Security in Cloud Computing: a Mapping Study

Belén Cruz Zapata¹, José Luis Fernández-Alemán¹, and Ambrosio Toval¹

¹ Faculty of Computer Science, Campus de Espinardo, 30100, University of Murcia, Spain
{b.cruzzapata, aleman, atoval}@um.es

Abstract. A number of cloud applications are currently widely used. However, one of the main reasons for the slowing down in the growth of cloud computing is that of security. Even though some research has been done in the security field, it is necessary to assess the current state of research and practice. This paper aims for the study of the existing research about security in cloud computing to analyze the state of art and to identify future directions. The method selected to investigate the security in cloud computing is a systematic mapping study. A total of 344 papers were selected and classified by security goal, research type and contribution type. The main security specific issues extracted are data protection (30.29%), access management (20.14%), software isolation (16.70%), availability (16.00%), trust (13.60%) and governance (3.27%). Our results demonstrate that cloud computing seems to be a promising area for security research and evaluation.

Keywords: cloud computing, security, mapping study.

1. Introduction

The US National Institute of Standards and Technology (NIST) describes Cloud Computing, from now on CC, as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” [1].

Many cloud applications are currently widely and successfully used (i.e. Google App Engine, Amazon's Computer Cloud Amazon Web Service and Microsoft Azure Service Platform). Although CC is a growing technology, no key player leads this revolution. The cloud saves money and has the backing of many large software vendors [2]. However, one of the main reasons why the growth of CC has slowed down is that of security, which is the greatest challenge of CC [3]. A few examples of this type of threats are two real incidents that occurred in 2009. One of them is Salesforce.com, which suffered an outage that locked more than 900,000 subscribers out of crucial CC applications and data needed to transact business with customers [4]. Another example is that of the smartphone known as “Sidekick” when users (over 800,000) temporarily lost personal data, which was accessed as a cloud service [5].

CC systems are secure if users can depend on them to behave as would be expected. Security in terms of information security is a condition that results from the establishment and maintenance of protective measures that enable an enterprise to
perform its mission or critical functions despite risks posed by threats to its use of information systems. Protective measures may involve a combination of deterrence, avoidance, prevention, detection, recovery and correction that should form part of the enterprise's risk management approach [6]. 5 goals have traditionally been associated with the achievement of adequate security: availability, confidentiality, data integrity, access control and audit [7]. The National Institute of Standards and Technology (NIST) has published the Guidelines on Security and Privacy in Public Cloud Computing, which provides an overview of the security and privacy challenges pertinent to public CC and points out considerations that organizations should bear in mind when using the cloud environment [8].

Despite the research that has been carried out in the field of CC security, it is necessary to assess the current state of research and practice in order to provide practitioners with evidence that will enable them to focus on its further development. The method selected to investigate the state-of-the-art is a systematic mapping study. A systematic mapping study provides an objective procedure with which to identify the nature and extent of the research that is available to answer a particular research question. These kinds of studies also help to identify gaps in current research in order to suggest areas for further investigation. Some surveys summarize risks and recommendations but do not follow procedures for a systematic mapping study [9] [10]. Silva et al. [11] performed a mapping study that focuses on accounting models for CC, classifying them into three categories: contribution type, research type and accounting models features. They provide an overview of the area, in addition to specific findings related to the taxonomy for accounting process, accounting models, pricing schemes and SLA (Service-Level Agreement) composition. The analysis done in [12] examines the different existing approaches in the literature about migration processes to CC while taking into account the security aspects that have to be also moved to Cloud. The review of cloud security proposals developed by Rebollo et al. [13] compare existing information security frameworks that have been specifically designed for the CC environment using the clauses from the ISO/IEC 27002 standard as evaluation criteria. The same authors also conduct a systematic literature review [14] to extract existing Information Security Governance frameworks that are suitable for application in CC. To the best of our knowledge, no systematic mapping study on security in CC currently exists.

The aim of this paper is therefore to study the existing research concerning security in CC in order to analyze the state-of-art and to identify future directions. The method consists of the application of a systematic mapping study to extract as much literature as possible. The 344 papers selected have been classified by their security goal, research type and contribution type.

The article is organized as follows: Section 2 shows how the mapping study was planned by defining the research questions and describing how the review was carried out: search strategy, inclusion and exclusion criteria and classification scheme. Section 3 synthesizes the results obtained by answering the research questions raised. In Section 4 we discuss the findings of our research and present the limitations of the study. Finally, in Section 5 we draw our conclusions.
2. **Mapping Study**

According to Petersen et al. [15], a systematic mapping study provides a structure of the type of research reports and results that have been published by categorizing them. It often provides a visual summary, a map, of its results. A number of research questions must be defined in order to obtain these objectives in a systematic manner. The main goals of a structured mapping study are to present an overview of a certain research area and to identify research gaps.

A systematic literature review is another kind of secondary study that answers specific research questions by identifying, analyzing and interpreting relevant evidence. A systematic mapping study, unlike a systematic literature review, does not analyze the articles identified in detail [15].

Petersen et al. propose five process steps when performing a systematic mapping study [15]. In the first step, the research questions are developed in order to define the scope. The next two phases include searching for and selecting primary studies by screening the articles found. The data is then extracted and synthesized. In a last phase the review results are obtained and the research questions are answered.

### 2.1. **Research Questions and Search Strategy**

The principal goal of this study is to analyze publications concerning the security in CC. This overall goal is defined in four research questions. Table 1 shows these research questions along with their motivation.

The main digital libraries used to search for primary studies were: IEEE Digital Library, IEEE-Xplore, ACM Digital Library, Science Direct and Wiley InterScience. These publication databases were chosen since they are some of the most relevant sources in software engineering.

Petersen et al. recommend defining the search string by using the PICO criteria: population, intervention, comparison and outcome. They also state that it should be driven by the research questions. When defining a search string, the resulting set should have the maximum possible coverage, but should also be of a manageable size. As the research questions suggested, a broad overview of the research area was desired, and some restrictions, such as temporal restrictions or specific security areas restrictions were therefore avoided in order to construct a complete map. In addition to the selected terms, boolean operators were used to form the complete search string which is composed of three parts. In the ACM Digital Library this search string is defined as:

```
“cloud computing” AND
(“security” OR “risk” OR “vulnerability” OR “threat”) AND Title: “cloud”.
```

The search process took place from November 2012 to December 2012. This process was conducted by applying the same search string to the same meta data (title or complete text) of each article for all the sources, although the search string's syntax was adapted to each digital library. A restriction on publication date was not established, but the papers obtained had all been published in the last few years.
Table 1. Research questions.

<table>
<thead>
<tr>
<th>No.</th>
<th>Research question</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>What are the main security areas investigated in cloud computing?</td>
<td>The first research question defines the basis of this systematic mapping study and provides an overview of recent studies.</td>
</tr>
<tr>
<td>RQ2</td>
<td>What kinds of research and contributions have been proposed?</td>
<td>This question provides information about the concrete techniques and solutions that have been proposed.</td>
</tr>
<tr>
<td>RQ3</td>
<td>How publications have evolved over time? What are the research and publication trends?</td>
<td>The third question shows the evolution of the publications concerning the subject under study.</td>
</tr>
<tr>
<td>RQ4</td>
<td>In which bibliographical sources were they published?</td>
<td>This last question examines the different sources in which the articles concerning security in CC are published.</td>
</tr>
</tbody>
</table>

2.2. Screening of Papers for Inclusion and Exclusion

The aim of the selection process was to identify those articles that are most relevant for the objective of this mapping study. Each study that was retrieved from the previous search was evaluated in order to decide whether or not it should be included by considering its title, abstract and keywords.

The studies that met at least one of the following inclusion criteria were included:
- Studies reporting some topic related to security in CC.
- Studies presenting actual security risks in CC.
- Studies proposing a solution in order to guarantee security in CC.

The following studies were excluded from the search results:
- Studies reporting a particular topic related to CC but not security in CC.
- Studies that mention security in CC but in a superficial manner.
- Studies concerning CC technology applied to a concrete application.
- Studies concerning security in CC but from a legal perspective.
- Studies concerning security in CC but from a business perspective.

The search phase obtained a set of 1757 papers, most of which were found in the IEEE Digital and ACM Digital libraries. As shown in Fig. 1, 500 articles were considered based on article title and on the inclusion and exclusion criteria. In the second phase, 84 duplicated studies were removed, resulting in 416 articles being included in the next phase. In the third phase, 72 articles were discarded and 344 were selected based on the abstract and the sub-section titles. The number of selected articles that remained was 344. The complete list of selected articles can be downloaded from [16].

The following step in the study was to read all the abstracts in order to assign keywords to each paper. With some papers, it was necessary to read the introduction, conclusion or even section titles to be able to extract these keywords.
2.3. **Classification Scheme**

The classification scheme proposed in this paper was adapted from Petersen et al. [15], with the exception of the Security Goal facet that was obtained from the NIST Guidelines on Security and Privacy in Public Cloud Computing [8]. The scheme consists of three facets: security goal, research type and contribution type (see Fig. 2). Classification schemes are rated on the basis of a set of quality attributes [17]:

− Orthogonality. There are clear boundaries between categories, which makes it easy to classify.
− Defined based on existing literature. The taxonomy/classification is created on the basis of an exhaustive analysis of existing literature in the field.
− Based on the terminology used in literature. The taxonomy uses terms that are used in existing literature.
− Complete. No categories are missing, thus allowing existing articles to be classified.
− Accepted. The community accepts and knows the classification/taxonomy.

---

**Fig. 1.** Selection process of primary studies.
Security goal

There are different approaches for classification of security aspects in cloud computing. Fernandes et al. [18] propose a taxonomy that covers eight categories: software, storage and computing, virtualization, Internet and services, network, access, trust, and compliance and legality. Gruschka and Jensen [19] present a taxonomy for attacks on cloud services. The ISO/IEC DIS 27017 [20] is a standard in progress, which is being specifically tailored to cloud services and defines guidelines to support the interpretation and implementation of cloud-specific information security controls supplementing the guidance in ISO/IEC 27002. The ISO 27799, an implementation of ISO/IEC 27002 in health informatics [21] [22], has been already used in secondary studies [23] [24]. The Security goal facet used in this article is based on the NIST Guidelines on Security and Privacy in Public Cloud Computing [8], which is a publication accepted by the community in the field. This facet includes six categories based on different key security and privacy issues identified by NIST (see Table 2).

The 4.2 section of the guidelines is about compliance. Compliance is the company responsibility to operate in agreement with established laws, regulations, standards and specifications, so as it was mentioned in the exclusion criteria, compliance is out of this study bounds. Section 4.4 about the architecture is really distributed among the rest of
the categories; due to the proposed architectures can have the objective of protecting data, ensuring the availability or keeping the software isolation.

- Governance. Governance implies control and oversight by the organization over policies, procedures and standards for application development and information technology service acquisition, as well as the design, implementation, testing, use and monitoring of deployed or engaged services [8].

- Trust. Trust in the cloud can be viewed as the customers’ level of confidence in using the cloud. Confidence can be increased by mitigating technical and psychological barriers to using cloud services [25].

- Identity and access management. This category includes the identity proofing and authentication aspects, and the prevention of unauthorized access to information resources in the cloud [8]. Access control systems provide identification and authentication services (who can log on), authorization (what a subject can do) and accountability (what a subject did) [6].

- Software isolation. High degrees of multi-tenancy over large numbers of platforms are needed for CC to achieve the envisioned flexibility of on-demand provisioning of reliable services and the cost benefits and efficiencies due to economies of scale. Multi-tenancy in IaaS (Infrastructure as a Service) CC environments is typically done by multiplexing the execution of virtual machines from potentially different consumers on the same physical server [8]. Virtualization is the simulation of the software and/or hardware upon which other software runs. This simulated environment is called a virtual machine (VM) [26].

- Data protection. Data stored in the cloud usually resides in a shared environment with data from other customers. Organizations moving sensitive data into the cloud must account for the means by which access to the data is controlled and is kept secure [8]. Papers which describe measures to ensure data protection and data security, in particular data confidentiality and integrity are classified in this category.

- Availability. This category refers to the availability of information resources and may be affected by technical issues, natural phenomena or human causes. The dynamic provisioning of a cloud simplifies the work of an attacker to cause a denial of service. Papers that describe investigations concerning availability in CC are mainly solutions with which to avoid denial-of-service attacks [6] [27].

The security categories are not mutually exclusive so that one paper can discuss both availability and data protection issues. The keywords used to classify the papers into these categories are listed in Table 3.

<table>
<thead>
<tr>
<th>Category/Guidelines section</th>
<th>Security element</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Governance</td>
<td>Audit</td>
</tr>
<tr>
<td>4.3 Trust</td>
<td>Audit</td>
</tr>
<tr>
<td>4.5 Identity and Access Management</td>
<td>Access control</td>
</tr>
<tr>
<td>4.6 Software isolation</td>
<td>Integrity, Confidentiality</td>
</tr>
<tr>
<td>4.7 Data Protection</td>
<td>Integrity, Confidentiality</td>
</tr>
<tr>
<td>4.8 Availability</td>
<td>Availability</td>
</tr>
</tbody>
</table>
Table 3. Keywords for security goal

<table>
<thead>
<tr>
<th>Security Goal</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>denial of service, intrusion, attack, malware, detection, infrastructure</td>
</tr>
<tr>
<td>Data protection</td>
<td>data protection, integrity, confidentiality, privacy, encryption</td>
</tr>
<tr>
<td>Governance</td>
<td>governance, policies, procedure, standard, requirement, rule</td>
</tr>
<tr>
<td>Identity and access</td>
<td>identity, access, privilege, authentication, key, cryptography, certificate</td>
</tr>
<tr>
<td>management</td>
<td></td>
</tr>
<tr>
<td>Software isolation</td>
<td>software isolation, multi-tenancy, virtualization, virtual machine (VM), hypervisor, environment</td>
</tr>
<tr>
<td>Trust</td>
<td>trust, risk management, transparency, credibility, reliability, data location, monitor</td>
</tr>
</tbody>
</table>

Research type

The research type facet is based on the schema proposed by Wieringa et al. [28]. This facet has six categories. The keywords used to classify the papers into these categories are listed in Table 4.

Table 4. Keywords for research type

<table>
<thead>
<tr>
<th>Research type</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation research</td>
<td>test, result, simulation, emulator, analysis, experiment, prototype</td>
</tr>
<tr>
<td>Evaluation research</td>
<td>evaluation, implementation, result, platform, case study, production</td>
</tr>
<tr>
<td>Solution proposal</td>
<td>solution, proposal</td>
</tr>
<tr>
<td>Philosophical paper</td>
<td>philosophical paper</td>
</tr>
<tr>
<td>Opinion paper</td>
<td>discuss, survey, suggests</td>
</tr>
<tr>
<td>Experience paper</td>
<td>experimental</td>
</tr>
</tbody>
</table>

- Validation research. A validation research has not been implemented in practice and focuses on the validation of the solution in the lab or in CC simulation scenarios. Validation studies should include discussions concerning limitations. A CC validation research states hypotheses, uses summary statistics and describes the main components of an experimental setup.

- Evaluation research. Evaluation research has been implemented in practice and shows the solution implementation and what the consequences of the implementation in a CC environment are in terms of benefits and drawbacks. Evaluation research can be excluded if no industrial cooperation or real world project is mentioned.

- Solution proposal. A solution proposal is a new technique or a significant extension to an existing one. Its benefits are exemplified or argued, but there is no complete validation in a CC environment.

- Philosophical proposal. This proposal shows a point of view as regards the subject without the preciseness of a solution proposal. These papers sketch a new way of looking at existing things by structuring the field in the form of a taxonomy or conceptual framework.
Opinion paper. An opinion paper reports the author's opinion of what is good or bad. In this study, opinion papers are those in which the author carries out a theoretical study or proposes CC security recommendations.

Experience paper. An experience paper reports on personal experiences from a real life CC project. Research methodology is not included.

Contribution type

These categories are extracted from Petersen's article [15]. They have been slightly adapted to CC security through, for example, the creation of the ontology category. An ontology represents the vocabulary of a domain, therefore, for a recent discipline as CC, it should be common to create ontologies to clarify the structure of knowledge [29]. The keywords used to classify the papers into these categories are listed in Table 5.

- Tool. Papers proposing a tool related to CC. A privacy manager or a decision support tool, are some examples of this.
- Method. Papers describing a new method. Encryption methods, secure protocols (such as non-repudiation protocols) or communication protocols were identified as methods.
- Model. Papers proposing models such as system architectures or CC frameworks (in the sense of an architecture framework, not a software framework).
- Survey. Papers exposing risks and vulnerabilities in CC, but in which no solution is proposed.
- Ontology. Papers proposing an ontology to identify and discuss the information that will be exchanged in order to preserve security in CC.
- Testbed. Papers proposing a testbed that enables researchers to study different aspects of CC.

Table 5. Keywords for contribution type

<table>
<thead>
<tr>
<th>Contribution type</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool</td>
<td>tool, demo, implementation, manager, execution</td>
</tr>
<tr>
<td>Method</td>
<td>method, mechanism, technique, protocol, algorithm</td>
</tr>
<tr>
<td>Model</td>
<td>model, framework, architecture, implementation, system, scheme</td>
</tr>
<tr>
<td>Survey</td>
<td>survey, categorizes, recommendation, best practices</td>
</tr>
<tr>
<td>Ontology</td>
<td>ontology, property, connection</td>
</tr>
<tr>
<td>Testbed</td>
<td>testbed, evaluate, debug</td>
</tr>
</tbody>
</table>

3. Results

Each article was classified into the categories of each facet in order to answer the four research questions. The results of the systematic mapping study are presented as follows.
3.1. **RQ1: What are the main security areas investigated in cloud computing?**

In order to be able to answer this question, the security goal facet was created. Its results are shown in Table 6. Most of the papers deal with data protection (30.29%), followed by identity management (20.14%), software isolation and availability with very similar percentages (16.70% and 16.00%), trust (13.60%) and the least studied topic is that of governance (3.27%).

199 of 344 papers (57.85%) discuss about just one topic at a time and just the NIST guidelines include all the security issues [8]. Those 199 papers keep similar percentages for the different categories so that they imply the major contribution to the global ranking: 75 are about data protection, 44 about identity management, 28 about software isolation, 28 about availability, 18 about trust and 6 about governance.

Data protection is related to privacy, integrity and confidentiality topics. To achieve data protection authors address privacy managers, encryption schemes (34 papers), audit practices (16 papers), noise generation strategies (5 papers) or storage methodologies (4 papers). For example, Cachin et al. surveyed some cryptographic tools and concluded that these kinds of solutions are academic and should be evaluated in practice [30]. With regard to noise generation strategies, Zhang et al. [31] propose a strategy based on historical probability as a means to reduce the number of noise requests. One example of the use of storage methodology to maintain data secure is that described by Subashini et al., which consists of a fragmentation technique based on the sensitivity and value of the data [32].

Identity management is covered in its most by access control architectures, even so, alternative issues can be found such as watermarking methods, authentication frameworks or encryption schemes. Software isolation is done by virtualization techniques. Availability involves intrusion and DoS (Denial-of-Service) attack detection systems and audit practices, such as [33] et al. that offer a solution to detect the source of a DoS attack and filter the attack traffic.

**Table 6. Number of papers by security goal**

<table>
<thead>
<tr>
<th>Security goal</th>
<th>Number of papers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data protection</td>
<td>176</td>
<td>30.29%</td>
</tr>
<tr>
<td>Identity and access management</td>
<td>117</td>
<td>20.14%</td>
</tr>
<tr>
<td>Software isolation</td>
<td>97</td>
<td>16.70%</td>
</tr>
<tr>
<td>Availability</td>
<td>93</td>
<td>16.00%</td>
</tr>
<tr>
<td>Trust</td>
<td>79</td>
<td>13.60%</td>
</tr>
<tr>
<td>Governance</td>
<td>19</td>
<td>3.27%</td>
</tr>
</tbody>
</table>

3.2. **RQ2: What kinds of research and contributions have been proposed?**

Upon examining the contribution type facet results in Table 7 it was discovered that most of the solutions are model proposals, that is, secure architectures for CC systems. Some examples are presented to illustrate the utility of these architectures. Zhang et al. [34] who present an information risk management framework in order to better understand critical areas of focus in a CC environment. This framework also helps as
regards identifying threats and vulnerabilities. Lin and Squicciarini [35] propose a new data protection framework that addresses challenges during the life cycle of a CC service. This framework consists of three key components: policy ranking, policy integration and policy enforcement. With regard to dataflow processing, Juan Du et al. present ROSIA, a robust service integrity attestation system whose purpose is to process stateful dataflow applications in CC [36]. There is also a framework proposal for mobile environments. Jarabek et al. [37] develop a cloud-based anti-malware system for Android. Performance and data consumption of this system is then evaluated using 50 Android applications downloaded from Google Play market.

Table 7. Number of papers by contribution type

<table>
<thead>
<tr>
<th>Contribution type</th>
<th>Number of papers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>143</td>
<td>41.57%</td>
</tr>
<tr>
<td>Method</td>
<td>96</td>
<td>27.91%</td>
</tr>
<tr>
<td>Survey</td>
<td>92</td>
<td>26.74%</td>
</tr>
<tr>
<td>Tool</td>
<td>7</td>
<td>2.03%</td>
</tr>
<tr>
<td>Ontology</td>
<td>5</td>
<td>1.45%</td>
</tr>
<tr>
<td>Testbed</td>
<td>1</td>
<td>3.27%</td>
</tr>
</tbody>
</table>

96 out of 344 papers propose methods (27.91%). The methods are secure communication protocols, cryptographic algorithms with which to encrypt data and measures to detect problems in system integrity or to measure the trustiness of the cloud resources. Some representative papers selected to show the goal of these methods are Kwon et al. and Wang et al.; Kwon et al. [38] propose measures to detect anomalies or abnormal activities. Wang et al. [39] propose a method to efficiently share confidential data on cloud servers by combining the hierarchical identity-based encryption system and the ciphertext-policy attribute-based encryption system. Zhang et al. [40] propose an integrity verification scheme that allows a third party auditor to verify the integrity of multiple stored replicas.

Around 26.70% of the papers are surveys. These papers are theory studies in which the actual situation of CC is explained. Some also make recommendations, such as Dahbur et al. [9] who make the necessary hints that can help promote the benefits of and mitigate the risks associated with CC. Hay et al. [41] suggest some future directions for security research and development to help advance the security posture of this technology.

Table 8. Number of papers by research type

<table>
<thead>
<tr>
<th>Research type</th>
<th>Number of papers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution proposal</td>
<td>134</td>
<td>38.95%</td>
</tr>
<tr>
<td>Validation research</td>
<td>107</td>
<td>31.10%</td>
</tr>
<tr>
<td>Opinion paper</td>
<td>83</td>
<td>24.13%</td>
</tr>
<tr>
<td>Evaluation research</td>
<td>13</td>
<td>3.78%</td>
</tr>
<tr>
<td>Philosophical paper</td>
<td>6</td>
<td>1.75%</td>
</tr>
<tr>
<td>Experience paper</td>
<td>1</td>
<td>0.29%</td>
</tr>
</tbody>
</table>

Only 7 papers propose a tool (2.03%). One of the first tools published was a client-based privacy manager that helps reduce the risk of data leakage or loss of privacy, and
provides additional privacy-related benefits by reducing the amount of sensitive information sent to the cloud [42]. D. Tancock et al. [43] propose the Privacy Impact Assessment (PIA) decision support tool that can be integrated within a CC environment. There are another 5 papers that create an ontology or taxonomy, and a further one which proposes a test bed [44].

Regarding the research type facet in Table 8, 38.95% of the papers are solution proposals, i.e., solutions that have not been evaluated in real practical scenarios. Solutions validated in experiments or prototypes represent around 31.10%, while evaluation research papers represent around 3.78%. In evaluation research, the solution has been implemented in practice, while the experiments validate the solution in the lab or in simulation scenarios. Of 13 evaluation papers, only 4 are case studies. To better understand the contribution of the papers located in these categories, some representative examples are provided. Balduzzi et al. conduct the evaluation of a real system [45], in which the security problems of public images that are available on the Amazon EC2 service are investigated. Lo et al. [46] propose a framework of cooperative intrusion detection system (IDS) to reduce the impact of denial-of-service (DoS) attack or distributed denial-of-service (DDoS) in a CC environment. The authors validate the proposed system with experimental results that indicate that the framework could resist DoS attack. Dhage et al. [47] propose an architecture capable of detecting intrusions in a distributed CC environment and safeguarding it from possible security breaches, but no validation results are mentioned.

![Fig. 3. Number of papers by contribution and security facets.](image)

Opinion papers represent around 24.13% (83 papers). These papers are reviews of security that provide personal recommendations. They are closely related to surveys from the contribution type facet. The remaining 1.75% of the studies belongs to philosophical papers. Only one experience paper has been found [48]. In this paper, Oza et al. identify and marry a range of issues related to user experience and security.

Some additional graphics have been created to allow the relations between the facets to be observed. The goal security facet is compared in Fig. 3 and Fig. 4. Most of the papers belong to the data protection category and most of them are solution proposals.
and opinion papers, respectively. Fig. 5 represents the contribution type and investigation type facets.

**Fig. 4.** Number of papers by research and security facets.

**Fig. 5.** Number of papers by research and contribution facets.

### 3.3. RQ3: How have publications changed over time?

Four different years could be identified, 2009, 2010, 2011 and 2012. More than 40% (139 articles) of the articles was published in 2012. Fig. 6 and Fig. 7 show the distribution of the papers by year. The evolution of publications by research type (see Fig. 7) is the same as the general evolution (see Fig. 6), the number of papers gets
bigger and bigger, doubling each year. This fact is observed in almost every single category: validation research, solution proposals, philosophical papers, opinion papers and evaluation research.

Fig. 6. Number of papers by year.

3.4. **RQ4: In which sources were they published?**

This question is answered by extracting the conference or the journal in which the papers were published. The Journal Citation Report (JCR) has been used to evaluate the recognition and stability of a journal with a systematic and objective system. JCR is an evaluation mechanism based on statistic information from reference data. The Computing Research and Education (CORE) system has been used to evaluate conferences. This is a Computing Research and Education Association of Australasia ranking mechanism.

More than 73.26% of the articles (252 articles) were presented at conferences and 23.84% of them were published in journals. The remaining articles were not published in either conferences or journals, but rather in books. Note the variety of different publishing sources: 132 conferences and 40 journals were identified.

In the journal ranking, the top one is IEEE Security and Privacy together with Future Generation Computer Systems: 9 papers. The next, with 6 papers, is the Journal of Computer Law and Security Review, but most of the remaining journals have just one or two selected papers. In the conference ranking, the top one is CLOUDCOM with 26 papers. CCSW is in second position with 11 papers; and DASC is in third position with 9 papers.
Discussion

The current situation of security in CC has been analyzed by examining the research and new advances published. Based on the results of this systematic mapping study, the first finding is that CC is a recent discipline since the papers found were only published in 2009, 2010, 2011 and 2012. Moreover the fact that not many papers report on tools or solutions evaluated in real scenarios (2.03% and 3.78%) also demonstrates this statement. Validations (31.10%) by way of experiments are more frequent than evaluations (3.78%) that involve case studies or implementations, but the most common research type consists of solution proposals, which only provide the scenario and direction of the solution. Case studies represent only 1.16% of papers. This situation is particularly critical for security in CC since conducting and examining real-world case studies allows best practices for CC to be learnt and established. These results coincide with those of authors who describe CC as a new technology. CC has gained increased attention during the past three years, and would appear to be a promising area for further research.

Each year the number of papers increases but the growth rate is lower (Fig. 6). The number of papers published in 2009 increased by 150% in 2010, by 75% from 2010 to 2011 and by 22% from 2011 to 2012.

We performed a regression analysis on this data and found there is a strong positive correlation between year and the number of publications. In order to predict the trend of
publications in the future, we have tried to determine a trend line for this data using several common regression models: Linear, Logarithmic, Polynomial, Power, Exponential and Moving average. The quadratic model is the best-fit line we found, which achieves a coefficient of determination of $R^2 = 0.9924$. To put the CC growth trend into a wider context, according to DBLP (Trier Universität, 2012), the general growth in computer science is exponential. Fitting an exponential model to our data, a coefficient of $R^2 = 0.9238$ is obtained.

Upon extracting the evolution of the research type over the years, some interesting information can be found (Fig. 10). The proportion of papers that only propose solutions has decreased: 50% (2009), 47.69% (2010), 39.47% (2011) and 32.37% (2012). Instead, validation research papers have increased over the years: 19.23% (2009), 23.08% (2010), 31.58% (2011) and 36.69% (2012). This provides some evidence to suggest that the CC is starting to move from foundational theory to practical validation, possibly a sign of puberty.

Of all the extracted security-related topics, data protection is that which is most frequently studied, owing to the fact that a significant barrier to the adoption of cloud services is the fear of leakage of sensitive data or loss of privacy [42]. Ponemon Institute conducted a study about security in CC in which 925 IT practitioners from US and Europe were surveyed in 2010 [54]. One of the results of this study was the rating from respondents who said they were confident that their organizations could achieve some security features. The security feature with the least confidence percentage was Secure sensitive or confidential information at rest (33%) which explains the great interest in studying data protection. The next security features with low confidence percentages were: Prevent or curtail external attacks, Secure vendor relationships before sharing information assets and Identify and authenticate users before granting access to information assets or IT infrastructure. This last one can be mapped to our identity and access management category, the second one in our results.

In contrast, relatively little attention (3.27% of the papers) has been paid to the governance in CC. Our finding also confirms the study performed by the Ponemon Institute, in which respondents identified governance as one of the least important technology to securing CC. In a follow-up study performed in 2012 [55] which goal was to determine what changes had occurred in the respondents answers since the first study [54], the results showed that there was no clear answer according to respondents about who should ensure the security of IaaS, so the authors suggested that it is important to improve governance by establishing responsibilities and policies that define the process of checking the CC security.

Trust is the next one with a low number of papers (13.60%), but unlike governance, trust is a key element to the real users of a CC system. The only experience paper extracted from this mapping study presents an analysis of user experience issues in CC [48]. The analysis was conducted by Nilay Oza et al. using an interview process. The 100% of the interviewees mentioned trust as one affector of their user experience. The confidence in a cloud service provider depends on factors such as the presence in social networks, friend referrals, position in search engines, language or cost of the service. These indicators are just a starting point. In fact the real service will build the confidence level of the provider. The confidence level depends on the security controls to protect the data and applications. Audits performed by third-parties may play a key role in achieving a minimum level of trust [8]. Our findings present 15 articles about audit, contributing with models, methods, surveys but just one of them is a tool. This
demonstrates the absence of audit reporting tools and, in general, CC specific tools, as Bryan Doerr also suggests [56].

From a practitioner's point of view, the results presented in this study provide an overview of existing security approaches in CC. This technology has not yet reached great levels of expansion, which provides new opportunities for innovation. The approaches and models proposed are in their beginnings and should be applied to real systems. Around 42% of contributions are proposals for new models and only 5.59% of them are evaluated. Working with existing CC security frameworks should be considered. The NIST, which has published the Guidelines on Security and Privacy in Public Cloud Computing [8], has also defined the NIST Cloud Computing Security Reference Architecture [57]. This architecture introduces a Risk Management Framework for applications or services in the cloud. The comparative criteria proposed by Rebollo et al. [14] could be adopted in order to analyze and compare CC related frameworks. This comparative framework includes the following characteristics: Policies and Processes Adaptation, Control and Audit, and Service Level Agreements.

From a researcher's point of view, multiple lines of investigation emerge as a result of the few years of experience in CC. Even the term 'cloud computing' itself is still vague, with no universal definition [58]. The main line of investigation focuses on studying more specific security mechanisms, such as audit, and on evaluating the solutions and proposing stable models and methodologies which are capable of tackling the real world problems in CC. 38.95% of the papers are solution proposals that have not been evaluated in real practical scenarios, which demonstrates that CC is a recent research field. There is a need for research on evaluation, and likewise experience papers and tools (only 7 papers propose a tool, 2%).

Alternative solutions are also required, such as ontology definitions, or software test suites, both of which are unexplored fields. Ontologies are the first step to perform an analysis of the domain. Having more articles proposing ontologies will enable knowledge sharing [29] and therefore facilitate the creation of missing universal terminology of CC. In this mapping study, just one paper proposing a testbed was found, the Open Cirrus Testbed [44], but there are some other important non-published testbeds: PlanetLab [59], EmuLab [60], IBM-Google NSF CluE [61], Amazon Cloud [62], Eucalyptus Public Cloud [63], Open Cloud Testbed (OCT) [64], VINI [65] or OneLab [66]. Our study shows the lack of articles about the use of testbeds and the lack of citations to these providers. A testbed in cloud computing is a network platform available to researchers and developers for experimentation, to develop, debug and evaluate their systems [60]. Since researchers have to validate and evaluate their solution proposals, testbeds can help them with that task. Nevertheless, notice that testbeds are emulators that are not entirely representative of the Internet [67].

4.1 Limitations of the Study

There is a set of threats to validity.

- Conclusion validity. Conclusion threats to validity in a mapping study are related to incorrect data extraction or missing studies [68]. One limitation of this systematic mapping study is the distortion of statistical analysis owing to the criteria used in the selection phase. The final results depend on the decisions made by the researcher who conducted the search process. This limitation was alleviated by defining the
inclusion and exclusion criteria in order to assemble the most extensive amount of papers conforming to the domain. In the steps of selecting and classifying, bias may affect the interpretation of the results. This limitation was alleviated by clearly describing every activity performed in these steps.

- Construct validity. Construct threats to validity in a mapping study are related to the identification of primary studies [69]. Security in CC is a wide topic and requires the consideration of more specific research questions, since the goal of systematic mapping studies is more oriented towards categorizing the papers selected. Different terms related to security (i.e. privacy, vulnerability or risk) were determined in order to find a wide range of articles. However, the results found could not be complete due, either to the manual research or the fact of missing other terms in the search string, which could has affected the final list of articles selected. To alleviate this threat, the PICO criteria was used and different terms were added to the search string. The search for primary studies was performed by using IEEE Digital Library, IEEE-Xplore, ACM Digital Library, Science Direct and Wiley InterScience. According to Zhang et al. [70], IEEE Xplore and ACM Digital Library are the main search portals in software engineering.

- Internal validity. Internal threats to validity in a mapping study are related to the extraction and data analysis [69]. A common view is that mapping studies are often conducted based on only the abstracts. However, we have detected that in the search and inclusion/exclusion phase, the extraction of relevant information was particularly difficult owing to the fact that some abstracts are often misleading and lack important information. Structured abstracts, which ensure that all important information is available such as background, research method and conclusions would help in the identification of studies [71]. This threat was alleviated by additionally allow more detailed study of papers for which it is not clear how they should be classified. The more parts of a paper one considers, the more effort is required and the validity of the results also increases [15]. Data extraction could also result in a misclassification, but this limitation was alleviated by creating a classification scheme on the basis of widely accepted guidelines [26].

- External validity. External threats to validity in a mapping study are related to the generality of the results [72]. The validity of the conclusions drawn in this paper concerns only this systematic mapping study. This threat is therefore not present in this context.

5. Conclusions

Various relevant conclusions have been extracted from this systematic mapping study, which provides a structure of the type of research reports, and results that have been published as regards security in CC. Different security issues are being investigated: data protection, access management, software isolation, availability, trust and governance. Most of the solution proposals are models, followed by methods. The proposed solutions are secure architectures for CC systems, communication protocols and encryption protocols. Although the first year of publication was 2009 and the number of publications would appear to be growing, 40.00% of the papers are proposals that do not contain any real evaluations and around 26.74% are surveys that contain no
proposals at all. The idea of CC as an emerging technology is supported by these results, thus creating a favorable area for research and evaluation, with special emphasis on security [73].

The total of 344 papers selected were found in a variety of 172 different publishing sources, signifying that this topic is the source of a huge amount of interest in different fields of computer science. Other disciplines such as health [74], business [75] [76], and education [77] [78], among others, are also important application areas in which a strong growth of publications is expected. The secure use of cloud services through mobile devices [79] and the security assurance mechanisms in green cloud computing environments [80] are also promising research areas.

Acknowledgements. This research is part of the projects PEGASO-PANGEA (TIN2009-13718-C02-02) financed by the Spanish Ministry of Science and Innovation (Spain), and GEODAS-REQ project (TIN2012-37493-C03-02) financed by both the Spanish Ministry of Economy and Competitiveness and European FEDER funds.

References


59. PlanetLab. Available at: http://www.planet-lab.org/

60. EmuLab. Available at: http://www.emulab.net/


62. Amazon web services. Available at: http://aws.amazon.com

63. Eucalyptus. Available at: http://www.eucalyptus.com/

64. Open Cloud Consortium. Available at: http://opencloudconsortium.org/

65. VINI. Available at: http://www.vini-veritas.net/

66. OneLab. Available at: http://www.onelab.eu/


68. Ampatzoglou, A., Charalampidou, S., Stamelos, I.: Research state of the art on GoF design


Belén Cruz Zapata received her Engineer's degree in Computer Science from the University of Murcia in Spain, specializing in software technologies and intelligent and knowledge technologies. She earned an M.Sc in Computer Science and is now working in her Ph.D on the Software Engineering Research Group from the University of Murcia. Her research is focused on mobile technologies in general and also applied to medicine. Belén is currently working as a mobile developer for Android and iOS in the San Francisco Bay Area at Groupon.
José Luis Fernández-Alemán is an Associate Professor at the University of Murcia (Spain), where he is a member of the Software Engineering Research Group. He received his B.Sc. (Hons.) degree in 1994 and his Ph.D. degree in 2002, both in Computer Science from the University of Murcia. He has published more than 20 JCR papers in the areas of requirements engineering and software engineering in the e-health, e-learning and mobile development domains and its application to the fields of computer science, medicine and nursing. Publications include articles in highly ranked international journals such as IEEE Computer, IEEE Software, REJ, IEEE Transactions on Software Engineering, JMI, JBI and NET.

Ambrosio Toval Álvarez is a full professor at the University of Murcia, in Spain. He holds a BS in Mathematics from the University Complutense of Madrid, and received a Ph.D. in Computer Science (cum laude) from the Technical University of Valencia. He is involved in a variety of applied research and development projects with industry and conducts research and technology transfer in the areas of requirements engineering processes and tools, privacy and security requirements and applications in the e-health, e-learning and mobile development domains. He has published in the same topics in international journals, such as IEEE Software, IST, REJ, Computer Standards & Interfaces, IET, IJIS, etc. Dr. Toval is currently the Head of the Software Engineering Research Group, at the University of Murcia.

Received: February 05, 2014; Accepted: October 05, 2014