The Challenges of the Data Privacy in the Cloud using Docker.
A Systematic Review

João Paulo Marques Silva[1], **Alaim Assis Jr**[1], **Mauricio Wanderley Martins**[1], **Reutman Oliveira**[1]

[1]CESAR School
Recife – PE – Brazil

contato@cesar.school

Abstract. *Cloud Computing is a technology that provides on demand services to individuals, companies or governments using the Internet. The shared cost of Cloud infrastructure has made this technology one of the most important in recent years. On the other hand, both customers and providers are increasingly concerned about the privacy of their data in these distributed environments. This work is a systematic review of 20 selected studies that report on the challenges and difficulties in maintaining data security and privacy in Cloud computing technologies and more recently in Docker container technology.*

Resumo. *Cloud Computing é uma tecnologia que provê serviços on demand para pessoas, empresas ou governos utilizando a Internet. O custo compartilhado da infraestrutura em Cloud tornou essa tecnologia em uma das mais importantes dos últimos anos. Por outro lado, tanto clientes como providers se preocupam cada vez mais com a privacidade dos seus dados nestes ambientes distribuidos. Este trabalho é uma revisão sistematica realizada em 20 estudos selecionados que relatam sobre os desafios e dificuldades em manter a segurança e privacidade dos dados nas tecnologias de Cloud computing e mais recentemente na tecnologia de containeres em Docker.*

1. Introduction

Cloud Computing has revolutionized and continues to revolutionize IT in many areas, from the way customers receive content from providers to how companies do business, and is therefore considered a disruptive technology [Yu et al. 2012]. Cloud Computing provides on-demand services, reducing costs, sharing and configuring computing resources, and high scalability and flexibility in services offered [Makkaoui* et al. 2016]. It also has the advantages of fast deployment and virtualization, which allows the separation of business services from the infrastructures required to perform these services and also it allows lower energy consumption, due to infrastructure sharing, is considered a Green technology [Radwan et al. 2017]. Cloud Computing can be used virtually anywhere and offers numerous benefits to businesses, government and individual users. Even so, with this breakthrough and acceptance by the IT industry, the use of Cloud still raises concerns about privacy and information security, especially for users with confidential data that would be harmed if they were stolen. These factors are the main inhibitors for the adoption and use of Cloud computing services [Mohammad and Asadullah 2015].

The objective of this study is to carry out a systematic review on the challenges of the data privacy observed in the use of containers in the Cloud. The paper is organized

into four sections. Section II we will describe the methodology adopted in the systematic review and in Section III we will present the results obtained in the research and Section IV we will present the conclusions and future work.

2. Applied Protocol

The main objective of this study is to conduct a systematic review of publications that address the challenges, threats and techniques employed in security in the Cloud Computing environment and in Docker containers. To carry out this review, we decided to follow an approach based on the guidelines proposed by Kitchenham [A. Kitchenham 2007]. Our review then consists of 4 steps illustrated in figure 1.

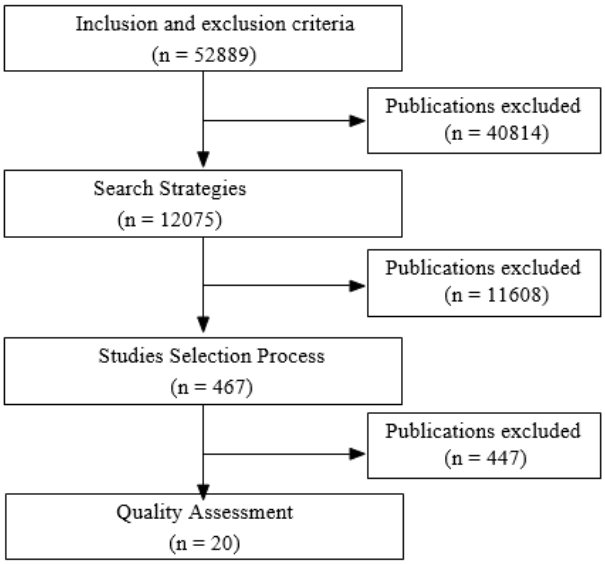

Figura 1. Kitchenham's Adapted Steps Protocol.

To guide us in the study we elaborate the main research questions:

- What are the impacts related to information security using containers in the Cloud?
- What are the difficulties related to data privacy in the Cloud?
- What security techniques and tools are being issued in the Cloud?
- What are the risks involved using Docker in the Cloud?

2.1. Inclusion and exclusion criteria

For this study we consider in our search only publications that address the challenges and problems faced, as well as techniques used in security and privacy in the Cloud as well as in publications that address the risks and problems faced in Docker container environments. We also restrict the search to a period of five years (2013 to 2017). Studies not published in English or that did not have all of their content available or that had as scope only to propose new security frameworks were excluded.

2.2. Search Strategies

Initially we conducted a search for "Cloud computing"in Scopus which resulted in 52,889 items in several repositories. Refining the Scopus search for "Cloud computing AND docker"we found 12,075 results.

Only the following scientific repositories were considered: ACM Digital Library (DL), IEEE Xplore (IEEE) and ScienceDirect (ScD).

In order to identify only those studies relevant to our review, we used the following combination of keywords:

- ”cloud computing”AND security;
- ”cloud computing”AND privacy;
- ”Docker”AND security;

In DL, ScD and IEEE the keywords were used to search for publications by title, abstract or keywords. Table 1 shows the studies found in each repository.

Tabela 1. Results in databases

Database	Amount of studies
ACM Digital Library	82
IEEE Xplore	203
ScienceDirect	182
Total	467

2.3. Studies Selection Process

In this stage, they were analyzed the studies about security and privacy in Cloud computing and Docker environments. At this stage we realized that several studies only dealt with security concepts in a general way applied to the Cloud and did not have the Docker technology in its scope. These studies found in scientific repositories addressed other issues such as frameworks, business technologies and performance comparisons. Therefore, we excluded studies that did not also relate to privacy or Docker technology. At this stage we eliminated 425 studies and identified 42 studies to be analyzed. After we perform the abstracts analysis of the 42 studies identified as potentially relevant to our review. In abstracts we identified studies that did not address the scope of our review. These studies dealt with comparisons of the use of Docker in test environments, virtualization simulations, discussions on distributed environments, performance comparisons and deployment in Cloud. Twelve studies were eliminated at this stage, remaining 30 studies for our systematic review.

2.4. Quality Assessment

In this stage a critical analysis of studies was carried out. This quality analysis took into account the relevance, quality, inclusion or exclusion criteria, validity, and data analyzed by the authors of the selected studies. According to Kitchenham [A. Kitchenham 2007], 8 questions were elaborated for quality analysis. The questions were:

- Are the objectives of the study clearly defined?
- Does the study report the threats that involve the use of technologies in cloud computing environments?
- Does the study report security techniques that can protect data in the cloud?
- Does the study report any risks involved in using Docker container technology in the cloud?

- Were the results presented adequately validated?

Of the remaining 30 studies, 10 were eliminated after the qualitative analysis - took into account the relevance, quality, inclusion or exclusion criteria, validity, and data analyzed by the authors of the selected studies- leaving 20 studies for our systematic review. The following section reports our results from the review of these studies.

3. Results and Discussion

Twenty studies were identified for this review. The studies report the key challenges and threats that write about security and privacy in the Cloud. Many of these studies are about security techniques that can be applied to maintain privacy in the Cloud. In these 20 studies selected for review, we have 58 different authors from 10 countries. Most of the publications are from India (4), the others are from United States (3), France (3), China (3), Turkey (2), Qatar (2). UK, Switzerland, Brazil, Malasy has one (1) study each. We found 59 different keywords in these studies. The main keywords found in the studies with their respective frequency were: cloud (40), security (33), computing (13), privacy (12), risk (5) and docker (4). All studies were analyzed according to the proposed quality issues.

3.1. What are the difficulties related to data privacy in the Cloud?

Overall, the reviewed studies demonstrate a common concern of users and companies that use Cloud services with the privacy of their data. There are difficulties in maintaining data security and privacy, especially because of the threats surrounding Cloud computing solutions. The studies point to several threats, which have a direct impact on the adoption of technology and its expansion. One of the main difficulties pointed out in the studies is related to the threats of insider attacks to organizations that can come from employees or people contracted to provide services in the premises of the company.

3.2. What are the risks involved in using Docker in the Cloud?

The Cloud container technology has successfully become the benchmark in the container and ecosystem market of DevOps. The studies analyzed demonstrate that Docker as Cloud technology also suffers from threats to data privacy. They report that there are vulnerabilities related to the insecure configuration of the production system, the distribution of images, verification, decompression, storage process, vulnerabilities of the images and the linux kernel and Docker itself. Another point questioned by the studies is that containers are less secure compared to virtual machines because the containers work heavily coupled with the host operating system and run on top of it, so if the container is compromised then the attacker can gain full access to the system and its resources.

3.3. What security techniques and tools are being used in the Cloud?

In general, the studies analyzed suggest security techniques based on the mentioned threats. For each type of threat the study defines the techniques to be used. In our review, the main techniques suggested in the studies are listed in table 2.

Tabela 2. Threats and Security Solutions in Studies

Threats	Security Solutions
Escape from container to host	Host updated and with strict access and security policies.
Get root access on the host	Secure network with TLS protocol.
Attack between containers	Avoid modifying docker default settings, which are already prepared for isolation between containers
DOS - Denial of Service	Host updated and with strict access and security policies, Secure network with TLS protocol.
Image Vulnerability	Use official images and always keep them updated according to the latest versions
Modifying container or host attributes	Do not give root privileges to Docker in daemon
Modification of Host	Do not deliver more than standard capabilities to the container capabilities.
Vulnerability Propagation	Apply security patch to the image before using it
Outdated images with permanent vulnerabilities	Regular updating of the image by the maintainers.
Buffer Overflow / Negação de Serviço	Apply security patch to the image before using it
Accidental Data Leakage	Maintain environment variables for sensitive data by maintainers.
Container Escape	Keeping the docker up to date, as well as the up-to-date host and strict security policies

4. Conclusion

Cloud Computing has played a very important role in today's technological revolution and modern society. With its way of delivering on-demand services, it reduces the initial costs for users and businesses and has therefore become increasingly the technology most widely used in IT business environments. Innovation does not stop and so new technologies are emerging in this environment, like Docker which with its container-based platform has become increasingly used by companies. However, the rapid growth of these technologies has also raised the concerns of providers and customers with the privacy of data stored in Cloud environments. Numerous research has been conducted to understand how to handle the challenges and difficulties of maintaining security and privacy in Cloud environments. In this systematic review we analyze and find some answers and also more questions about the theme. As a limitation of this study, we observed that studies on Docker technology are still recent and are in the early stages of development, and that many studies do not address solutions to data privacy issues, but only point to threats and vulnerabilities present in Cloud environments. In future work we will focus on Docker's security and privacy by proposing a layered model to help providers deal with these aspects in Docker.

Referências

A. Kitchenham, B. (2007). *Kitchenham, B.: Guidelines for performing Systematic Literature Reviews in software engineering. EBSE Technical Report EBSE-2007-01*.

Jian, Z. and Chen, L. (2017). A Defense Method against Docker Escape Attack. *Proceedings of the 2017 International Conference on Cryptography, Security and Privacy - ICCSP '17*, pages 142–146.

Kumar, K. and Kurhekar, M. (2017). Economically Efficient Virtualization over Cloud Using Docker Containers. *Proceedings - 2016 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2016*, pages 95–100.

Makkaoui*, K. E., Ezzati, A., Beni-Hssane, A., and Motamed, C. (2016). Cloud Security and Privacy Model for Providing Secure Cloud Services.

Manu, A. R., Patel, J. K., Akhtar, S., Agrawal, V. K., and Murthy, K. N. B. S. (2016). A Study , Analysis and deep dive on Cloud PAAS security in terms of Docker Container security A Study , Analysis and deep dive on Cloud PAAS security in terms of Docker Container security. *IEEE - 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT]*, (March):1–13.

Martin, A., Raponi, S., Combe, T., and Di Pietro, R. (2018). Docker ecosystem – Vulnerability Analysis. *Computer Communications*, 122:30–43.

Mohammad, A. and Asadullah, Z. (2015). Factors Influencing Information Privacy Concern in Cloud Computing Environment. 2015:238–242.

Pai, S. J., Kumar, A., and Gopal, A. (2016). Linux hardening techniques. (1):94–99.

Radwan, T., Azer, M. A., and Abdelbaki, N. (2017). Cloud computing security: challenges and future trends. *International Journal of Computer Applications in Technology*, 55(2):158.

Shu, R., Gu, X., and Enck, W. (2017). A Study of Security Vulnerabilities on Docker Hub. *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy - CODASPY '17*, pages 269–280.

Sun, Y., Zhang, J., Xiong, Y., and Zhu, G. (2014). Data Security and Privacy in Cloud Computing. *International Journal of Distributed Sensor Networks*, 2014:687–692.

Werner, J., Westphall, C. M., and Westphall, C. B. (2017). Cloud identity management: A survey on privacy strategies. *Computer Networks*, 122:29–42.

Yu, H., Powell, N., Stembridge, D., and Yuan, X. (2012). Cloud computing and security challenges. *Proceeding ACM-SE '12 Proceedings of the 50th Annual Southeast Regional Conference*, page 298.