is the dominant model, it is critical to control the client's access to the internal cloud entities (resources and services) by defining which users and groups have access to a specific entity. In the case of BigCloud, these entities are represented in the direct environment. A cloud gateway in this context is a single sign-on layer that links the client requests, to the authorized entity of a cloud SW stack securely and efficiently. This layer verifies the external client's access to the system using their user ID and passwords. Every client username and IP address has to be in the client's host file (/etc/hosts) or DNS table, and it has to match the client's given password that can use to map a connection between an IP address and domain names. Hence, it assists in addressing network nodes in a computer network. This process may also include Apache Knox, a unified gateway framework for Hadoop services and ecosystems that can be utilized as an SSO gateway. When connecting to a BigCloud cluster, there are several methods of authenticating the access. For instance, using a simple username/password identification approach. Another method is an authentication using Kerberos protocol (authentication based on tokens). Each client and service must be authenticated by using the Kerberos keytab file (binary containing the information needed to log) to initialize trust between a client/application and the BigCloud components. Authentication for access to the Hadoop services web console requires enabling HTTP SPNEGO protocol as a backend for Kerberos credentials. Thus, the two approaches prevent unauthorized access to the stored data.

Access Control

In a BigCloud-based Business Intelligence environment, several user roles need to be enforced at the service level. These roles must be provisioned dynamically to ensure large-scale participation while maintaining access control. This process improves security controls for authentication and authorization and enforces access discussions to meet BigCloud regulatory compliance. For instance, after users log-on to the cluster, the system must assign authoriza-

tions (i.e., access rights over a given service). The system manages access in the context of a specific service, resource, and data functionality provided by the cloud service provider. BigCloud should support a robust set of role-level security that can be utilized to configure the right level of application authorization for different user types, such as defining the users and groups who are authorized to make service calls to cloud storage service. The call will pass the authorization check only if the user making call belongs to an authorized service entity. In general, the BigCloud platform security model supports three levels of permissions within IaaS:

- Application level: Controls which users and groups are able to create, modify, and publish data within a BD application run within a specific execution engine (e.g., Hadoop). A client can submit jobs and query results of a predefined framework with limited access to the data.

- Framework level: Controls which users and groups are able to deploy, configure, and administrate a BD framework (e.g., MapReduce, Spark, Storm) over the given cloud instances. A client may access the runtime variables, paths, add/remove processing features, and change the scheduler (e.g., fair or capacity) and the resource manager (Hadoop Yarn, Apache Mesos, etc.).

- Runtime environment level: Controls which users and groups are able to query and manage the runtime environment (VM, containers, OS, etc.). However, the client does not control the underlying infrastructure but has limited control (based on the SLA policy) to select networking components (e.g., virtual networks) and to select the OS and VM capacities and configurations.

Data Governance

BD sources and types can vary in their nature with multiple data processing patterns generally formulated as trees, graphs, or workflows. BigCloud should enable a client to maintain high data quality throughout the complete lifecycle of the data, with flexible mechanisms that store and access such data sources independently from their specific format. Moreover, metadata formalisms should be defined and used to describe the relevant information associated with data sources (e.g., location, type, format), enabling their access, use, and administration. A common platform is also essential for metadata exchange and storage within the different

elements. This design will assist in supporting policies consistently across the BigCloud components. Apache Atlas provides data governance capabilities for the Hadoop stack and helps in searching, classifying, and managing data [95].

Data Integrity

Storing clients' critical data over a cloud model requires robust data integrity and availability mechanisms. Cloud clients want to ensure that BigCloud provides appropriate data privacy and integrity between all components of the system, as well as the data source with which they communicate. Supporting the appropriate security for these connections is imperative by ensuring adequate consistency and accuracy of data-in-transit. A block of data fetched from the storage file system or database (e.g., an HDFS DataNode) could arrive corrupted due to faults in the storage device, network faults, or buggy software, as well as abuse or attack. Several approaches are implemented to tackle this issue, such as checksum checking on the transit data or wire encryption. Wire-security, for data transfer between web console and client, may be managed via Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), cryptographic protocols HTTP communications. The cloud provider can configure BigCloud Business Intelligence so that all communication between every component of the cloud system, as well as with the web-client traffic, is secured using TLS/SSL. In Hadoop environment, clients are protected by using SSL(HTTPS). SSL configuration is recommended but not required to configure Hadoop security with Kerberos for data encryption on HTTP.

Another implementation of end-to-end encryption relies on providing secure communication over public networks. In this case, all REST APIs offered by BigCloud components (like Apache HBase, Hive [57], and Oozie [59]) are enforced to pass cryptographic protocols. Doing so requires creating a key store to hold the TLS/SSL certificate and set up environment variables. It takes two stages to set up a secure connection. The first one uses digital signatures and asymmetric cryptography for authentication, while the second stage is for data transmission. Figure 2.7 summarize client connection to BigCloud services using a public network. After a secure session is established (steps one and two), both the client (e.g., result query) and BD frameworks (e.g., HBase data fetching) may access the data securely. The same approach applies to authenticate the internal components of communication upon an SLA policy. For instance, SSL certification to secure connection between the access control and the data storage requires either a self-signed or an authority-signed certificate. Thus, admins need to configure SSL on REST server and a universal key-store to hold the SSL certificates.

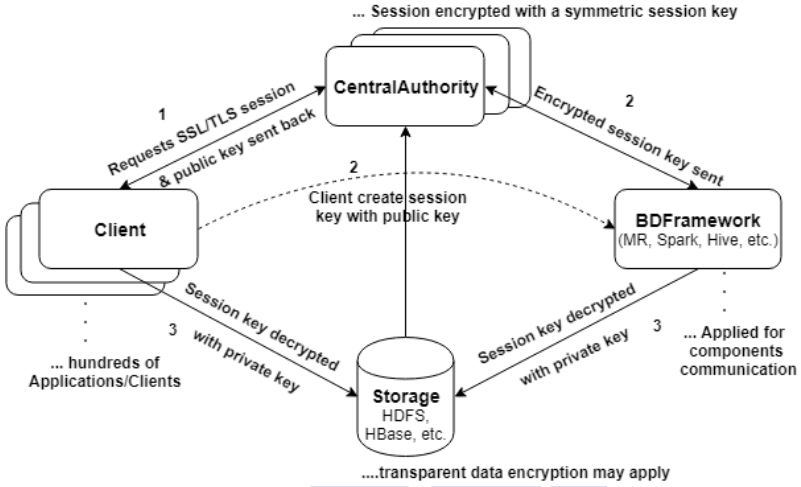


Figure 2.7: Implementing data encryption discussion over BigCloud components.

Data Confidentiality

This requirement assures that a given stored data cannot be reached by any client/application except those who hold permission. Thus, data confidentiality in general aims at preventing protected data from being inappropriately accessed. It preserves authorized restrictions on data access, including job metadata. Cryptographic encapsulation enforcement by using distributed cryptographic protocols, such as PKI and identity/attribute-based encryption, is a common trend. This security layer also includes validity and recoverability approaches as Hadoop 3x starts utilizing erasure coding for fault tolerance. However, aiming for data confidentiality, Hadoop's HDFS implement end-to-end encryption with the so-called Transparent Data Encryption (TDE) [92]. These HDFS encryption sets are at the file-level of on-disk data and are stored as NameNode metadata. Further, HDFS TDE operations rely on encryption zone level of all components of a path, which means all files designated zones are encrypted on disk. In context, transparent "at-rest" encryption implies that the client/application access data without being aware the data was encrypted. It also indicates that data is automatically encrypted and decrypted on-the-fly as it is read or written. However, it is not meant to hide sensitive data (e.g., data masking technique). Nevertheless, security policies like masking can be implemented on top of TDE data as a post-decryption.

Figure 2.8 shows the confidentiality layer components, stages, and granularity levels

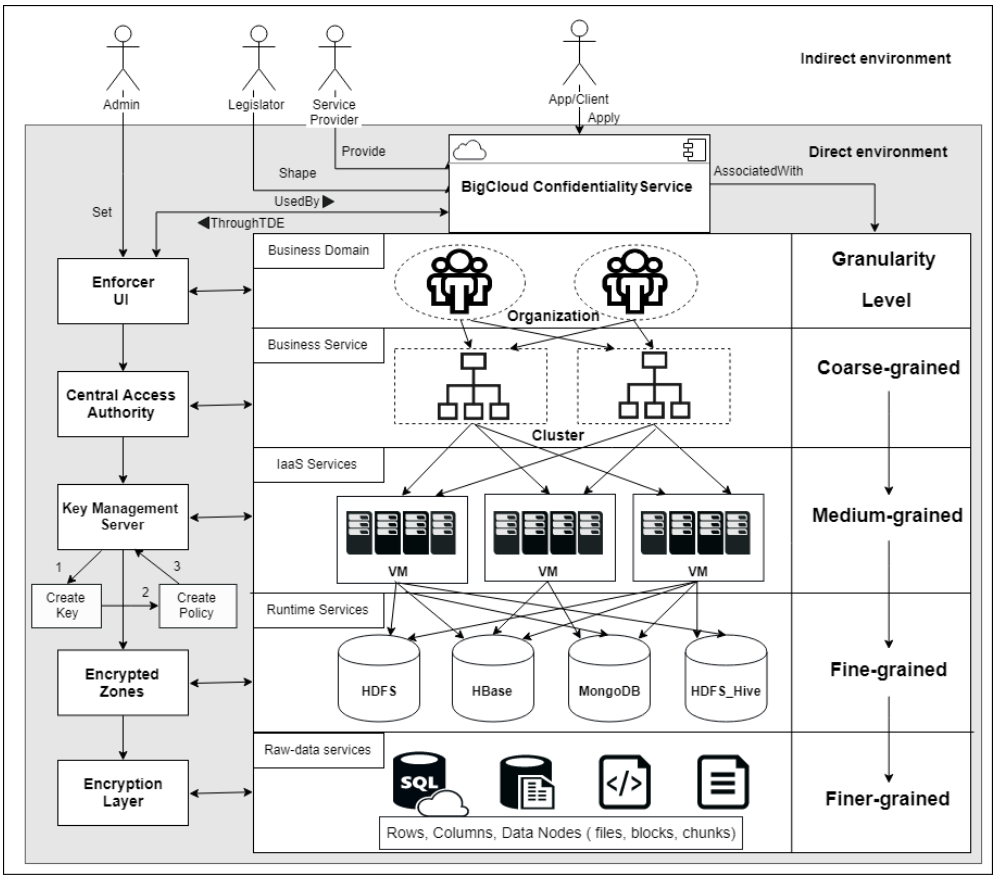
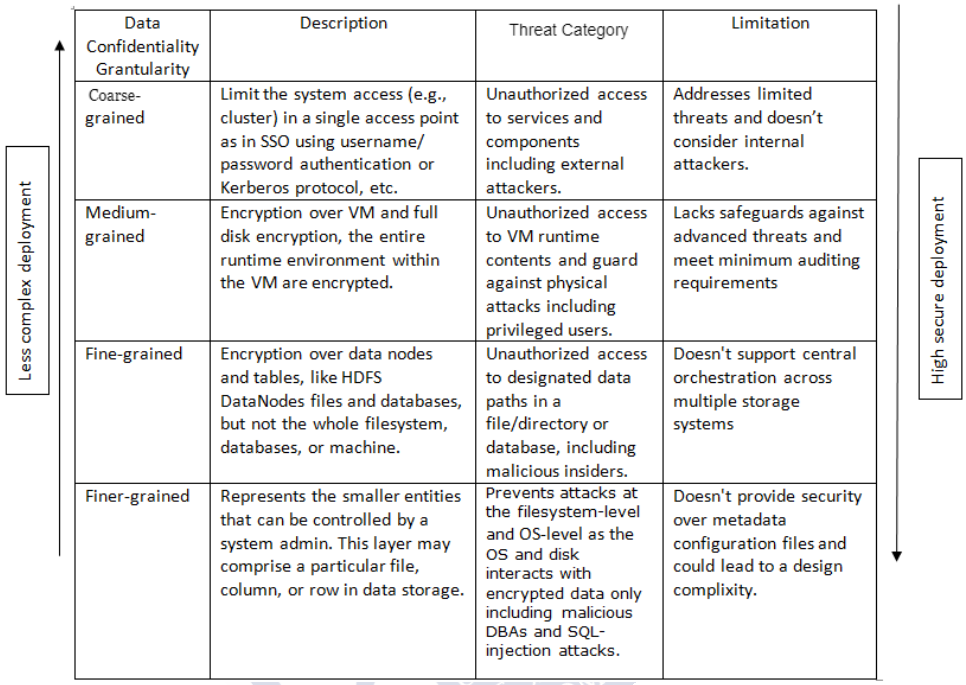


Figure 2.8: Transparent Data Encryption Analysis Pattern with Encryption Zones Granularity in a BigCloud System Architecture.

within a BigCloud ecosystem. Ensuring adequate services and data confidentiality while the client is signed into a privileged BigCloud session requires several layers of control. These layers for controlling data confidentiality are separated across four essential strategies. Figure 2.9 discusses these strategies by grouping similar techniques and mechanisms into the same layer of granularity. The figure further specifies both the solution domain (threat category) and the solution limitation over a BigCloud security stack.

On the other hand, TDE encrypts HDFS data at rest (on disk) using an interaction of multiple components and security keys. Next, we illustrate these components and stages by



Data Confidentiality Grantularity	Description	Threat Category	Limitation
Coarse-grained	Limit the system access (e.g., cluster) in a single access point as in SSO using username/ password authentication or Kerberos protocol, etc.	Unauthorized access to services and components including external attackers.	Addresses limited threats and doesn't consider internal attackers.
Medium-grained	Encryption over VM and full disk encryption, the entire runtime environment within the VM are encrypted.	Unauthorized access to VM runtime contents and guard against physical attacks including privileged users.	Lacks safeguards against advanced threats and meet minimum auditing requirements
Fine-grained	Encryption over data nodes and tables, like HDFS DataNodes files and databases, but not the whole filesystem, databases, or machine.	Unauthorized access to designated data paths in a file/directory or database, including malicious insiders.	Doesn't support central orchestration across multiple storage systems
Finer-grained	Represents the smaller entities that can be controlled by a system admin. This layer may comprise a particular file, column, or row in data storage.	Prevents attacks at the filesystem-level and OS-level as the OS and disk interacts with encrypted data only including malicious DBAs and SQL-injection attacks.	Doesn't provide security over metadata configuration files and could lead to a design complixity.

Figure 2.9: Transparent Data Encryption Analysis Pattern with Encryption Zones Granularity in the BigCloud System Architecture.

defining these components and encryption steps:

- Enforcer UI is a panelboard that is a subset of the general management panel to provide connectivity among the security services and customers. It also acts as a registration authority for all of the external logins and all REST APIs. As a first confidentiality stage for verifying client's calls, it employs either simple username/passwords or a third-party protocol such as Kerberos authentication. Additionally, system admins manage the process (create, edit, and delete policies) for the clients, groups, and applications that can use the service through the UI tool.

- Central Access Authority (CAA) is a policy-based authority that keeps issuing the encryption service that controls the client access for BigCloud customers. However, it represents the primary stage of the data confidentiality process that releases access

policies and functionalities by matching each user/group with its granted permission key and enforcing encryption discussions.

- Key Management Service (or Server) (KMS) is a validation entity that approves client/application reading and writing permissions to the service (encrypted zone). Upon passing the KMS, it affords the master key to encrypt or decrypt the data. In this context, all of the policy creation, encryption, and decryption processes of data encryption key zones are managed in the KMS layer. Apache Ranger may be utilized as a third-party KMS.

- Encrypted Zone (EZ) is a unique file path (a directory or a database) the contents of which are transparently encrypted. When establishing a new encryption zone, a single encryption key is associated with each of these zones. Moreover, the content files within these zones hold a private data encryption key. These keys are never handled directly by the CAA, as they only see a stream of encrypted chunks. This process can be managed as follow:

    1. Create Key: After creating the targeted EZ, the admin creates a key for each particular zone (EZK).
    2. Create Policy: The admin launch policy against each EZK, which spawns service inclusion (who can read/write to the EZ) and add clients/applications to that policy.

- Encryption Layer: The client/application informs the CAA it wants to write a file (e.g., SQL client accessing Hive) to a particular EZ. The CAA requests the KMS to return an encrypted data encryption key from the key store by establishing a trusted connection between the server and the key management server. The client may use that key to write/encrypt to the EZ and read/decrypt from the file. The CAA stores the encrypted data encryption key in the metadata store.

It worth mentioning that the CAA does not control the data encryption keys (encrypted data in the files) directly, but it uses an encrypted data encryption key that can only be decrypted by the clients' data encryption system.

Confidentiality must be maintained throughout the complete data lifecycle. However, herein we highlight the BigCloud confidentiality challenges in terms of data halt, where confidentiality is delivered typically via data encryption techniques. Figure 2.9 shows a com-

parison of data confidentiality granularity with related approaches to data security within the BigCloud system. Medium-grained encryption (MGE) enforces the decision of which files and directories to encrypt on clients' behalf (i.e., clients' discretion)—thereby protecting the swap space, OS, containers, and temporary files as well. MGE, however, does not replace the fine-grained encryption (FGE) in all scenarios. The VM encryption may be employed in conjunction with the file-based encryption, seeking secure multi-layer encryption implementation.

On the other hand, FGE management operates over individual files, directories, and tables (i.e., accessible HBase/Hive DB table columns, Kafka queues, and HDFS file-level of access). With the ability to encrypt each of its components with a separate encryption key, the FGE provides flexible policy decisions and high-performance encryption. Therefore, FGE provides greater overall protection as it stays encrypted through the rest of the layers. However, this protection is at the cost of increased complexity (i.e., it is more comfortable to encrypt a hard drive than a specific cell for instance). FGE has to be associated with a robust access control mechanism and enabled wire encryption.

HDFS is a Java-based framework that requires a JVM environment and contact with the Linux kernel file system before reaching the stored data in the disks. Hence, the machine kernel security (of OS-level) is concerned with security design when processing sensitive data over a public cloud. Utilizing HDFS TDE could potentially cause a slight performance degeneration as additional process layers are attached. However, doing so is justified and acceptable compared to potential threats. Alternatively, direct access to the multi-tenant database may be restricted to specific admins (e.g., DBA). The client's service calls may be routed through an intermediate business layer, which enforces security checks, including means for protecting personal identity and proprietary information. Moreover, visibility labels may be utilized by tagging cells in a table (e.g., Apache HBase) and controlling access to them. This method restricts the access to specific subsets of labeled data in a fine-grained manner.

In Figure 2.9, the coarse-grained encryption is the easiest to implement and manage, and it is the most flexible security approach. This layer limits the system access (e.g., cluster) in a single access point, such as an SSO using username/password authentication or Kerberos protocol. In contrast, the medium-grained encryption works on the runtime environment (i.e., VM access control and encryption). This reasonable compromise shifts the complexity of a solution to the required level of isolation, especially when implemented with other confidentiality layers. The next two layers (fine-grained) are the most secure yet complex approaches

to acquire; they require very detailed policy definitions, including the DB , data node, and even control decisions on the file paths and the specific rows, columns, and cells of the target storage.

Auditing and Analysis

To cope with the modern security demands of large-scale distributed clusters, as in a BigCloud, any security architecture should be able to perform security auditing and analysis at the service level. Security auditing and analysis aggregates log files and reports and provides a robust audit capability within different components of the BD ecosystem. This layer may also afford granular insights into pieces of information by performing security and risk assessments, tracking data pipeline audit logs, and examining behavioral analytics to meet their compliance demands within BigCloud. Examples of this may include incident reporting, behavioral and data activities analytics, daemon (processes starting under the framework and running in the background) logs, and risk assessment of the system components regularly. This feature does not only identifying security issues but provides a sophisticated alert engine that identifies security vulnerabilities and shows insights. Apache Eagle is an open source analytics solution for the Hadoop frameworks and applications [14].

Orchestration and Automation

With the increasing number of different security frameworks, policies, and products in a cloud stack, the connection and integration of these tools is a cornerstone behind inclusive security. This process is called security orchestration, and it brings together these various technologies to work in harmony for the benefit of its customers. It is crucial to standardize and model security to enable interoperability among the various security subcomponents and products. This effort aids in supporting the heterogeneity of security deployment over IaaS using various security layers and tools. By bringing together security components consistency it improves efficiency and effectiveness of security management and the processes surrounding them. Orchestration, hence, is critical for managing the heterogeneity of security deployment over IaaS using various security layers and tools. By bringing together security components consistency, the orchestration process improves the efficiency and effectiveness of security management and the processes surrounding them. Moreover, the synchronization of the security ecosystem helps security admins and clients to make more informed decisions and aids in better specification development. Security orchestration involves advanced automation procedures by

assembling security alerts across the ecosystem. By enabling universal alert repository, security implementers may execute automatable mitigation policies and standardized reactions scenarios. This feature strengthens the overall security operations and supports the right incident response. Apache Knox provides a common platform for frameworks interaction by abstracting the policy exchange. Likewise, Apache Bigtop [13] equips the stack with comprehensive packaging, testing, and configuration of big data frameworks. Bigtop supports a wide range of components/tools for continuous integration using a Jenkins server.

2.7 Summary and open challenges

In this chapter of the thesis, we propose a systematic research methodology for the security of BigCloud adoption. By capturing the methodology stages, we design four primary models that guide the security deployment of any BigCloud solution. First, we design a reference architecture to summarize the relationship between the security service and other cloud services as well as their functions. Second, we offer a solution to the main research question, which deals with the security elements associated with a BD deployment over the IaaS cloud model. Our solution is the design of a security component model that consists of main actors that emphasize the separation of concerns with respect to the service functionality (data service security, IaaS security, etc.) and non-functional security requirements at the beginning of the design, through a reuse-based approach and specifications. Third, we propose a security analysis pattern that refines the cloud context-pattern [24] in synergy to an extended CIA triad [68]. It provides a set of guidelines for the structuring of BD specifications, which relates a cloud design to its security environment. Finally, we suggest a structured election method for BigCloud-specific security selection. It delivers various insights regarding the latest ongoing developments and cutting-edge frameworks by mapping each security domain to its solution knowledge.

2.7.1 Recommendations

Although this chapter has offered an acceptable start towards BigCloud security evaluation and has enhanced the overall understanding of the BigCloud security process. Aiming to analyze further and discuss our results, we draft some recommendations and suggestions that converge on the idea of BigCloud security by design for both service providers and clients from the knowledge obtained in this study.

BigCloud Provider

Several strategic recommendations maybe suggested and given to the service providers to improve and fill the unimproved gaps. Among these recommendations is BigCloud Security Evaluation Framework. The main goal of this evaluation framework is to help BigCloud providers to identify the unimproved gaps according to particular security control elements. An example of such effort is represented in section 4.

Deploying BD frameworks in a cloud environment, whether private or public, demands proactive thinking regarding security ramifications. Security is magnified when considering the impact on clients' sensitive data. This is especially true when IaaS cloud providers control the underlying infrastructure (storage, servers, and networks), while clients have no control over these assets. This shift of responsibility requires providing capabilities to assure the functional properties of BigCloud security and the trust concerns between the BD owner and the IaaS cloud providers. These concerns are based on a lack of control, visibility, and governance while outsourcing the client's data computation. Ensuring the security of BD frameworks over cloud deployment architectures is a keystone to sustain the porting of BD applications to cloud deployment architectures.

A substantial effort has been made to solve the problem of cloud quality-of-service evaluation [6]. Data security and reliability are first-class considerations that play an essential role in most cloud-computing contexts. However, there is a remarkable research gap regarding the evaluation of the BD security service within the cloud [48]. A security evaluation framework can maximize the level of trust (between resource provider and user) and minimize the risk to an acceptable level. Hence, BD application implementers not only have a clear sense of whether the provided service security level is high or low, but can also assist in improving the trust level among them. In this regard, a security evaluation framework that identifies unimproved gaps (according to security control elements) serves to converge BD operations to vast cloud environments. Clients will consider an IaaS cloud provider trustworthy if they fulfill the security requirements of a rigorous security evaluation framework.

It is still challenging to put forward such an improvement plan unless it is based on the results of a security analysis that establishes clear security components and requirements. The BCSC model accommodates the previous security evaluation framework requirements. Cloud adopters, who are involved in the development of BD solutions, may leverage the BCSC to perform a security analysis that maps the installed/needed security components. Further, the BCSC guides the security designers in selecting the required security controls that best suit

their demands. Overall, an evaluation framework with interest in BigCloud security would help in the following:

- Meeting client satisfaction: BigCloud service providers can provide adequate information regarding their system security, which indeed raises client satisfaction.
- Improving BigCloud security services: Security evaluation can play a vital role in meeting client demands by providing the IaaS cloud with security improvement initiatives and gap analysis.
- Managing security risk: Security evaluation results can guide providers in detecting unimproved security gaps between their current IaaS cloud state and the ideal security state.
- Guide porting new BD execution environments: A secure BD execution environment would serve to converge BD operations with the vast cloud paradigms (e.g., edge cloud, decentralized cloud, etc.).
- Gaining competitive advantages: IaaS cloud providers could use the results of the security evaluation framework to remain competitive in the market.

BigCloud Client

Several recommendations can be concluded for BigCloud clients. For instance, clients are recommended to use password managers, which offer greater security and convenience for the use of passwords to access BigCloud services. Typically, clients' passwords are managed using an encryption mechanism, which is achieved principally through storing such passwords in an encrypted database. The clients have to ensure that the cloud providers protects their identity correctly and that they enforce strong and unique passwords. Other features used for automating the filling of the password and sharing credentials, which are integrated to the function of the browser over services (e.g., XML or REST) and allow users to change and randomize passwords, are required to be testified by the client for better cloud service selection. For instance, Azure Active Directory (AD) uses REST instead of the traditional LDAP, which is meant for running applications over SaaS and providing identity management services. Meanwhile, AWS provides fine-grained access control to AWS resources.

CHAPTER 3

COMMON SECURITY CRITERIA FOR VEHICULAR CLOUDS

Internet of Things (IoT) is becoming increasingly crucial to intelligent transportation system stakeholders, including cloud-based vehicular cloud (VC) and Internet of Vehicles (IoV) paradigms. This new trend involves communication and data exchange between several objects within different layers of control. Security in such a deployment is pivotal to realize the general IoT-based smart city. However, the evaluation of the degree of security regarding these paradigms remains a challenge. This chapter provides an insight into addressing this concern.

3.1 Introduction

Cloud Computing offers virtualized technology and networking resources over the Internet to store, process, and analyze petabytes of organizations and individual data dynamically [54]. Over the last years, advancements in this trending technology have shaped the modern application delivery model and the functions they perform [113, 116, 27]. IoT continues to be one of the fastest growing computing models. According to [58], the estimated IoT security market of this model will reach 4.4 Billion USD by 2022.

With its great potential to provide cheap and straightforward access to large amounts of scalable computing power, the advantages of adopting cloud computing to IoT applications are indisputable [1]. The vehicular cloud offers to combine the best of both vehicular ad

hoc networks (VANETs) and cloud computing. Therefore, by utilizing cloud capabilities to extend the inter-networked vehicles, clients can enjoy efficient services (e.g., entertainment) as well as receive real-time alerts about road incidents or traffic status. However, when a VC system is realized, important security factors should be considered. These factors include the architectural design of the system and the utilization of underlying security technologies and services.

A main issue is related to client trust in data transfer in and out of VANETs. This connectivity concern mainly involves communication among vehicles (V2V), vehicle to infrastructure (V2I), and vehicle to cloud (V2C) connection, besides storing and processing critical data within an off-premise (public) datacenter. Other security threats associated with this adoption include privacy, integrity, confidentiality, and the availability of data in its whole life cycle. Fulfilling the highest security and data protection requirements, among all architecture delivery layers, is a common objective. Nevertheless, various effects can make it challenging to ensure the security properties of intelligent transportation systems [60]. For this reason, security and privacy of data management are considered among the leading next-decade research directions in the VC field [28, 38].

This work contributes to the VC security deployment body of knowledge by facilitating the realization of secure Internet of Vehicles systems in the VC model by identifying the common security criteria (CSC). First, we start examining in detail the building blocks of the VC security stack for supporting IoV science. Then, we classify the different layers of security based on their supported service models into a conceptual model. Second, we provide many insights into VC privacy by introducing a novel security analysis pattern. This pattern maps the current VC architecture layers into privacy granularity specifications of the solution domain. Finally, we propose a security evaluation conceptual model that consists of six criteria that influences VC security formation. The model methodology consists of security control elements (SCEs) that generally influence both the client and provider security formation within the VC environment. Next, we extend these SCEs by identifying the most relevant security control components (SCCs) and security control subcomponents (SCSs) that aimed to develop an effective VC trust solution.

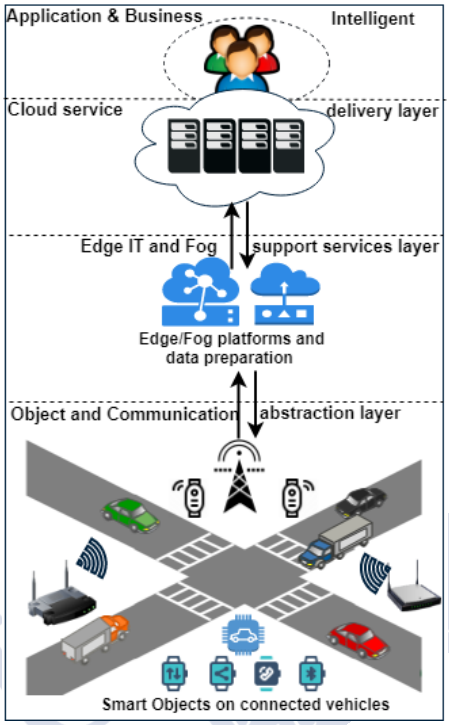


Figure 3.1: A conceptual layered abstraction architecture for connected vehicles model.

3.2 Vehicular Clouds Security

Cloud computing is a set of enabling technologies, broader services, and more flexible solutions for an enterprise to deploy its frameworks. This section discusses the relationship between IoV and cloud-computing services. Further, we specify a vocabulary to design elements associated with VC actors by presenting a VC conceptual model, which outlines the main components of the VC environment. It abstracts the relationship between the security service and other VC services as well as their functions. It, therefore, relates to companion security requirements and features that are the basis for conducting reliable VC implementation.

Overall, Figure 3.1 illustrates the major actors within the VC ecosystem to refer as: objects and communication layer, Edge/Fog IT support service layer, Cloud service delivery layer, and finally, applications and business intelligent layer. In such architecture, the data life cycle

is composed of several stages, with the data flowing from sensors and smart objects to the cloud. It begins with data generation (e.g., collected from vehicles and IoV infrastructure) by different objects, where connection security is required. The allocated data are then assembled (using compression and aggregation) and transferred to Edge IT and the cloud. Following this, metadata are reserved, and the data are materialized in the cloud storage within a data lake, which can be a multi-tenant storage serving several applications.

In the same context, data life cycle flows among different VC service layers. First, from the data generation point (very bottom) is the sensor mesh network, named *object and communication abstraction*, which comprises physical smart devices such as sensors. This layer outlines hardware and firmware that provide car–to–car and car–to–infrastructure (defined earlier as V2V and V2I) connectivity using different communication technologies like Bluetooth or Wi-Fi. Subsequently, with the vast number of IoV devices spawning, the storage and computation of this data will take place on the cloud. Hence, different applications can harness it to make valuable decisions. It is worth mentioning that some applications utilize stream data processing on the edge nodes (also including Fog deployments) such as traffic ahead or parking nearby. This will be further discussed in Section 3. Finally, VC systems grant end users with applications in a SaaS layer (e.g., machine learning and ITS). These applications, however, leverage the services and functionalities of the lower cloud services layer. Based on an analysis of these applications, users and administrators can remotely send commands to smart devices at the bottom layer. Supporting architecture security involves several approaches, spanning from physical and network security to edge and cloud platform security. It also includes the framework and application security, security analytics (e.g., behavioral analysis, data flow analysis, Trojan detection, etc.) and continues security testing, besides identification and access management, such as authentication and authorization.

3.2.1 Cloud Data Security

Cloud Data operations include data protection and monitoring in the three stages of the data-security life cycle: data at rest, data in motion, and data in use [32, 72, 53]. Table 2.2 maps each stage with its primary security mechanism. These stages are further discussed as follows:

Data storage (data-at-rest) security: The primary capabilities of data storage include managing the storage required by VC applications. However, modern cloud storage components can implement backups (including virtual storage). These backups may be considered a remediation technique that encourages working with the backup to create snapshots at reg-

ular intervals. Another way that the storage component can collaborate with the hypervisor or container is to allow for workload migration and metadata storage among the datacenter nodes. Data storage security services must cover file systems (FS), database (DB), and object storage (OSt) sytems security scanning (data content discovery) to identify and locate sensitive content (e.g., credit card numbers). This method supports data compliance and audit efforts by conducting comprehensive reporting on the effectiveness of data storage protection mechanisms and guides decisions on security measurements for implementing data encryption (disk-level encryption) and masking, removing, or warning the file owner. In general, data at rest is considered more vulnerable than data in transit [53].

Data transfer (data-in-motion) security: It can be classified, based on the connection domain zone, into cloud and edge data movement. The first one involves internal data transfer between storage capacities and processing units on the cloud. This communication usually takes place at the platform layer. Businesses must accept the reality of data insecurity in-motion and take proactive steps to remediate the security risk that's inherent with sending data off-premise. All data transfers (including the metadata) should be handled with the same level of minimal trust. Measurements to prevent an expensive and embarrassing data breach include:

- Private WAN service or encapsulating the data workloads;
- Implement robust network security controls and sniffing the traffic on the network using proxies (to identify the content);
- IPsec VPN gateway and transport layer security protocols.

Second, external data transfer occurs between the vehicles and the cloud provider, causing the network to act as an intermediary that provides data transport services using different communication methods, from dedicated network channels to ad-hoc networks and the open Internet. The cloud may require the network provider to provide secure connections between it and the IoVs to keep vulnerabilities (e.g., a man-in-the-middle attack) in Internet transmission channels to a minimum. The network service provider should maintain security control points, conduct security testing, and prevent suspected tasks.

Data processing (data-in-use) security: This refers to securing the processing environment. High-reliability data execution may be achieved by (i) harnessing robust distributed identification and access management and (ii) enforcing isolation among different IoV zones,

edges, and applications. Appropriate access control is required for protecting IoV configuration and its applications [51]. Doing so will prevent the attacker from modifying these critical assets [11]. Data blocks are duplicated across multiple nodes to support high reliability and availability of VC operations. This traditional approach is considered costly and achieved moderate performance in large scales of operation [118], which makes the advancement of modern large-scale distributed storage systems with erasure coding techniques necessary. This storage technique provides the same level of fault tolerance with much less storage space and has been already implemented successfully in the Hadoop Distributed File System.

3.3 Vehicular Clouds Security Analysis Pattern

Hosting data in IoT-based systems increases the number of potential security risks. In considering security, as mentioned earlier, we divided privacy within such a deployment into several granularity levels. The following sections provide insights in this regard using a security analysis pattern. These analysis patterns capture an abstraction of the targeted security solution, which is represented as a group of related, generic objects with stereotypical attributes and expected interactions [91, 24].

Figure 3.2 shows a model designed to guide policies for the information security life cycle within VC systems. The model is composed of two environments: indirect and direct. On the one hand, the indirect system environment is related to the legislator, service provider, external customers, and other stakeholders that interact indirectly with the VC; that is, they are not connected to the IoV by association. Nevertheless, data, patterns, and model sharing apply at this layer. The business domain, as an intermediate layer, acts as a bridge between indirect and direct environments.

The direct environment, meanwhile, has been divided into five interacted layers, based on the VC service layers in Figure 3.1. The object and communication abstraction have been divided into the VANET and the backend communication channel. Meanwhile, application and business intelligence have been divided into business services and business domain. Finally, the support service layer is represented by edge IT services including the same functionality.

These service layers have been designed by refining the layered architecture for the IoT in [97, 31] to meet the IoV specifications. Our model also refines the virtual object layer that bridges the gap between the physical and the virtual world [88] by extending its security functionalities. This layer separation aims at analyzing the security objects and attributes of

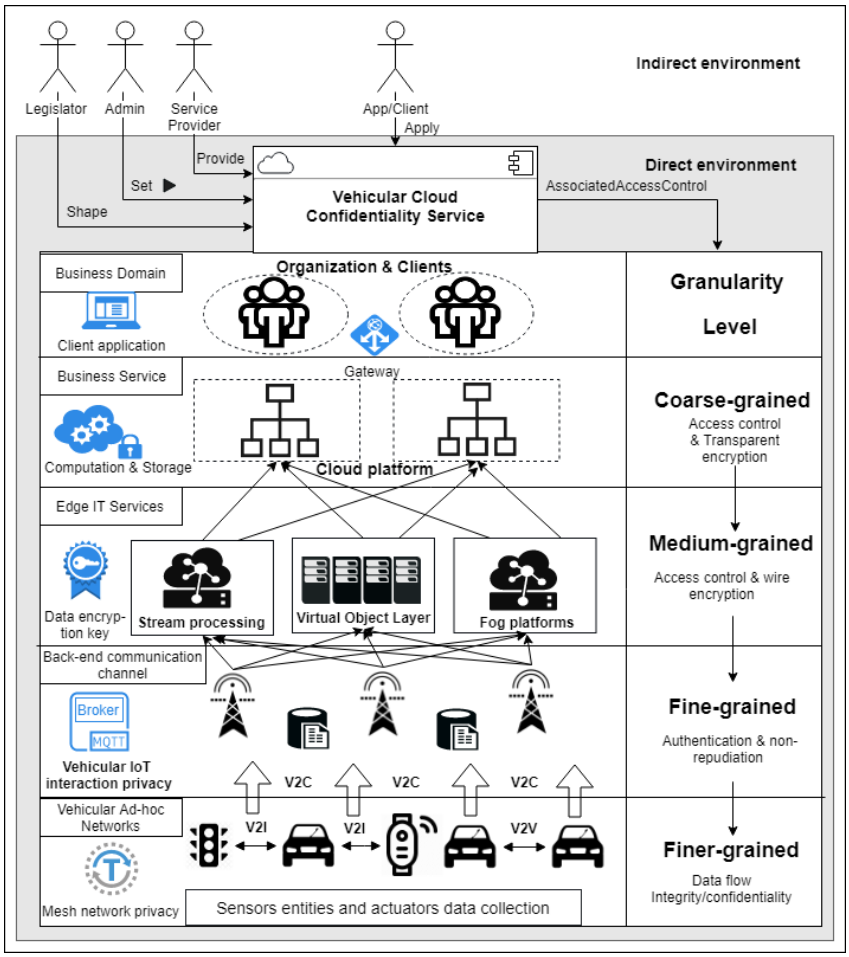


Figure 3.2: Confidentiality control as a security analysis pattern in a vehicular cloud system architecture.

the VC ecosystem with a higher granularity.

The data flow from the mesh network objects up to the client's applications in the cloud. The bottom layer comprises physical smart devices, which enables efficient information exchange by expressing communication as V2V, V2I, and V2C. Overall, data integrity and confidentiality are the primary concern. In the upper layer, the back-end composite communication channel aggregates data to meet the IoV application requirements. These communica-

tions can be implemented as web resources using RESTful Web Services or other protocols such as MQTT. It thus provides universal methodologies that link all physical objects to the edge IT. In this case, authentication policies (attribute based, time based, zone based, etc.) as well as nonrepudiation mechanisms are substantial. RFID-based identification enables the implementation of a wide range of applications by linking the physical objects and the virtual world [88].

3.3.1 Fog and cloud computing

Fog computing extends the cloud computing paradigm to the edge of the network [26, 56]. This is critical in time and location-sensitive IoV applications, as modern ITS demands real-time processing in many scenarios where VC resources are inter-networked via purely peer-to-peer connections [46]. Therefore, the association of edge/fog computing and IoV stream processing is imperative [26]. This trend aims at addressing the associated high latency, low bandwidth, and communication delays pertinent to using the central cloud. Access control, wire encryption (end-to-end), data governance, and co-existence of heterogeneous network architecture over a shared infrastructure are the main barriers to improving trust.

With the broad participation of IoV and other ITS devices, the allocation and execution of data will mainly be done in the cloud. In the final stage, different applications can harness business intelligence to make valuable decisions. These applications leverage the services and functionalities of the lower cloud services layer. Data integrity is considered under these circumstances, and software integrity should be considered, too. In general, software integrity is divided into protecting virtual objects in the second and third layers (communication between entities and edge/cloud resources by mean of RESTful APIs). Clients and administrators can remotely forward commands and instructions to smart devices at the bottom layer using these applications. They also can prepare reports, conduct in-depth analysis, or even utilize visualization applications. The admins set ID access control and transparent data encryption at rest (on disk) using the interaction of multiple components and security keys.

3.3.2 VC Confidentiality Service

The confidentiality layer components, stages, and granularity levels within a VC ecosystem are shown in Figure 3.2. Ensuring adequate services and data confidentiality (over its complete life cycle) while the client is signed into a privileged VC session is the main goal. These

layers are distributed across five essential strategies for controlling data confidentiality with different levels of control. Figure 2.8 illustrate these components and stages by defining the components and encryption steps. However, the encryption engine is the foundation layer in any security solution. Utilizing a hardware security module provides the highest level of physical security. In this model, it is recommended to implement strong authentication and cryptoprocessing at the IoV communication stage (backend communication or data aggregation stations). This mechanism provides the most secure option using a fine-grained approach and acceleration of cryptographic processes to improve performance. This approach is well recognized in large-scale systems such as ATM security [70] and vehicular hardware security models [117]. It is also worth mentioning that wire encryption protocols may be utilized (SSL/TLS based authentication).

Another technology for maintaining the confidentiality is the Blockchain. Blockchain as an enabling technology of blocks cryptography, which has been established as a solution for the large-scale distributed agents of the Bitcoin system security [64]. In its core functionality, Blockchain technology relies on providing robust cryptographic proof for data authentication and integrity — by providing a list of all transactions and a hash to the previous block. This list is verified by a majority agreement of nodes which are actively involved in verifying and validating transactions [39].

3.4 Security Control Elements

The majority of current VC systems are deficient in conducting trust analysis that aids in improving the understandability, decomposability, and reliability of their systems. Therefore, it is suggested to provide highly trustworthy evidence of the proposed evaluation criteria, with a robust methodology that assists the exploration of the subjective nature of security evaluation criteria within the targeted VC solution.

This research addresses the previous issue by capturing security services used in VC environments to describe the security items and the relationships amongst them. It further examines security oriented to VC trustworthy design, aiming to represent its glossary and landscape techniques and to define research gaps. Security election, in general, is a control element that shapes the policies, practices, procedures, and responsibilities of the service provider.

The question we are addressing regarding the verity of security solutions in VC is as follows: *What are the Common Security Criteria (CSC) in context-based security that can be*

used to evaluate the degree of security in VC environment?

To answer the question, we collected all criteria of trust that have appeared in the literature-based security. The most critical criteria are: (i) physical and environmental protection; (ii) logical access control; (iii) data and service confidentiality; (iv) data and service integrity; (v) data and service availability; and, finally, (vi) life cycle control (auditing, governance, and compliance).

To facilitate using the systematic criteria selection methodology, we summarize the proposed models using tables, which consist of SCEs that generally influence security. These SCEs are designed to cope with the security services mentioned earlier in Section 3.2 (Figure 3.2). Namely, physical, network, edge IT, cloud, and application security, respectively. We have further extended these SCEs from different research contexts into VC context by identifying SCCs and security control subcomponents (SCS) with the aim of developing an effective VC security solution.

Physical and environmental protection is a subcategory of security. This control area involves the core facilities and equipment of VC (sensors, storages, and networks). As the base for all the above layers, it is critical to ensure an adequate level of security design and implementation. The analysis results, as shown in Table 3.1, showed that the physical and environmental protection areas can be divided into hardware and software SCC.

Table 3.1: Identifying physical and environmental protection and its TCC of Security.

SCE	Vehicular Cloud physical and environmental protection
SCC Type	Refines both Hardware and Software Security
SCS and their importance	- Smart objects, Sensors, and Actuators Security - Communication channels security - Network Security design (segmentation and isolation). - Firmware Security. To fortify infrastructure and resources (internal equipment and devices like network devices) with roles and privileges and prevent any unauthorized access. Additionally, to detect any abnormal behavior while controlling Firmware accessibility by entities to ensure business continuity.

Logical access control is a set of rules and procedures that control access to VCs content and/or data. It can be sub-divided into authentication, authorization, and service gateway. The analysis results, as shown in Table 3.2, point out that many authentication mechanisms may be applied to increase the level of access security.

Table 3.2: Identifying Logical Access Control and its SCC of Security.

SCE	Vehicular Cloud logical Access Control
SCC Type	Refines ID and access management, and Gateway
SCS and their importance	Authentication and Authorization: - LDAP-based and Kerberos protocol - Third party approach. - Dual-stage authentication (at mesh network and Edge layers). Make sure that access to the VC is controlled with only the commands necessary to perform authorized functions and increase the level of access security. Furthermore, to manage the IoV roles and privileges and prevent any unauthorized functions.

Encryption as a security tool, fortifying data confidentiality by ensuring that an authorized entity can only access a piece of the given information (data flow). Building trust in VC is required by service providers to ensure some specific encryption control criteria. The analysis results, as shown in Table 3.3, indicate that several encryption methods can be applied to maintain the confidentiality.

Table 3.3: Identifying Confidentiality and its TCC of Security.

SCE	Vehicular Cloud Confidentiality
SCC Type	Refines Data Encryption control at rest.
SCS and their importance	- Transparent Data Encryption - Blockchain of Things - End-to-end wire encryption - Identity-based encryption and attribute-based encryption. -Data masking and tokenization. To ensure that data are encrypted at the whole life-cycle. Key management, key distribution, and encryption engine are associated interests when implementing VC confidentiality.

The creation and location of virtual objects are primarily proposed in the VC, and their communication uses RESTful technologies. Building trust in VC requires specific integrity control areas and control criteria, especially when using Wi-Fi or public networks to reach a cloud service, by applying validation or error checking. The analysis results, as shown in Table 3.4, indicate that integrity can be divided into data at transition and software/hardware integrity using different approaches to tackle this issue. Protecting edge IT by developing

Table 3.4: Identifying Integrity and its TCC of Security.

SCE	Vehicular Cloud Integrity
SCC Type	Refines Data at transition and Software/Hardware Integrity control.
SCS and their importance	- Data transmission protection, e.g., SSL/TLS protocol. - Data isolation. (encryption) - Communication Integrity (objects and edge/cloud computation) - Scalability and compatibility. Integrity ensures that data and related HW/SW can only be accessed and modified by those who authorized smart objects and applications. Integrity includes controlling the physical environment, that is, network and servers. Data integrity can be indicated by any malicious action, while data is transmitted (e.g., a man-in-the-middle attack) between the smart objects to the cloud or between the IoV components.

a method for granting its resource privileges to enforce information flow between V2I and V2C requires trusted computing techniques such as the virtual trusted platform module as a supplementary security measurement.

Table 3.5: Identified Availability and its TCC of Security.

SCE	Vehicular Cloud Availability
SCC Type	Refines Data and service availability control.
SCS and their importance	- Data replication - Failure history and recovery - Redundancy of disasters and incidents May also be referred to as resilience and fault tolerance, which aims to guarantee that data/service remains available at a required level of performance in the event of the failure of some VC components.

Availability is a sub-category of trust. It is used to argue that VC can be made available to end users (clients and vehicles) from anywhere at any time by using any gadget. Furthermore, VC should ensure the availability of the operating time of the system as specified in the SLA. The analysis results are shown in Table 3.5.

Finally, compliance is a sub-category of accountability. It refers to verifying the commitment of VC providers to follow specific legal requirements, standards, contracts, and policies. Additionally, auditing is a sub-category of analysis. Any secure model should not only secure

Table 3.6: Identifying Life-cycle management and its TCC of Security.

SCE	Vehicular Cloud Life-cycle management
SCC Type	Refines Auditing, Governance, and compliance controls
SCS and their importance	- Behavioral analytics and monitoring - Managing service/operation log information and file - Tagging (data labeling for governance) and filtering - Automation of service. - Compliance with legalisation (data and software). - Compliance with SLA. To make sure that all VC stages meets the auditing, governance, and compliance control requirements.

its entities but should be able to perform security auditing at the whole service layers (from sensors to the cloud). Correspondingly, governance is a capability that ensures efficient manageability and data/service quality. This feature leverages information to help implementers gain insight and build confidence in business decisions and operations. The analysis results are shown in Table 3.6.

3.5 Summary

The main objective of the study was to review and further discuss the various criteria that influence degree of security in the VC context. This chapter first outlines a conceptual model that provides an overview of the criteria that influences both the client and provider security formation. Next, we create a security analysis pattern that envisions VC to have both back-end data aggregation channels (e.g., MQTT broker) and edge/fog data services to supplement the sensor mesh network (e.g., VANET) and cloud architecture layers. It also represents VC privacy as a multi-criteria construct affecting the data life cycle.

This research realized that the majority of current VC systems fail to provide the security-level for their system. We also advocate that, when evaluation and selection is realized, VCs can lead to a significant enhancement in the system security that sustains its adoption by the clients. Thus, we explore the various criteria that influence the security degree in the IoV-cloud context. In the next chapter, we intend to extend the implementation of this model to an evaluation framework of IoT-based VC environments using a fuzzy modified VIKOR approach [89] with a systematic evaluation and selection method.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 4

EVALUATION FRAMEWORK FOR VEHICULAR CLOUDS

4.1 Introduction

The vast amount of data generated by the Internet of Things (IoT), especially VANETs (vehicular ad hoc networks), needs a scalable resource, which can be provided by cloud computing on a rental basis. Accordingly, the cloud has attracted the most significant interest in IoT-based applications [27, 69], particularly vehicular clouds (VCs) [94]. The transmitted data in such a realm should be located securely throughout the whole life cycle to guarantee high data privacy. Security is a crucial aspect of spreading the adoption of cloud capabilities among industrial connected vehicles (CVs) [103] and industrial cyber-physical systems [109, 40]. In this regard, security by design can mitigate many of these imposed challenges [84]. Without adequately addressing this concern, a VC would not gain the clients' trustworthiness and, hence, acceptance. This facility can be achieved by regularly evaluating the security components of the VC to put together an adequate plan of improvement. However, a dedicated work that evaluates the trustworthiness of this framework remains an open challenge.

When a cloud-based connected vehicle (CV) system is realized, significant security concerns should be considered [85, 44, 115]. Security features (criteria) are not equal, which means that they should not be governed and managed at the same level. It is essential to note the importance of creating a shared understanding of security-related criteria and be able to assign priorities based on each security criteria impact and potential for mitigation. These

considerations include security by design of the system and utilization of underlying security technologies and services. This research captures security services used in industrial CVs that describe the VC security items and relationships among them. It also presents landscape techniques to define security gaps (distance from an ideal point of security) and best practices. This work aims at facilitating the realization of vehicles securely connected to cloud computing in an industrial environment. First, we analyze the security criteria of data analytics in VC computing and propose three-level security evaluation elements. Namely, Level 1 consists of 6 common security criteria (CSCs). Level 2 consist of 10 security control components (SCCs). Level 3 consist of 36 security control subcomponents (SCSs). Next, the framework uses the proposed security criteria as a measure to comprehensively evaluate and rank the security criteria using a multicriteria decision-making algorithm. Finally, the framework proposes to compare the evaluated criteria to an ideal level and report (visualize) the evaluation results in order to update the security by design.

Deploying a secure industrial IoT solution has only been recently proposed to provide security analysis and mitigate vulnerabilities at the design/modeling phase [84]. However, this proposal did not offer a practical solution for optimizing the design phase by evaluating the system security features as this study does. Additionally, this study copes with the previous limitation and aids in better security updates and patches at the runtime/simulation phase. The proposed methodological approach combines the use of criteria importance and performance rates for determining trust service attributes that a designer or policy maker should devote more attention to. It also labels which feature should be lower priority to keep the focus on the high-priority ones. Trustworthiness evaluation of vehicular cloud (TrustE-VC) [8] offers a useful and practice-ready tool for designers and industrial CV practices to better evaluate and select industrial CV trust requirements.

The remainder of this chapter is organized as follows. Section 4.2 provides the background of this study along with the underlying motivation. Section 4.3 presents the building blocks of the vehicular cloud security stack, the evaluation criteria extraction method, and the related work. A discussion of the proposed framework (TrustE-VC) and its main components is presented in Section 4.4. Section 4.5 highlights the TrustE-VC framework outputs and results that need to be addressed within the next-generation industrial CV and IoV-cloud platforms. A comprehensive discussion and the future direction are discussed in 4.6, and finally, we conclude this chapter with Section 4.7.

4.2 Background

The focus on significant requests for sensitive data allocated from various sources drives us to pay more attention to data security [5]. These confidential data can reside near the device (on the edge/fog side) or in the cloud. This implies that the data can be attacked during transmission or at their site. Data security and privacy are considered to be main barriers for full acceptance of the IoT paradigm [107]. Most security threats and extensive privacy issues stem from the lack of well-investigated security and privacy guidelines. However, these guidelines and security requirements (confidentiality, integrity, availability, privacy, audibility, accountability, and trustworthiness) will give the IoT stakeholders the vision to build secure IoT systems during the design phase, which aims to enhance IoT data security and privacy and prevent data scams [64].

4.2.1 Vehicle-to-Cloud Connection

VC technology relies on vehicles' onboard computing capabilities, storage, and sensing power and leverages cloud computing services. Cloud computing represents a practical model that supports the scalable deployment of management of large-scale collected data on the edge of the network with a cost-effective and sophisticated approach of storing and processing big datasets. In this context, the security of vehicle-to-cloud data exchange and communication is a first-class concern among industrial CV practitioners. Overall, Figure 3.1 illustrates the major layers associated with the industrial CV ecosystem, such as the connected vehicle layer, edge/fog support service layer, and cloud service delivery layer [9]. In such an architecture, the data life cycle is composed of several stages, with the data flowing from sensors and smart objects to the cloud. It begins with data generation (e.g., collected from vehicles and IoV infrastructure) by different objects, where connection security is required. The allocated data are then assembled (using compression and aggregation) and transferred to edge IT and the cloud. Following this, metadata are reserved, and the data are kept in the cloud storage within a data lake, which can be multitenant storage serving several applications.

In the same context, the data life cycle flows among different VC service layers [101]. First, at the data generation point (very bottom) is the sensor mesh network, named *object and communication abstraction*, which comprises physical smart devices such as sensors [93]. This layer includes hardware and firmware that provide car–to–car and car–to–infrastructure (defined earlier as V2V and V2I) connectivity using different communication technologies

such as Bluetooth and WiFi. Subsequently, with the vast number of IoV devices spawning data, the storage and computation of these data will take place on the cloud. Hence, different applications can harness it to make valuable decisions. It is worth mentioning that some applications utilize stream data processing on the edge nodes (also including fog deployments) such as traffic ahead or parking nearby. Finally, VC systems assist with the client applications in the SaaS model. Both clients and administrators can remotely send commands to smart devices at the bottom layer. It is vital to grant and sustain the architecture security in different layers, including physical and network security, to edge and cloud platform security. This also includes the framework and application security, security analytics (e.g., Trojan detection), besides identification and access management control like authentication and authorization with continuous security testing.

4.2.2 The necessity of VC security evaluation

The Internet of Vehicles in the cloud paradigm represents a responsibility transfer of data hosting, software control, and infrastructure management [43, 124]. IoV implementers are always seeking to secure their operational environments, as security is always a first concern. On the other hand, cloud providers improve their competitiveness in the cloud market by ensuring that appropriate security expectations for their services are met. Security engineering (planning, designing, and assessment) that minimizes the vulnerability surface and meets clients' security satisfaction is expected. However, it is challenging to meet these expectations unless they are based on the results of systematic gap analysis and assessment. This analysis could aid in evaluating the security level of every component individually as well as their integration. Evaluating the security level of a VC system enables the suggestion of an efficient and effective plan that maximizes the security and minimizes the level of risk to an acceptable level. Consequently, VC practitioners identify the unimproved gaps and have a clear sense of the system security level and how to address the challenges.

IoV-to-cloud security evaluation aids in shaping the security policy of the service provider as follows:

Achieve client satisfaction: Based on the security evaluation, the VC provider will provide adequate information regarding service enhancements to achieve the client's highest level of satisfaction. A security evaluation report will reduce the vulnerabilities among the system components by allowing the security designers to control undesirable behavior. Additionally, it will improve the security level and reduce the risk to the clients' satisfaction.

Improve VC services: A security evaluation can lead to a significant enhancement in the system security that sustains its adoption by the clients. It can also guide service providers to improve their service, in addition to determining and meeting the client's requirements. This information will assist VC providers in establishing a new service level that can meet IoV needs.

Gain competitive advantages: The intelligent industrial CV business is rapidly growing. A competitive and growing market share is vital for the industry. By providing a detailed report on the service security status by conducting a security evaluation, customers can be reassured regarding the security measures that maintain their data privacy and confidentiality. This report, in return, leads to competitive advantages. High rates on a security evaluation after classifying the standard security criteria improve the system reliability in the market. Therefore, industrial CV providers could use this information to convince potential customers.

4.3 Material and Methods

This research contributes to the deployment of the VC security body of knowledge by first examining in detail the building blocks of the vehicular cloud security stack architecture. Second, it analyzes and classifies state-of-the-art security frameworks, which are mainly available today as open-source platforms for sophisticated criteria selection. Third, a rigorous and robust evaluation framework based on evaluation theory, which guides VC service providers to identify security gaps, is proposed. Finally, we highlight some open challenges and recommendations for both service providers and customers for a comprehensive discussion toward achieving the vision of providing IoV-cloud secure services.

4.3.1 Security Criteria Extraction

In this study, we conducted a systematic literature review of more than 500 papers, focusing on the security, confidentiality, and privacy of IoT-based and cloud deployments during the last ten years. The collection of articles was deployed in two different stages; initially, the needed scientific papers were retrieved according to a scholar query executed in Publish or Perish software[1]. These search queries analyzed academic citations using a variety of sources of academically impactful conferences and journals (from Google Scholar and Microsoft Academic Search). The results were exported to a CSV file for further analysis on the basis of

[1] https://harzing.com/resources/publish-or-perish

various metrics (number of citations, venue impact factor, etc.) to determine their impact and perform simple statistical analysis. Afterward, Python code was developed to read the papers' URLs from the previous record and download them in local files. Finally, we classified those papers manually based on their main research idea.

This intensive review led to the exploration, identification, and understanding of six common security criteria (CSCs) that generally influence IoT security. Next, the interview study successfully extended these CSCs (level one) from different research contexts into the VC context by identifying ten security control components (SCCs) (level two) and 36 security control subcomponents (SCSs) (level three), aiming to develop an effective IoV-secure solution in the cloud environment. The main evaluation criteria in our VC security framework are presented in Table 4.1. These CSCs are further defined and described in the next section.

4.3.2 Evaluation target and criteria

A deductive content analysis method has been utilized to explore the qualitative data to avoid random security criteria, which manipulates the evaluation process. The deductive content examination ensures that researchers are unlikely to be working from naive perspectives and concepts, which are inspected as the hallmark of interviews. This phase led to the following conclusions:

Physical and Environmental Protection: This describes technologies used to safeguard infrastructure assets against physical attack, i.e., sensors, actuators, network cables, etc., to deter unauthorized physical access to provide security during data movement.

Logical Access Control: This is a subcategory of security that refers to a set of rules and procedures that control access to infrastructure content or data. Authentication mechanisms may be applied to increase the level of access security. In general, it is composed of two main entities: First, authentication ensures that data are from legitimate sources by affording perimeter security capabilities. It also checks that data are not from malicious sources or nodes. Second, authorization ensures that only authorized objects are employed in collecting, streaming, analyzing, and modifying the data of an IoV system. Additionally, it ensures that access to the VC is controlled with only the commands necessary to perform authorized functions to increase the level of access security. Furthermore, managing the IoV roles and privileges and preventing any unauthorized functions are considerations. A supplementary stage that improves the access control auditing includes nonrepudiation techniques. These

Table 4.1: Diagrammatic Vehicular Cloud Security Level in Industrial Connected Vehicles

CSC	SCC	SCS
Physical and Environmental Protection (C1)	Infrastructure Security Design (C13)	-Physical access control rights and roles (C131) -Multiple barriers to physical access (C132) -Monitoring physical access (C133) -Lockable physical casings (C134) -Disaster and incident management (C135) -Mirroring and redundancy of the infrastructure (C136)
	Network Security Design (C11)	-Communication channels security (C111) -Network security design (C112) (segmentation and isolation).
	Smart Object, Sensor, and Actuator Security (C12)	-Malicious insiders (C121) -Firmware security (C122)
Logical Access Control (C2)	Authentication (C22)	-LDAP-based and Kerberos protocol (C221) -Third party approach (C222) -Dual-stage authentication (C223) (mesh network and edge) -Smart object access control rights and roles (C224)
	Authorization (C21)	-Identity access policies (C211) -Penetration testing (C212)
Communication Confidentiality (C3)	Wire Encryption (C31)	-Multiple-level encryption (C311) -Blockchain encryption (C312) -Transparent data encryption (C313) -End-to-end wire encryption(C314) -Identity-based encryption and attribute-based encryption(C315) -Data privacy (C316)
Communication Integrity (C4)	Communication Integrity (C41) (objects and edge/cloud computation)	-Data transmission protection (C411), e.g., SSL/TLS protocol -System vulnerabilities/exploitable bugs (C412) -Data and wire encryption (C413) -Hardware compatibility (C414) -Monitoring production environment (C415) -Scalability and compatibility(C416)
Data and Service Availability (C5)	Data Availability (C52) Service Availability (C51)	-Data replication (C521) -Failure history and recovery (C511) -Redundancy of disasters and incidents (C512)
Data Privacy and Governance (C6)	Data Privacy and Governance (C61)	-Behavioral analytics and monitoring (C611) -Managing service/operation log information and file (C612) -Tagging (data labeling for governance) and filtering (C613) -Automation of service (C614) -Compliance with legalization (C615) (data and software) -Compliance with SLA (C616).

approaches denote that the source cannot deny transferring the packages (i.e., data) it sent earlier.

Communication Confidentiality: This is an associated process that guarantees that a given data stream, i.e., from edge to cloud or V2V communications, can be recognized only by the desired recipients. Encryption is a security tool for fortifying data confidentiality by ensuring that an authorized entity can only access a piece of the given information (data flow). Establishing confidentiality in the VC is required by service providers to ensure some specific encryption control criteria. Several encryption methods can be applied to maintain confidentiality and ensure that data are encrypted during the whole life cycle. Key management, key distribution, and the encryption engine are associated concerns when implementing VC confidentiality.

Communication Integrity: This is the status of ensuring that a malicious intermediate does not modify data migration between its source and the data center (cloud). Regarding data transmission, there is integrity confidence that data can only be obtained and edited by entities that have access to a key (authorized client). The data integrity can be indicated by

any malicious action while data are transmitted (e.g., a man-in-the-middle attack) between smart objects to the cloud or between IoV components. Actions are taken to ensure integrity, including managing the physical infrastructure, i.e., network and servers, restricting access to data blocks and sustaining rigorous authentication processes. Service providers may enforce encrypted end-to-end protocols to link connected vehicles and cloud service. Protecting edge IT by developing a method for granting its resource privileges to enforce information flow between industrial CVs requires secure computing techniques such as a virtual trusted platform module as a supplementary security measurement. This link guarantees that the transmitted data between the edge node side and cloud service side through public networks remain private and integral.

Data and Service Availability: This may also be referred to as resilience and fault tolerance, which aims to guarantee that data remain available at a required level of performance in the event of the failure of system components. IoT-based paradigms mainly depend on the availability of massive datasets when and where the data are needed by the client or an application, which could be a workload on a private cloud or a service running on a public cloud. Availability is commonly realized by providing data replication at the back-end of the service. However, a more recent trend is to utilize edge/fog backups instead, seeking a performance and time beneficial approach.

Data Privacy and Governance: This is an enabling technology that authorizes enterprises to effectively and efficiently meet the compliance requirements within the IoT environment. It also allows integration with the whole enterprise data ecosystem (cloud or data center) and exchanges metadata with other tools and processes within and outside of the VC paradigm.

4.3.3 Related Work

The security of IoV deployment architectures and VC service security have always been a concern and, hence, are a research trend. A large body of research aims to address this concern in the literature with various insights [44, 105, 86, 85, 42, 107]. Recently, Gupta et al. [51] proposed an authorization framework relevant to IoV and vehicular clouds; they discussed the need for access control within such a sensitive environment. They extended their work with the CV-ABACG [50] model, a formalized dynamic group, and attribute-based access control for a smart car ecosystem. In [43], the integration of cloud computing and fog computing for securing the data storage in IIoT deployment architectures was proposed. Meanwhile,

realization of service-oriented models to securely access the underlying resources of cloud manufacturing based on IoT technologies was attempted in [108].

A recent study by Yang Lu and Li Da Xu indicated that the security quality of service-based design has the potential, as a leading research trend, to protect the IoT network [75]. This vital aspect was further investigated to enable security analysis and mitigation of security threats [84]. A risk assessment for wired networks using attack graphs was studied in [99]. Security analysis of IoT systems based on the generic behavior by formalizing the interactions among various IoT things and capturing IoT-specific threat classifications was reported [83]. However, all previous studies did not aim to evaluate the industrial IoT framework trust or provide a ranking and selection approach among the security features. For this, TrustE-VC aims at addressing this research gap and copes with the modern evaluation and selection methodology.

Kayes et al. proposed a context-sensitive access control that supports control decisions when there are dynamic changes to the context [61] and context-aware access control using fuzzy logic [63]. Meanwhile, in [62] the context-aware access control policies at the runtime is specified, and a pluggable single-sign-on authentication module is suggested in [23]. While addressing the trust challenges of the ITS using cutting edge, big data frameworks was discussed in [22] with a multi-tier VC architecture. The work in [4] proposed a machine learning algorithm for relay attack detection in the VC.

On the other hand, securing a vehicular network using a fuzzy trust model based on experience and plausibility to ensure the reliability of vehicle communications was reported [105]. While many evaluation models of cloud computing have been proposed in the literature [45, 100, 7], they do not yet provide practical analysis for security designers of cloud-based IoT and VC systems. Aiming at improving the intelligent transportation system, Bui and Jung [29] used a dynamic decision-making approach for CVs. However, evaluating multiple conflicting criteria in decision making has not been a subject of intensive studies in the literature. This advanced analytic method aids in better decision making to choose prioritized security improvement actions and, hence, ensure the trustworthiness of the industrial CV environment. In this chapter, we address this main open issue by proposing a novel evaluation approach.

4.4 VC Security Evaluation Framework

Defining the evaluation criteria is crucial to evaluating the performance parameters of the target system (i.e., VC). In this section, diagrammatic VC security levels are proposed in Table 4.1. These levels are used to assess the VC and then identify the unimproved gaps for further enhancements. In total, this study proposes 52 different evaluation metrics, six in level 1, ten in level 2, and thirty-six in level 3. Level 1, namely, common security criteria (CSCs), are hierarchically divided into 10 security control components (SCCs) in level 2. Those SCCs are broken down into 3 specific security control subcomponent (SCS) techniques in level 3. The final aim of this table is to provide a unified benchmark (evaluation criteria) to evaluate the security of VC services within any proposed solution.

The TrustE-VC methodology contains a sequence of structured processes, which are described using well-defined activities (i.e., inputs and outputs). In Figure 4.1, we propose a matching methodology for instantiating trust requirements in a context pattern (structural descriptions) of the VCs. The main processes of this method include preparation, examination, and decision. The preparation stage starts with identifying the evaluation target (with suitable variables) and describing objects that are subject to evaluation. Evaluation criteria point out the characteristics and constraint parameters of this target by weighting these criteria (optimal weight vector of the criteria). These evaluation criteria are further aggregated and associated with a group of decisions to obtain the collective trust weight and start the data gathering (of the available solution domain), which evaluates the rates of each criterion. This aggregation must be assigned a fuzzy rate of fuzzy best and worst values to be evaluated properly [110]. Based on the evaluation results, the correlation among the criteria weight and ideal point (the security gap) is scored and rated. In other words, evaluate the normalized fuzzy difference to evaluate the fuzzy index value. This score can be utilized to report particular pieces of information and prepare a mitigation plan that, eventually, updates any security breach or vulnerability by determining the rank of the trust criteria mode.

Our evaluation framework is composed of three main components to assist with the security gaps of a typical VC environment: (1) Aggregation of the evaluation values of the security levels, i.e., SCCs and SCSs in the decision-making method (group decision makers (GDMs)), is proposed. Evaluation of the security level of the industrial CV based on a singular perception framework (one DM) can return poor decisions [112]. Hence, to acquire a reasonable resolution, the use of GDMs is a suitable and relevant approach to knowledge synthesis and collection. In [112], the authors' results obtained from GDMs are more objective,

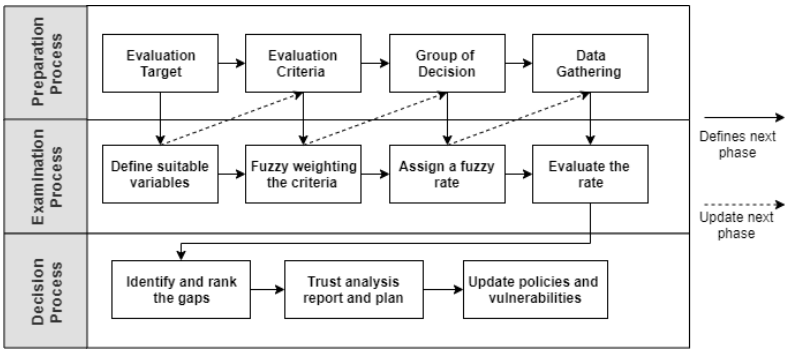


Figure 4.1: Context-based pattern matching for a trust by design of TrustE-VC framework

as they combine different experiences and views. (2) Fuzzy set theory and fuzzy aggregation techniques are used to evaluate the security level of industrial CV criteria according to the GDMs as a fuzzy multicriteria decision-making (MCDM) approach. (3) Simple additive VIKOR associated with the performance analysis and performance rate is used to visualize the framework findings, as proposed in [7].

4.4.1 Multicriteria Group Decision Making

To evaluate security solutions in a VC system using several components, we have to assess all of its criteria. Usually, any computing system cannot function well with all evaluation criteria, and hence, the investigator or security analyst has to map the trade-off among them. This is especially relevant within large-scale architectures as in a typical industrial CV. Evaluating and selecting the best security criteria tuning is the primary goal of our framework. Herein, we propose a multidecision algorithm that feeds the fuzzy ranking model of our proposed evaluation framework.

Next, we review some concepts regarding realization of the preliminaries of the criterion space, fuzzy set theory, and the decision making approach, which combined represent the basic principle of our proposed framework.

1. Representing the criterion space:

$$\max i; \quad \text{subject to} \quad i \in I \tag{4.1}$$

where i is the vector of $\mathbf{Y}$ criteria and I is the feasible set, $I \subseteq GY$. If G is defined explicitly (a set of choices), the result is called a multiple-criteria (MC) analysis. If I is defined implicitly (by a set of constraints), the result is named an MC process.

2. Representing the fuzzy sets:

To present the main concepts of fuzzy sets that link elements to define their membership to a function, which is usually [0,1], the membership degree is generally a figure (special fuzzy set), where χ pick rate on the real line, and $\tilde{q}((\chi))$ is a continuous mapping of U to the closed rate in the interval [0,1] as follows:

$$\tilde{q} = (\chi, \bar{q}(\chi)), \chi \in U \tag{4.2}$$

3. Representing the decision making:

The decision-making process matches criteria to a set of potential decisions accessible to the security designers and analytics. The criteria weight are the values of the security evaluation. Hence, investigators set identical security metrics in the decision model based on the specified security criteria. For example, when designing and implementing a VC solution, the implementer chooses the design parameters (security criteria). Each of these influences the security measures that evaluate the system components. Mathematically, an evaluation framework can be described as shown below:

$$\begin{aligned} &\max i = f(e) = f(e_1, \ldots, e_n) \\ &\text{subject to}, e \to E; \text{ as follows:} \\ &i \in I = f(e) : e \in E, E \subseteq I^n \end{aligned} \tag{4.3}$$

where E is the feasible set, and e is the decision variable vector of size n. A well-developed specific case is achieved when E is a polyhedron defined by linear inequalities and qualities. Different definitions are fundamental in **TrustE-VC**; these are closely linked to the nondominance and efficiency defined based on both space and variable representations.

Definition 1: if there exists $e^* \in E$, then $e^* \in E$ is not dominated when $e \geq e^*$ and $e \neq e^*$.

Definition 2: $e^* \in E$ is efficient if there does not exist another $e \in E$ such that $f(e) \geq f(e*)$ and $f(e) \neq f(e*)$. A key factor for a successful **evaluation framework** is a good representation of the decision situation by a problem domain.

Definition 3: if there does not exist another $i \in I$ where $i > I^*$, then we can say that $i^* \to I$ is weakly nondominated.

Definition 4: if there does not exist another $e \in E$ such that $f(e) > f(e^*)$, then $e^* \to E$ is weakly efficient as well, including all nondominated and some other particular points.

These unique points appear in practice, which makes them vital. Moreover, it is essential to distinguish them from nondominated points. To illustrate this case, consider that we maximized a particular objective, i.e., security criteria. Doing so may return a weakly nondominated point that is dominated. The dominated points of the weakly nondominated set are located either on vertical or horizontal planes (hyperplanes) in the criterion space.

Ideal point: shows the highest (the best for the maximization weight) of each criteria and compares favorable to an unfeasible decision.

Nadir point: shows the lowest (the worst for the maximization weight) of each criteria among the evaluation set. The ideal point and the nadir point are useful in the evaluation process to obtain the "quality" of the range of solutions.

Algorithm 1 illustrates our evaluation framework approach. Furthermore, it can indeed be realized by a GDM model that feeds the fuzzy ranking model of our proposed framework. Mathematically, problems corresponding to the above arguments can be represented as the following formal definitions:

By defining the most suitable ϕ^* and the worst $\widehat{\phi}$ values of all values of all test criteria, $i = 1,2,\ldots,n$, $\phi^* = \max(\phi_j, j = 1,\ldots,J)$ and $\widehat{\phi} = \min(\phi_j, j = 1,\ldots,J)$ if the i^{th} function is a benefit, whereas $\phi^* = \min(\phi_j, j = 1,\ldots,J)$ and $\widehat{\phi} = \max(\phi_j, j = 1,\ldots,J)$ if the i^{th} function is a cost.

Here, $S^* = \min(S_j), j = 1,\ldots,J, \widehat{S} = \max(S_j, j = 1,\ldots,J), R^* = \min(R_j), j = 1,\ldots,J$, and $\widehat{R} = \max(S_j, j = 1,\ldots,J)$. Eq. 4.5 is used to compute the group utility S_j, which is used to normalize the distance, and Eq. 4.4 is used to calculate the individual regret R_j in order to normalize the Chebyshev distance.

$$T_j = \sum_{i=1}^{n} w_i \left(\frac{\phi^* - \phi_j}{\phi^* - \widehat{\phi}} \right) \tag{4.4}$$

$$S_j = \max \left[w_i \left(\frac{\phi^* - \phi_j}{\phi^* - \widehat{\phi}} \right) \right] \tag{4.5}$$

Then, the comprehensive sorting index I_j is computed using the following equation:

$$I_j = v \left(\frac{S_j - T^*}{\widehat{T} - T^*} \right) + (1 - v) \left(\frac{S_j - S^*}{\widehat{S} - S^*} \right) \tag{4.6}$$

where $j = 1,2,\ldots J$, $R^* = \min(R_j)$, $\widehat{R} = \max(R_j)$, $S^* = \min(S_j)$ and $\widehat{S} = \max(S_j)$.

Algorithm 1: Multidecision approach for weighting security criteria in the VC framework

Require: Criterion functions $f(y)_i$
Determine best ϕ_i^* and worst $\widehat{\phi_i}$ values $\forall \quad f(y)_i$;
if i^{th} *function represents a benefit* **then**
  $\phi^* = \max(\phi_j)$, $\widehat{\phi} = \min(\phi_j)$
else
  $\phi^* = \min(\phi_j)$, $\widehat{\phi} = \max(\phi_j)$
end
Compute both values S_j and R_j.
Calculate the comprehensive sorting index I_j
Rank the fuzzy values R, S and I in ascending order.
Select the $\min(I_j)$ that represent the best ranked; the alternative $A^{(1)}$ is proposed as a compromise solution.
if $C1$ *is Acceptable advantage* **then**
  $I(A^{(2)} - I(A^{(1)}) \geq \frac{1}{m-1}$, where $A^{(2)}$ is the alternative with the second position in the ranking list by I.
end
if $C2$ *is Acceptable stability in decision making* **then**
  The alternative $A^{(1)}$ must also be the best ranked by T or/and S.
end

This approach is introduced as a weight for the strategy of maximum group utility, whereas $1-v$ is the weight of the individual regret. A compromise between these strategies could be reached by setting $v = 0.5$, and here, v is modified as $\frac{(n+1)}{2n}$ (from $v + 0.5\frac{(n-1)}{n} = 1$) since the criterion (1 of n) related to R is also included in S.

4.4.2 Fuzzy Set Theory Modeling

Due to the subjectivity of the anonymous values of the security criteria, the evaluation of the security level of industrial CV is imprecise and arguably vague. This imprecision issue requires novel decision-making approaches that address the subjective evaluations. Fuzzy set theory has been proposed as a pioneering solution, which aids in different areas [121]. In this study, we employed this theory to express and manage ambiguity in decision making. The linguistic variables in the fuzzy theory (e.g., very high, very low, low, and high) can afford a powerful connection tool—by assigning a numerical variable within a binary set (0,1). These

linguistic variables effectively model the vagueness or fuzziness inherent in decision-making problems [112].

In this research, we utilized the fuzzy triangular numbers [65] that describe linguistic variables connected with a membership degree of 0 or 1. This criteria enables modeling of fuzzy operations with both convenience and simplicity. A triangular fuzzy number is a fuzzy number represented by three points (K_1^L, K_2^M, K_3^H), where $(k_1 < k_2 < k_3)$.

According to [65], in fuzzy triangular numbers, any membership functions of fuzzy number A can be defined as follows:

$$\left\{ \begin{array}{cc} 0, & x < K_1^L \\ x - \frac{K_1^L}{K_2^M} - K_1^L, & K_1^L \leq x \leq K_2^M \\ K_3^L - \frac{x}{K_3^M} - K_2^L, & K_2^L \leq x \leq K_3^M \\ 0, & x \leq K_3^H \end{array} \right\} \tag{4.7}$$

A fuzzy multicriteria selection approach for analyzing the performance of industrial CV communications to the cloud was implemented by employing intuitionistic fuzzy numbers. In the proposed model, linguistic terms are used to rate the alternatives via criteria weighting and their corresponding fuzzy numbers.

4.4.3 Adaptive evaluation approach

To handle the problem of portfolio selection and aggregation of decisions, the SAW approach is proposed and the method in [33, 76] is used. Due to its simplicity and ability to identify unimproved gaps of alternatives, SAW has become the most popular decision-making (DM) approach. According to [112, 33] SAW is considered to be straightforward and can easily handle DM queries, motivated by its linear additive function, which can individually represent DM decisions. An empirical study [123] applied SAW and found superiority in both performance and simplicity. The essential principle of SAW is to calculate and categorize the weighted sum of the performance degrees for every criterion group.

The following equation describes this process:

$$A = \sum_{i=1}^{u} x_y x_{ij}, \quad j = 1, 2, \ldots, u-1, q = 1, 2, \ldots, v-1 \tag{4.8}$$

For identifying evaluation criteria to achieve the ideal level of any tested standard, a multiattribute model called importance-performance analysis (IPA) was reported [81]. It is com-

posed of a dimension matrix, namely, "Importance" and "Performance", to explain evaluation criteria graphically. IPA has been utilized for analysis in different studies to allocate unimproved gaps between various services. The use of IPA aims to describe the security criteria associated with each industrial CV component. The IPA map is beneficial for deciding how best to allocate unimproved gaps between an actual industrial CV system and an ideal point depending on the evaluation criteria. In this study, we use the "performance rate" (PR) represented by the x-axis instead of "performance", while "global weight" (GW) constitutes the y-axis instead of "importance."

The GW is used to demonstrate the importance of the criteria sample and examine the reliability. Hence, the GW represents the importance and performance realization to improve the system reliability. The weights range from 0 to 1 according to the following equation:

$$W = W_x \in (W_1, W_2, \ldots, W_n) \tag{4.9}$$

$$BestPR = [(K-3-K_1)+(K-2-K_1)]/3+K_1, \forall i \tag{4.10}$$

$$GW = \frac{W}{\sum_{i=1}^{W_x} W_n,}, n = 1,2,\ldots,n_i \tag{4.11}$$

4.5 Analysis and Results

The proliferation of IoT devices had led to the generation of a considerable amount of heterogeneous data. The allocated data have to be kept in appropriate storage (e.g., the cloud), which is remotely accessible for processing. Consequently, these data can be used to learn a new pattern of behavior using a machine learning algorithm that embeds the intelligence into any system. With this growing trend, new security issues have arisen. Tremendous security criteria have been proposed to cope with these issues. However, our results demonstrate that all security criteria are not equal, which means that such criteria should not be governed and managed at the same level. Our novel framework TrustE-VC provides a trust measure of these different security criteria. It may be utilized to classify the importance of each criterion based on its distance from an ideal security point of the system.

Based on the results presented in Table 4.2, the average performance rate (APR) of TrustE-VC and the ideal point for each SCC are converted to crisp values by applying Eqs. 4.9 and 4.10, respectively, which identify the best PR among the set. The GWs related to all of these

Table 4.2: Highlighted TrustE-VC findings for the SCCs and SCSs.

SCC Criteria	**GW**	**APR**	**SCS Evaluation**	
		(*CV*)	Lowest PR	Highest PR
Infrastructure Design_C_{13}	0.820	0.556	$C_{111(0.580,0.886)}$	$C_{113(0.926,0.946)}$
Network Security_C_{11}	0.800	0.426	$C_{121(0.506,0.820}$	$C_{122(0.853,0.886}$
Object Security_C_{12}	0.840	0.648	$C_{132(0.605,0.811}$	$C_{131(0.820,0.926)}$
Authentication_C_{22}	0.875	0.485	$C_{212(0.459,0.506)}$	$C_{213(0.800,0.886)}$
Authorization_C_{21}	**0.600**	**0.416**	$C_{222(0.760,0.240)}$	$C_{221(0.926,0.946)}$
Wire Encryption_C_{31}	**0.604**	**0.343**	$C_{316(0.126,0.300)}$	$C_{313(0.906,0926)}$
Integrity_C_{41}	0.820	0.420	$C_{414(0.420,0.686)}$	$C_{415(0.820,0.906)}$
Data Availability_C_{52}	0.784	0.560	$C_{521(0.766,0.806)}$	—
Service Availability_C_{51}	0.755	0.546	$C_{511(0.820,0.840)}$	$C_{512(0.866,0.886)}$
Privacy & Governance_C_{61}	0.738	0.369	$C_{614(0.346,0.646)}$	$C_{613(0.811,0.926)}$
Overall Average	**0.799**	**0.478**	—	—

values are converted to fuzzy numbers using Eq. 4.11. Next, the main APRs are mapped against their GWs to graphically present a map depicting the SCC that is most in need of improvement (see Table 4.2). It can be observed that TrustE-VC achieved poor results in terms of authorization (C_{22}) and encryption (C_{31}). Thus, TrustE-VC should pay more attention to and find the best strategy to improve these criteria. Moreover, the GW and APR datasets in Table 4.2, with overall averages of 0.799 and 0.478, respectfully, were assigned to form an IPA analysis for the VC. This aims at defining the linkage of both the x- and y-axes for each IPA map.

Sensitive data should be located and transmitted securely throughout the whole life cycle to guarantee high data privacy of such a deployment architecture, for instance, integrating the IoT paradigm with the cloud for storage, processing, and data security purposes [106]. For example, encryption researchers have traditionally responded by focusing on multiple CSCs, such as transparent data encryption, end-to-end wire encryption, and data masking and tokenization. TrustE-VC can provide a blockchain of thing encryption services as a supplementary service to clients. Such an approach could let the client choose from the recommended encryption software and techniques within different layers. Moreover, TrustE-VC needs to raise the encryption level to include most elements existing at the TrustE-VC cloud level. Multiple level encryption, blockchain encryption, transparent data encryption, identity-based

encryption, attribute-based encryption, and wire encryption migration are all encryption routines to improve this SCC.

In contrast, C_{11}, C_{51}, C_{13}, and C_{52} were the SCCs that obtained the most significant attention from TrustE-VC. These four criteria were estimated as satisfactory in meeting customer needs, as they had the most leading APR rates. This analysis leads TrustE-VC to a significant realization that these criteria should be retained from the VC customer points of view. This assessment should help in attracting new VC customers and increasing business shares.

Table 4.2 shows that TrustE-VC has high-performance rates in most SCCs except for C_{22} and C_{31}, as they are far from the ideal point. Depending on these conclusions, recommendations are expected to enhance the unimproved gaps of TrustE-VC for these lower SCCs through various strategies. For instance, a strategy could be considered to improve the privacy with an appropriate access control level to include most components in the industrial CV system. Fault tolerance and recovery mechanisms, to operate appropriately under an incident or a failure, are other examples, to name a few. Only four out of six SCCs have a high APR for TrustE-VC, namely, data (C_{11} and C_{13}) and access control (C_{51} and C_{52}), in that they perform fairly against the ideal point. The remaining APRs of the SCCs are lower than the target point (ideal), in which each SCA criteria performs poorly against the ideal point. A primary assumption would be that these SCCs need an updated strategy (including installation of new hardware and tools) to be improved. Additionally, the table shows the highest and lowest value of each SCS criteria, except for data availability C_{52} which contains one SCS: replication C_{521}. Different SCCs within TrustE-VC need additional refinement to achieve the target level. These SCCs have a lower APR than the other SCCs, which does not imply that all SCCs have to be urgently improved.

In other words, it is necessary to identify the SCCs that need improvement action and those SCSs under each SCC that need further refinement. An IPA diagram for TrustE-VC is illustrated in Figure 4.2, 4.3, and 4.4 to achieve the above objectives. In this 3D plot of the framework findings, each CSC represents the x-axis. Meanwhile, the PR and GW of each SCS represent the y-axis and z-axis, respectively. To represent the fuzzy weight variance, $Cn = (CSC(Cn), PR(Cn), GW(Cn))$, and the final IPA$(Cn) = \sum CSC + PR(Cn) * GW(Cn)$.

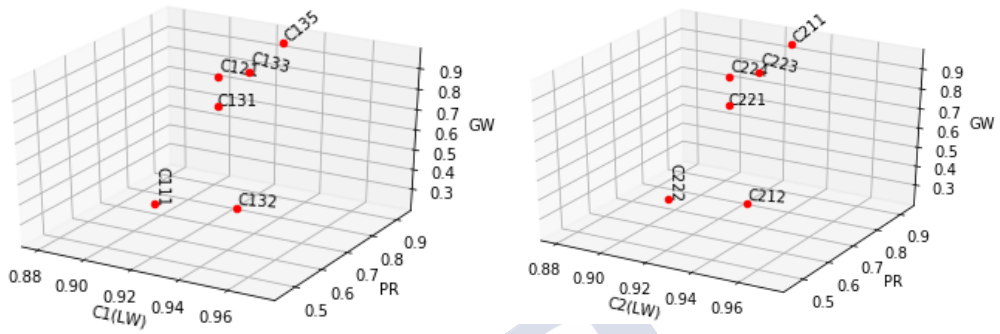


Figure 4.2: A fuzzy weight-variance 3D plot diagram for Physical Protection and Logical Access Control CSC and SCC in TrustE-VC.

4.6 Discussion

Standardizing the industrial IoT is an essential measure to widely accept and support its technology [104]. However, the standardization process of the trust methods faces several challenges, among which is the lack of an evaluation approach of these standards and solutions. A study that aims at evaluating the trust of current industrial IoT solutions with a focus on the security criteria selection is indispensable. Along this line, this study proposes a methodological approach to comprehensively assess a leading application in this realm, i.e., industrial CV.

This research contributes to the deployment of the industrial VC security body of knowledge by first examining in detail the building blocks of the industrial VC trust architecture. Second, a rigorous and robust evaluation framework based on evaluation theory is presented, which guides VC service providers to identify security gaps. Finally, we highlight some open challenges and recommendations for both service providers and customers for a comprehensive discussion toward achieving the vision of providing trustworthy IoV-cloud secure services.

The proposed approach combines the use of criteria importance and performance rates for determining those trust service attributes to which a designer or policy maker should devote

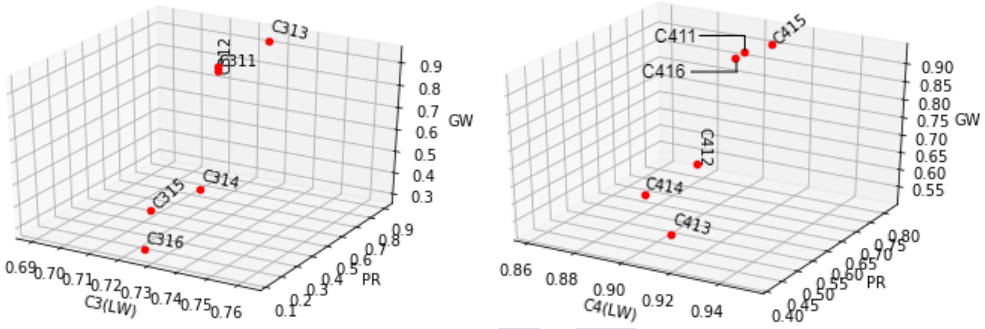


Figure 4.3: A fuzzy weight-variance 3D plot diagram for Communication Confidentiality and Communication Integrity CSC and SCC in TrustE-VC.

more attention. Also it labels which feature should be a lower priority to keep the focus on the high-priority ones. First, the proposed approach uses the three-level security evaluation elements as a classification tree to analyze all the security elements in an industrial CV environment. Next, a multicriteria decision-making algorithm is applied for comprehensively weighing these criteria. Together, they offer a useful and practice-ready tool for designers and industrial CV practitioners to better evaluate and select industrial CV trust requirements.

4.6.1 Security by Design

Security by design is a development approach that ensures that security is considered from the start of system deployment, and not as an additional late phase, to operate and maintain the trustworthiness of industry 4.0 technologies [84]. Due to its inherently remote operations, resource co-tenancy, distributed management, and administrative control, ensuring the privacy of IoT-based workloads while outsourcing computation is crucial. Industrial CV clients do not have direct control over the systems that utilize their data because of the cloud's black-box nature. In this context, modeling and optimization of IoT feature selection is an emerging trend [3].

To this end, our study addresses an emerging research gap of optimizing the security

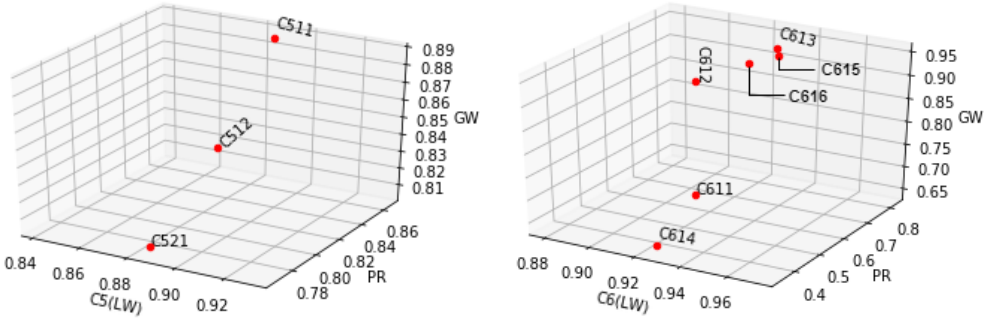


Figure 4.4: A fuzzy weight-variance 3D plot diagram for Data and Service Availability, as well as Data Privacy and Governance CSC and SCC in TrustE-VC.

features of an implemented industrial VC solution by evaluating and ranking these features for optimal selection in the design phase.

Based on our test observation, the most pressing challenges in assessing IoT data protection before a move to the cloud are as follows:

- Data residency: This refers to the physical geographic location where the data stored in the cloud reside. When deploying a cloud-based IoV system, the physical location of the data is no longer known or fully trusted. Data residency also includes data flow channels, data stream processing, and edge data input/output.

- Data privacy: This describes the ability to limit data sharing in industrial CV systems, including with third parties, through an organization or individuals. Maintaining an appropriate data privacy level can be achieved by exploring various technologies and tools, including encryption. Other solutions include modifying policies and legislation to prevent unauthorized access or use of data. Defining the legal ownership, responsibilities, and privileges of data between the owner and data custodian can alleviate privacy threats.

- Data ownership: Defining data ownership is a serious concern within IoV-to-cloud data processing. When a client transfers his or her data to the cloud, the primary processor of that data is then not the physical owner, but the provider. Consequently, a new threat parameter is raised regarding trust for that provider. The client cannot be sure how the cloud system manipulates his or her data or whether the processing complies with his or her demands.

4.6.2 Future trends

In this chapter, we address one of the leading open issues regarding industrial CV adoption. Namely, we evaluate the security criteria during the design phase of a cloud-based IoT application. We present a rigorous and robust evaluation model called TrustE-VC. The proposed framework formalizes and generalizes the main ideas proposed in the literature for empirical evaluation and selection of security criteria. TrustE-VC's main contribution is to help VC platforms identify the standard security criteria and provide security gap analysis according to a novel multicriteria decision-making evaluation theory for supporting industrial CV systems as a commodity service in the cloud. Diagrammatic security levels, a novel evaluation theory, and a fuzzy ranking approach based on additive weight and CV security analysis comprise the critical framework components.

Future studies can focus on the following aspects. First, it would be interesting to employ TrustE-VC on other VC implementations (frameworks) to capture qualitative risk evaluation information in a highly complex decision environment. In particular, the approach could be implemented with industrial CV providers to compare and evaluate the results of those providers to identify the most trustworthy providers in today's market. Second, utilization of new MCDM utility approaches, such as SWARA (step-wise weight assessment ratio analysis) and WASPAS (weighted aggregated sum product assessment) [80], can be applied to extend TrustE-VC to cope with other scenarios and trust analyses. Finally, the proposed TrustE-VC can be readily applied to extend the 52 security and trust criteria this study investigates. In fact, due to the wide trust area to evaluate, this study focused on the threats associated with industrial IoV-to-cloud security threats. Hence, it is worthwhile to conduct broader research that focuses on other trust criteria, such as privacy, governance, auditability, compliance, availability, and competence. The extension of these trust evaluation criteria should involve as much as possible the criteria that influence the trust of a suitable industrial CV solution. Finally, integration of this evaluation framework with the edges of the vehicular network modes [41],

e.g., V2V and V2I communication, is also considered a possible research trend.

4.7 Summary

A large body of research aims to address the security concern posed by the data transmission of connected vehicles with various insights. They do not yet provide practical analysis for security by design of cloud-based IoT and VC systems. Ensuring sustainable security integration of industrial CVs in the cloud environment with systematical security evaluation and selection has been limited in this context. This study intends to investigate a VC evaluation to offer assurances of the functional security properties of VC deployment architectures. The proposed TrustE-VC framework aims to express imprecise trust evaluation information to facilitate multiple-criteria decision analysis within industrial CV environments.

This framework provides a theoretical contribution based on the evaluation theory outlined by (1) categorizing a diagrammatic security taxonomy for security evaluation criteria in industrial CV clouds, (2) promoting a GDM ranking technique for better security criteria selection, and (3) introducing a fuzzy evaluation and ranking technique to evaluate and classify the unimproved security vulnerabilities in IoV-to-cloud deployment architectures. It also contributes to leveraging fuzzy sets and fuzzy IPA to accurately use the DM responses to prioritize the CSCs, SCCs, and SCS to achieve a better coverage for enhancing unimproved security vulnerabilities in current and future industrial CV environments. TrustE-VC aids in better decision making to choose prioritized security improvement actions and, hence, ensure the trustworthiness of the VC environment. Overall, the use of TrustE-VC as a basis to develop VC applications has proven the robustness and reliability of such a framework. As future work, the usage of the proposed methodology and framework will be considered to evaluate other IoT-applications and cyber-physical systems in the industrial IoT with larger-scale security features.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 5

CONCLUSIONS

This chapter provides an overview of the study, presents conclusions drawn from results, and outlines the future research.

Cloud computing is being steadily adopted as one of the dominant paradigms of BD platforms. The new concepts offered by the cloud—such as computation outsourcing, resource sharing, and external data warehousing—increase privacy concerns and security threats. BD frameworks, as an emerging technology domain, lack model-driven engineering to secure IaaS clouds. The software development methodology offered in this work focuses on creating conceptual models that abstract the security-solutions domain—delivering reference modules, basic requirements, main characteristics, and best practices for securing BD-cloud operations. These concerns and security threats have been addressed in this research by drafting security models for BD cloud adoption. This study proposes a component model that manages to standardize terminology, define key components and their relationships, collect relevant solution patterns, and categorize existing technologies. It also presents a reference architecture for big data systems that is focused on addressing security concerns associated with IaaS cloud deployment architectures. This research demonstrates how to use this model for the development of practical security solutions.

5.1 Extensive Summary

This thesis has successfully achieved the targeted objectives listed in Section 1.5 and has achieved the following contributions:

In Chapter 2, we analyze the security services used in IaaS cloud environments and describe BD security items and relationships amongst them. We discuss security systems oriented to BigCloud design in order to present their glossary and landscape techniques and to define research gaps and best practices. At first, this chapter extensively examines the building blocks of the cloud security stack for supporting BD science and classifies the different layers of security based on their supported service models into a reference architecture. An additional contribution is to classify the related security management technologies and services, which is based on the Hadoop stack, and to survey related work. Second, it examines the vulnerabilities associated with the BigCloud adoption by providing the security components of a secure design pattern and its attributes. Third, it provides many insights into BigCloud security specifications by refining the cloud context-pattern into a novel security analysis pattern. This pattern maps the current technologies to the solution domain by extending the CIA (Confidentiality, Integrity, and Availability) triad. Next, the study analyzes and classifies the state-of-the-art security frameworks available today, mostly as open source, for a detailed criteria election. Finally, it highlights some open challenges and recommendations for both service providers and customers, for a comprehensive discussion towards achieving the vision of providing a secure BigCloud service.

In Chapter 3, we identify the common security criteria (CSC) from a context-based analysis pattern and discuss, compare, and aggregate a conceptual model of these CSC impartially. A privacy granularity classification that maintains data confidentiality alongside the security selection criteria is proposed. This work contributes to the VC security deployment body of knowledge by facilitating the realization of secure internet of vehicles (IoV) systems in the VC model by identifying the CSC. The chapter starts by examining in detail the building blocks of the VC security stack for supporting IoV science. Then, it classifies the different layers of security based on their supported service models into a conceptual model. Next, it provides various insights into VC privacy by introducing a new security analysis pattern, which maps the current VC architecture layers into the privacy granularity specifications of the solution domain. Finally, we propose a security evaluation conceptual model that consists of 6 criteria that influences VC security formation. The model methodology consists of security control elements (SCEs) that generally influence both the client and provider security formation within the VC environment.

In Chapter 4, we address one of the leading open issues regarding VC adoption: evaluating the security criteria during the design phase. We present a rigorous and robust evaluation

model called TrustE-VC. The proposed framework formalizes and generalizes the main ideas proposed in the literature for empirical evaluation and selection of security criteria. The main contribution of TrustE-VC is to help VC platforms to identify the standard security criteria, providing security gap analysis according to multicriteria decision-making evaluation theory for supporting industrial CV applications as a commodity service in big data as a service cloud. Critical framework components entail diagrammatic security levels, a novel evaluation theory, and a fuzzy ranking approach based on an additive weight and VC security analysis. Before we conclude this chapter, we extend these SCEs through identifying 39 security control components (SCCs) and security control subcomponents (SCSs) that aimed to develop an effective VC trust solution.

5.2 Future Work

This research opens different directions and works in the field of BigCloud security management. The thesis and the proposed security by design provide new insights and benefits to the security implementation in BigCloud. Also, it contributes to the security development of BigCloud and to improve the understanding of security concepts in its associated frameworks. The underlying goal is to help make sure that BigCloud security is developed systematically with scientific validation principles. The result of this plan will be a set of tools to improve approaches to securing the system features in such an environment. Continuing research is aimed at developing an enhanced BigCloud implementation.

Future work for this study would also involve studying the evaluation framework over different use-cases and scenarios. It includes refining the syntax and defining the semantics of the proposed reference model as well as mapping the reference architecture to different cloud security architectures and big data frameworks. In the future, more security evaluation criteria are needed to be carried out on different BigCloud delivery models such as IaaS, PaaS, and SaaS, based on specific customer requirement scenarios. This goal will ensure the robustness of the framework, which can be adapted to suit the need of other BigCloud as for IoT-to-Cloud practices and data channels. Further adaptation of the framework is also being considered to ensure that the classification of security requirements and provisions can be automated into the mapping matrix to generate output for security analysis.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Bibliography

[1] Mohammad Aazam, Imran Khan, Aymen Abdullah Alsaffar, and Eui-Nam Huh. Cloud of things: Integrating Internet of Things and cloud computing and the issues involved. In *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th-18th January, 2014*, pages 414–419. IEEE, 2014.

[2] Nazri Abdullah, Anne Hakansson, and Esmiralda Moradian. Blockchain based approach to enhance big data authentication in distributed environment. In *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 887–892. IEEE, 2017.

[3] Awais Ahmad, Murad Khan, Anand Paul, Sadia Din, M Mazhar Rathore, Gwanggil Jeon, and Gyu Sang Choi. Toward modeling and optimization of features selection in Big Data based social Internet of things. *Future Generation Computer Systems*, 82:715–726, 2018.

[4] Usman Ahmad, Hong Song, Awais Bilal, Mamoun Alazab, and Alireza Jolfaei. Securing smart vehicles from relay attacks using machine learning. *The Journal of Supercomputing*, pages 1–18, 2019.

[5] Milad Ahvanooey, Qianmu LI, Xuefang Zhu, Mamoun Alazab, and Jing Zhang. ANiTW: A novel intelligent text watermarking technique for forensic identification of spurious information on social media. *Computers & Security*, 2019.

[6] Hamzeh Alabool, Ahmad Kamil, Noreen Arshad, and Deemah Alarabiat. Cloud service evaluation method-based multi-criteria decision-making: A systematic literature review. *Journal of Systems and Software*, 139:161–188, 2018.

[7] Hamzeh Mohammad Alabool and Ahmad Kamil Bin Mahmood. A novel evaluation framework for improving trust level of infrastructure as a service. *Cluster Computing*, 19(1):389–410, 2016.

[8] Mohammad Aladwan, Feras Awaysheh, Mamoun Alazab, Sadi Alawadi, Tomás Pena, and José Cabaleiro. Truste-vc: Trustworthy evaluation framework for industrial connected vehicles in the cloud. *IEEE Transactions on Industrial Informatics*, 2020.

[9] Mohammad Aladwan, Feras Awaysheh, José Cabaleiro, Tomás Pena, Hamzeh Alabool, and Mamoun Alazab. Common security criteria for vehicular clouds and internet of vehicles evaluation and selection. In *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications*, pages 814–820. IEEE, 2019.

[10] Mohammad N. Aladwan, Feras M. Awaysheh, Sadi Alawadi, Mamoun Alazab, Tomás F. Pena, and José C. Cabaleiro. TrustE-VC: trustworthy evaluation framework for industrial connected vehicles in the cloud. *IEEE transactions on industrial informatics*, 16(9):6203–6213, 2020.

[11] Asma Alshehri and Ravi Sandhu. Access control models for cloud-enabled Internet of Things: A proposed architecture and research agenda. In *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*, pages 530–538. IEEE, 2016.

[12] Amazon. Amazon EMR Web service manage cluster Cloud platform. `https://aws.amazon.com/emr/`, 2019. Accessed: 20/11/2019.

[13] Apache Bigtop. Comprehensive packaging, testing, and configuration of the Hadoop stack. `//https://bigtop.apache.org/`, 2019. Accessed: 20/11/2019.

[14] Apache Eagle. open source security analytics. `https://eagle.apache.org/`, 2019. Accessed: 20/11/2019.

[15] Apache Hadoop. Apache Hadoop 3.0 platform. `https://hadoop.apache.org/docs/r3.0.0/hadoop-project-dist/hadoop-hdfs/HDFSErasureCoding.html`, 2019. Accessed: 20/11/2019.

[16] Apache Hadoop. Hadoop in Secure Mode. `https://hadoop.apache.org/docs/r3.1.1/hadoop-project-dist/hadoop-common/SecureMode.html`, 2019. Accessed: 20/11/2019.

[17] Apache Hadoop. Transparent Encryption in HDFS. https://hadoop.apache.org/docs/r3.1.1/hadoop-project-dist/hadoop-hdfs/TransparentEncryption.html, 2019. Accessed: 20/11/2019.

[18] Apache Knox. EST API and Application Gateway for the Apache Hadoop Ecosystem. https://knox.apache.org/, 2019. Accessed: 20/11/2019.

[19] Apache Ranger. Comprehensive security managment for Enterprise Hadoop. https://ranger.apache.org, 2019. Accessed: 20/11/2019.

[20] Apache Sentry. Fine grained role based authorization. https://sentry.apache.org/, 2019. Accessed: 20/11/2019.

[21] Timothy Arndt. Big data and software engineering: prospects for mutual enrichment. *Iran Journal of Computer Science*, 1(1):3–10, 2018.

[22] Feras Awaysheh, José Carlos Cabaleiro, Tomás Fernández Pena, and Mamoun Alazab. Big data security frameworks meet the intelligent transportation systems trust challenges. In *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 807–813. IEEE, 2019.

[23] Feras M Awaysheh, José C Cabaleiro, Tomás F Pena, and Mamoun Alazab. A pluggable authentication module for Big Data federation architecture. In *Proceedings of the 24th ACM Symposium on Access Control Models and Technologies*, pages 223–225. ACM, 2019.

[24] Kristian Beckers, Holger Schmidt, Jan-Christoph Kuster, and Stephan Faßbender. Pattern-based support for context establishment and asset identification of the ISO 27000 in the field of cloud computing. In *2011 Sixth International Conference on Availability, Reliability and Security*, pages 327–333. IEEE, 2011.

[25] Gurjit Singh Bhathal and Amardeep Singh. Big data: Hadoop framework vulnerabilities, security issues and attacks. *Array*, 1:100002, 2019.

[26] Flavio Bonomi, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. Fog computing and its role in the Internet of Things. In *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pages 13–16. ACM, 2012.

[27] Alessio Botta, Walter De Donato, Valerio Persico, and Antonio Pescapé. Integration of cloud computing and Internet of Things: a survey. *Future Generation Computer Systems*, 56:684–700, 2016.

[28] Azzedine Boukerche and E Robson. Vehicular cloud computing: Architectures, applications, and mobility. *Computer Networks*, 135:171–189, 2018.

[29] Khac-Hoai Nam Bui and Jason J Jung. ACO-based dynamic decision making for connected vehicles in IoT system. *IEEE Transactions on Industrial Informatics*, 2019.

[30] Rajkumar Buyya, Satish Narayana Srirama, Giuliano Casale, Rodrigo Calheiros, Yogesh Simmhan, Blesson Varghese, Erol Gelenbe, Bahman Javadi, Luis Miguel Vaquero, Marco AS Netto, et al. A manifesto for future generation cloud computing: research directions for the next decade. *ACM computing surveys (CSUR)*, 51(5):105, 2018.

[31] Moumena Chaqfeh, Nader Mohamed, Imad Jawhar, and Jie Wu. Vehicular cloud data collection for intelligent transportation systems. In *2016 3rd Smart Cloud Networks & Systems (SCNS)*, pages 1–6. IEEE, 2016.

[32] Hongbing Cheng, Chunming Rong, Kai Hwang, Weihong Wang, and Yanyan Li. Secure big data storage and sharing scheme for cloud tenants. *China Communications*, 12(6):106–115, 2015.

[33] C West Churchman and Russell L Ackoff. An approximate measure of value. *Journal of the Operations Research Society of America*, 2(2):172–187, 1954.

[34] Cloudera. Cloudera Big Data cloud service provider. `https://hortonworks.com/`, 2019. Accessed: 20/11/2019.

[35] D. Cutting and M. Cafarella. Apache Hadoop. `http://hadoop.apache.org/`, 2016. Accessed: 20/11/2019.

[36] Jeffrey Dean and Sanjay Ghemawat. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.

[37] Yuri Demchenko, Cees De Laat, and Peter Membrey. Defining architecture components of the big data ecosystem. In *2014 International Conference on Collaboration Technologies and Systems (CTS)*, pages 104–112. IEEE, 2014.

[38] Manuel Díaz, Cristian Martín, and Bartolomé Rubio. State-of-the-art, challenges, and open issues in the integration of Internet of Things and cloud computing. *Journal of Network and Computer applications*, 67:99–117, 2016.

[39] Ittay Eyal, Adem Efe Gencer, Emin Gün Sirer, and Robbert Van Renesse. Bitcoin-ng: A scalable blockchain protocol. In *13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16)*, pages 45–59, 2016.

[40] Faezeh Farivar, Mohammad Sayad Haghighi, Alireza Jolfaei, and Mamoun Alazab. Artificial intelligence for detection, estimation, and compensation of malicious attacks in nonlinear cyber physical systems and industrial IoT. *IEEE transactions on industrial informatics*, 2019.

[41] Yixiong Feng, Bingtao Hu, He Hao, Yicong Gao, Zhiwu Li, and Jianrong Tan. Design of distributed cyber–physical systems for connected and automated vehicles with implementing methodologies. *IEEE Transactions on Industrial Informatics*, 14(9):4200–4211, 2018.

[42] Mohamed Amine Ferrag, Leandros A Maglaras, Helge Janicke, Jianmin Jiang, and Lei Shu. A systematic review of data protection and privacy preservation schemes for smart grid communications. *Sustainable cities and society*, 38:806–835, 2018.

[43] Jun-Song Fu, Yun Liu, Han-Chieh Chao, Bharat K Bhargava, and Zhen-Jiang Zhang. Secure data storage and searching for industrial IoT by integrating fog computing and cloud computing. *IEEE Transactions on Industrial Informatics*, 14(10):4519–4528, 2018.

[44] Iván García-Magariño, Sandra Sendra, Raquel Lacuesta, and Jaime Lloret. Security in vehicles with IoT by prioritization rules, vehicle certificates and trust management. *IEEE Internet of Things Journal*, 2018.

[45] Saurabh Kumar Garg, Steve Versteeg, and Rajkumar Buyya. A framework for ranking of cloud computing services. *Future Generation Computer Systems*, 29(4):1012–1023, 2013.

[46] Mario Gerla, Eun-Kyu Lee, Giovanni Pau, and Uichin Lee. Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds. In *2014 IEEE world forum on Internet of Things (WF-IoT)*, pages 241–246. IEEE, 2014.

[47] Ali Gholami and Ervin Laure. Big data security and privacy issues in the cloud. *International Journal of Network Security & Its Applications (IJNSA), Issue January*, 2016.

[48] Dan Gonzales, Jeremy M Kaplan, Evan Saltzman, Zev Winkelman, and Dulani Woods. Cloud-trust—a security assessment model for infrastructure as a service (IaaS) clouds. *IEEE Transactions on Cloud Computing*, 5(3):523–536, 2015.

[49] Google. Google Cloud Dataproc. `https://cloud.google.com/dataproc/`, 2019. Accessed: 20/11/2019.

[50] Maanak Gupta, James Benson, Farhan Patwa, and Ravi Sandhu. Dynamic groups and attribute-based access control for next-generation smart cars. In *Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy*, pages 61–72. ACM, 2019.

[51] Maanak Gupta and Ravi Sandhu. Authorization framework for secure cloud assisted connected cars and vehicular Internet of Things. In *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*, pages 193–204. ACM, 2018.

[52] Mohammad Hamdaqa, Tassos Livogiannis, and Ladan Tahvildari. A reference model for developing cloud applications. In *CLOSER*, pages 98–103, 2011.

[53] Liu Hao and Dezhi Han. The study and design on secure-cloud storage system. In *2011 International Conference on Electrical and Control Engineering*, pages 5126–5129. IEEE, 2011.

[54] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. The rise of "Big Data" on cloud computing: Review and open research issues. *Information systems*, 47:98–115, 2015.

[55] Benjamin Hindman, Andy Konwinski, Matei Zaharia, Ali Ghodsi, Anthony D Joseph, Randy H Katz, Scott Shenker, and Ion Stoica. Mesos: A platform for fine-grained resource sharing in the data center. In *NSDI*, volume 11, pages 22–22, 2011.

[56] Cheol-Ho Hong and Blesson Varghese. Resource management in fog/edge computing: a survey on architectures, infrastructure, and algorithms. *ACM Computing Surveys (CSUR)*, 52(5):1–37, 2019.

[57] Yin Huai, Ashutosh Chauhan, Alan Gates, Gunther Hagleitner, Eric N Hanson, Owen O'Malley, Jitendra Pandey, Yuan Yuan, Rubao Lee, and Xiaodong Zhang. Major technical advancements in Apache Hive. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1235–1246, 2014.

[58] IoT Analytics. Report on market insights for the security Internet of Things. `https://iot-analytics.com/new-iot-security-report/`, 2019. Accessed: 20/11/2019.

[59] Mohammad Islam, Angelo K Huang, Mohamed Battisha, Michelle Chiang, Santhosh Srinivasan, Craig Peters, Andreas Neumann, and Alejandro Abdelnur. Oozie: towards a scalable workflow management system for Hadoop. In *Proceedings of the 1st ACM SIGMOD Workshop on Scalable Workflow Execution Engines and Technologies*, pages 1–10, 2012.

[60] Alireza Jolfaei and Krishna Kant. Privacy and security of connected vehicles in intelligent transportation system. In *2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks–Supplemental Volume (DSN-S)*, pages 9–10. IEEE, 2019.

[61] ASM Kayes, Jun Han, Wenny Rahayu, Tharam Dillon, Md Saiful Islam, and Alan Colman. A policy model and framework for context-aware access control to information resources. *The Computer Journal*, 62(5):670–705, 2018.

[62] ASM Kayes, Wenny Rahayu, and Tharam Dillon. An ontology-based approach to dynamic contextual role for pervasive access control. In *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, pages 601–608. IEEE, 2018.

[63] ASM Kayes, Wenny Rahayu, Tharam Dillon, Elizabeth Chang, and Jun Han. Context-aware access control with imprecise context characterization for cloud-based data resources. *Future Generation Computer Systems*, 93:237–255, 2019.

[64] Minhaj Ahmad Khan and Khaled Salah. IoT security: Review, blockchain solutions, and open challenges. *Future Generation Computer Systems*, 82:395–411, 2018.

[65] George J Klir and Bo Yuan. Fuzzy sets and fuzzy logic: theory and applications. *Upper Saddle River*, page 563, 1995.

[66] Gregg Kreizman and Bruce Robertson. Incorporating security into the enterprise architecture process. *Gartner Research*, 2006.

[67] Kyriakos Kritikos and Philippe Massonet. An integrated meta-model for cloud application security modelling. *Procedia Computer Science*, 97:84–93, 2016.

[68] Ronald L Krutz, Russell Dean Vines, and Glenn Brunette. *Cloud security: A comprehensive guide to secure cloud computing*. Wiley Indianapolis, 2010.

[69] Eun-Kyu Lee, Mario Gerla, Giovanni Pau, Uichin Lee, and Jae-Han Lim. Internet of vehicles: From intelligent grid to autonomous cars and vehicular fogs. *International Journal of Distributed Sensor Networks*, 12(9):1550147716665500, 2016.

[70] Christian Lesjak, Daniel Hein, and Johannes Winter. Hardware-security technologies for industrial iot: Trustzone and security controller. In *IECON 2015-41st Annual Conference of the IEEE Industrial Electronics Society*, pages 002589–002595. IEEE, 2015.

[71] Yibin Li, Keke Gai, Longfei Qiu, Meikang Qiu, and Hui Zhao. Intelligent cryptography approach for secure distributed big data storage in cloud computing. *Information Sciences*, 387:103–115, 2017.

[72] Yibin Li, Keke Gai, Longfei Qiu, Meikang Qiu, and Hui Zhao. Intelligent cryptography approach for secure distributed big data storage in cloud computing. *Information Sciences*, 387:103–115, 2017.

[73] Chang Liu, Jinjun Chen, Laurence T Yang, Xuyun Zhang, Chi Yang, Rajiv Ranjan, and Ramamohanarao Kotagiri. Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates. *IEEE Transactions on Parallel and Distributed Systems*, 25(9):2234–2244, 2013.

[74] Chang Liu, Rajiv Ranjan, Chi Yang, Xuyun Zhang, Lizhe Wang, and Jinjun Chen. Murdpa: Top-down levelled multi-replica merkle hash tree based secure public auditing for dynamic big data storage on cloud. *IEEE Transactions on Computers*, 64(9):2609–2622, 2014.

[75] Yang Lu and Li Da Xu. Internet of Things (IoT) cybersecurity research: a review of current research topics. *IEEE Internet of Things Journal*, 6(2):2103–2115, 2018.

[76] Kenneth R MacCrimmon. Decisionmaking among multiple-attribute alternatives: a survey and consolidated approach. Technical report, The RAND Corporation, Santa Monica, CA, 1968.

[77] Sunilkumar S Manvi and Gopal Krishna Shyam. Resource management for infrastructure as a service (IaaS) in cloud computing: A survey. *Journal of Network and Computer Applications*, 41:424–440, 2014.

[78] MapR. MapR Big Data cloud service provider. `https://mapr.com/`, 2019. Accessed: 20/11/2019.

[79] Samuel Marchal, Xiuyan Jiang, Radu State, and Thomas Engel. A big data architecture for large scale security monitoring. In *2014 IEEE International Congress on Big Data*, pages 56–63. IEEE, 2014.

[80] Abbas Mardani, Mehrbakhsh Nilashi, Norhayati Zakuan, Nanthakumar Loganathan, Somayeh Soheilirad, Muhamad Zameri Mat Saman, and Othman Ibrahim. A systematic review and meta-analysis of swara and waspas methods: Theory and applications with recent fuzzy developments. *Applied Soft Computing*, 57:265–292, 2017.

[81] John A Martilla and John C James. Importance-performance analysis. *Journal of marketing*, 41(1):77–79, 1977.

[82] Microsoft. Microsoft AzureHDInsight. `https://azure.microsoft.com/`, 2019. Accessed: 20/11/2019.

[83] Mujahid Mohsin, Zahid Anwar, Ghaith Husari, Ehab Al-Shaer, and Mohammad Ashiqur Rahman. IoTSAT: A formal framework for security analysis of the Internet of Things (IoT). In *2016 IEEE Conference on Communications and Network Security (CNS)*, pages 180–188. IEEE, 2016.

[84] Haralambos Mouratidis and Vasiliki Diamantopoulou. A security analysis method for industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, 14(9):4093–4100, 2018.

[85] Mujahid Muhammad and Ghazanfar Ali Safdar. Survey on existing authentication issues for cellular-assisted v2x communication. *Vehicular Communications*, 12:50–65, 2018.

[86] Jianbing Ni, Aiqing Zhang, Xiaodong Lin, and Xuemin Sherman Shen. Security, privacy, and fairness in fog-based vehicular crowdsensing. *IEEE Communications Magazine*, 55(6):146–152, 2017.

[87] NIST. NIST Big Data Working Group (NBD-WG). `http://bigdatawg.nist.gov/`, 2019. Accessed: 20/11/2019.

[88] Michele Nitti, Virginia Pilloni, Giuseppe Colistra, and Luigi Atzori. The virtual object as a major element of the Internet of Things: a survey. *IEEE Communications Surveys & Tutorials*, 18(2):1228–1240, 2015.

[89] Serafim Opricovic and Gwo-Hshiung Tzeng. Compromise solution by mcdm methods: A comparative analysis of vikor and topsis. *European journal of operational research*, 156(2):445–455, 2004.

[90] Pekka Pääkkönen and Daniel Pakkala. Reference architecture and classification of technologies, products and services for big data systems. *Big Data Research*, 2(4):166–186, 2015.

[91] Alexander Palm, Zoltán Ádám Mann, and Andreas Metzger. Modeling data protection vulnerabilities of cloud systems using risk patterns. In *International Conference on System Analysis and Modeling*, pages 1–19. Springer, 2018.

[92] Raj R Parmar, Sudipta Roy, Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, and Tai-Hoon Kim. Large-scale encryption in the Hadoop environment: Challenges and solutions. *IEEE Access*, 5:7156–7163, 2017.

[93] Haixia Peng, Le Liang, Xuemin Shen, and Geoffrey Ye Li. Vehicular communications: A network layer perspective. *IEEE Transactions on Vehicular Technology*, 68(2):1064–1078, 2018.

[94] Johannes Pillmann, Benjamin Sliwa, Jens Schmutzler, Christoph Ide, and Christian Wietfeld. Car-to-cloud communication traffic analysis based on the common vehicle information model. In *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2017.

[95] Butch Quinto. Big data governance and management. In *Next-Generation Big Data*, pages 495–506. Springer, 2018.

[96] Noëlle Rakotondravony, Benjamin Taubmann, Waseem Mandarawi, Eva Weishäupl, Peng Xu, Bojan Kolosnjaji, Mykolai Protsenko, Hermann De Meer, and Hans P Reiser. Classifying malware attacks in IaaS cloud environments. *Journal of Cloud Computing*, 6(1):26, 2017.

[97] Chayan Sarkar, Akshay Uttama Nambi SN, R Venkatesha Prasad, Abdur Rahim, Ricardo Neisse, and Gianmarco Baldini. Diat: A scalable distributed architecture for iot. *IEEE Internet of Things journal*, 2(3):230–239, 2014.

[98] P Satheesh, B Srinivas, PRS Naidu, and B Prasanth Kumar. Study on efficient and adaptive reproducing management in hadoop distributed file system. In *Internet of Things and Personalized Healthcare Systems*, pages 121–132. Springer, 2019.

[99] Amartya Sen and Sanjay Madria. Risk assessment in a sensor cloud framework using attack graphs. *IEEE Transactions on Services Computing*, 10(6):942–955, 2016.

[100] A Shameli-Sendi, M Shajari, M Hassanabadi, M Jabbarifar, and M Dagenais. Fuzzy multi-criteria decision-making for information security risk assessment. *The Open Cybernetics & Systemics Journal*, 6(1), 2012.

[101] Joshua E Siegel, Dylan C Erb, and Sanjay E Sarma. A survey of the connected vehicle landscape—architectures, enabling technologies, applications, and development areas. *IEEE Transactions on Intelligent Transportation Systems*, 19(8):2391–2406, 2017.

[102] Ashish Singh and Kakali Chatterjee. Cloud security issues and challenges: A survey. *Journal of Network and Computer Applications*, 79:88–115, 2017.

[103] Saurabh Singh, Young-Sik Jeong, and Jong Hyuk Park. A survey on cloud computing security: Issues, threats, and solutions. *Journal of Network and Computer Applications*, 75:200–222, 2016.

[104] Emiliano Sisinni, Abusayeed Saifullah, Song Han, Ulf Jennehag, and Mikael Gidlund. Industrial Internet of things: Challenges, opportunities, and directions. *IEEE Transactions on Industrial Informatics*, 14(11):4724–4734, 2018.

[105] Seyed Ahmad Soleymani, Abdul Hanan Abdullah, Mahdi Zareei, Mohammad Hossein Anisi, Cesar Vargas-Rosales, Muhammad Khurram Khan, and Shidrokh Goudarzi. A secure trust model based on fuzzy logic in vehicular ad hoc networks with fog computing. *IEEE Access*, 5:15619–15629, 2017.

[106] Christos Stergiou, Kostas E Psannis, Byung-Gyu Kim, and Brij Gupta. Secure integration of IoT and cloud computing. *Future Generation Computer Systems*, 78:964–975, 2018.

[107] MingJian Tang, Mamoun Alazab, and Yuxiu Luo. Big data for cybersecurity: vulnerability disclosure trends and dependencies. *IEEE Transactions on Big Data*, 2017.

[108] Fei Tao, Ying Zuo, Li Da Xu, and Lin Zhang. IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing. *IEEE Transactions on Industrial Informatics*, 10(2):1547–1557, 2014.

[109] Hai Tao, Md Zakirul Alam Bhuiyan, Md Arafatur Rahman, Tian Wang, Jie Wu, Sinan Q Salih, Yafeng Li, and Thaier Hayajneh. TrustData: Trustworthy and secured data collection for event detection in industrial cyber-physical system. *IEEE Transactions on Industrial Informatics*, 2019.

[110] Zhang-Peng Tian, Jian-Qiang Wang, and Hong-Yu Zhang. An integrated approach for failure mode and effects analysis based on fuzzy best-worst, relative entropy, and VIKOR methods. *Applied Soft Computing*, 72:636–646, 2018.

[111] Ankit Toshniwal, Siddarth Taneja, Amit Shukla, Karthik Ramasamy, Jignesh M Patel, Sanjeev Kulkarni, Jason Jackson, Krishna Gade, Maosong Fu, Jake Donham, et al. Storm@ twitter. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 147–156. ACM, 2014.

[112] Gwo-Hshiung Tzeng and Jih-Jeng Huang. *Multiple attribute decision making: methods and applications*. Chapman and Hall/CRC, 2011.

[113] Blesson Varghese and Rajkumar Buyya. Next generation cloud computing: New trends and research directions. *Future Generation Computer Systems*, 79:849–861, 2018.

[114] Vinod Kumar Vavilapalli, Arun C Murthy, Chris Douglas, Sharad Agarwal, Mahadev Konar, Robert Evans, Thomas Graves, Jason Lowe, Hitesh Shah, Siddharth Seth, et al. Apache Hadoop YARN: Yet Another Resource Negotiator. In *Proceedings of the 4th annual Symposium on Cloud Computing*, page 5. ACM, 2013.

[115] Tian Wang, Md Zakirul Alam Bhuiyan, Guojun Wang, Lianyong Qi, Jie Wu, and Thaier Hayajneh. Preserving balance between privacy and data integrity in edge-assisted Internet of Things. *IEEE Internet of Things Journal*, 2019.

[116] Md Whaiduzzaman, Mehdi Sookhak, Abdullah Gani, and Rajkumar Buyya. A survey on vehicular cloud computing. *Journal of Network and Computer applications*, 40:325–344, 2014.

[117] Marko Wolf and Timo Gendrullis. Design, implementation, and evaluation of a vehicular hardware security module. In *International Conference on Information Security and Cryptology*, pages 302–318. Springer, 2011.

[118] Mingyuan Xia, Mohit Saxena, Mario Blaum, and David A Pease. A tale of two erasure codes in HDFS. In *13th USENIX Conference on File and Storage Technologies (FAST 15)*, pages 213–226, 2015.

[119] Tian Xia, Hironori Washizaki, Takehisa Kato, Haruhiko Kaiya, Shinpei Ogata, Eduardo B Fernández, Hideyuki Kanuka, Masayuki Yoshino, Dan Yamamoto, Takao Okubo, et al. Cloud security and privacy metamodel-metamodel for security and privacy knowledge in cloud services. In *MODELSWARD*, pages 379–386, 2018.

[120] Xianqing Yu, Peng Ning, and Mladen A Vouk. Enhancing security of Hadoop in a public cloud. In *2015 6th International Conference on Information and Communication Systems (ICICS)*, pages 38–43. IEEE, 2015.

[121] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[122] Matei Zaharia, Reynold S Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J Franklin, et al. Apache spark: a unified engine for big data processing. *Communications of the ACM*, 59(11):56–65, 2016.

[123] Stelios H Zanakis, Anthony Solomon, Nicole Wishart, and Sandipa Dublish. Multi-attribute decision making: a simulation comparison of select methods. *European journal of operational research*, 107(3):507–529, 1998.

[124] Li Zhu, Fei Richard Yu, Yige Wang, Bin Ning, and Tao Tang. Big data analytics in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1):383–398, 2018.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Figures

List of Tables