Security in Legacy Systems Migration to the Cloud: A Systematic Mapping Study

Luis Márquez Alcañiz¹, David G. Rosado², Daniel Mellado³ and Eduardo Fernández-Medina²

¹National Competition Commission, Madrid, Spain
²GSyA Research Group, University of Castilla-La Mancha, Dept. of Information Systems and Technologies, Ciudad Real, Spain
³Spanish Tax Agency, Madrid, Spain

Abstract. While cloud computing emerges as a major trend in IT industry, early providers and adopters are paving the path with concerns and solutions. One of the most worrisome challenges that face the corporate clients of this new form of IT provision is how to maintain the security of their most important every day apps in the new environment, that is how to migrate securely their legacy systems that run on data centres fully controlled by the organization's IT department to a less clearly controlled infrastructure that is managed at least partly outside the scope of the clients premises and even completely off-shore. This paper presents a Systematic Mapping Study on the issue as the first step to analyze the different existing approaches in the literature about migration process to Cloud computing where taking into account the security aspects that have to be also moved to Cloud. We propose four research questions dealing with the existing strategies to migrate legacy, how they relate to common security issues as well as security issues specific to the cloud environment, and how the proposals are aligned with security standards.

1 Introduction

In early papers relating cloud computing (CC) we could find different definitions of what CC is, likely referencing one in [1-4]. But, probably, the most widespread today is the one given by Mell and Grance in [2]; CC is “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction”. The Gartner's special report by Smith in [5] says in its very opening statement: “cloud computing is maturing, but it continues to be the latest, most hyped concept in IT”. The Gartner’s report also shows how the general issue has just gone over the top of the hype cycle, with general CC security and private CC at the verge of the peak.

From the very beginning, IT executives have shown a top concern: security [6]. So it was in the first years after the concept emerged [7, 8] and so it is still today [9].
For many experts CC poses new risks to the computing environment since it is “at odds with traditional security models and controls” [10].

Other people, though, see in the CC paradigm a good chance to improve security of legacy software. For instance, Winkler states in [11] that the migration from legacy systems to the cloud “gives us hope that we can regain control […] from poorly integrated or after-thought security.”

According to the survey carried out by MeriTalk [12] among 166 Federal IT leaders “47% of IT applications are based on legacy technology in need of modernization” and the Federal spending to support legacy apps can be estimated around $35B.

Given that the security is mostly questioned by very same practitioners that will have to decide, the apparent absence of an acknowledged migration model for legacy systems with specific security concerns could deter the widespread adoption of the model. Thence, this seems to be a valuable opportunity for research. As a first step in the research process we decided to perform a systematic mapping study (SMS) of the existing CC migration models, with a focus on concrete references to either issue of security, privacy or system information assurance. Also, to avoid the bias of our research as much as possible we decided to follow a structured method as the ones introduced in [13] and [14]. Systematic literature reviews and systematic mapping studies in software engineering are meant to bring some of the better qualities and sound practices in other research areas like social sciences to the first steps of any research in the software engineering field. This paper is the result of this systematic mapping study.

The rest of the paper is structured in 3 more sections. The next section, section 2, defines with some detail the background, the research question and describes the SMS process that we have followed. Section 3 presents a mapping study of the selected pieces with links to the research question defined and shows the analysis results of the comparison performed. And, finally, we close in section 4 with our conclusions and agenda for future work.

## 2 Background and Method

To carry on our SMS we decided to follow the method presented in [13]. The goal of a SMS is to collect and evaluate as much research as possible related to a given question so that the result is as repeatable, auditable and unbiased as possible. We adapted slightly the method presented in [13] to our specific needs.

It might look that the process is linear in nature. Actually, the results are produced through some iterations of the main cycle. Each cycle allows us to refine the process until we get the appropriate results.

### 2.1 Background: Legacy Systems Migration

The problem of software with poor quality from the perspective of its maintainability and adaptability and how it can evolve along its lifetime is not new. But the problem changes along with new technologies. When a new technology emerges, the products
that were developed or deployed with the preceding technologies can become legacy. Legacy is an ambiguous adjective: for the business management it means a valuable asset, for the IT department it means risk.

There are many definitions for what a legacy system is, but probably the most acknowledged is the one shared in [15] and [16]: systems become legacy systems when they begin to resist modification and evolution. Some authors like Somerville point out that legacy systems are somehow socio-technical systems that include not only software and hardware, but business processes and people.

The frame for legacy evolution is presented in [17], where Bisbal et al. classify the evolutionary solutions when dealing with legacy from mild wrapping to redevelopment, being the final result of the evolutionary process a mix of different components each one with a particular solution. Migration happens to be in between the extremes and, actually, takes techniques from both of them. In [16], Secord et al also give an approach for a risk driven modernization as part of a spiral development process and delivers a model that depicts a horseshoe and that is in use at the Carnegie Mellon SEL. This horse model is applicable also to the cloud. There is little work on migration from legacy to service oriented architectures (SOA). In [18], Heckel et al justify this little work in the fact that the SOA reengineering area is quite new. But, when focused on the more specific world of CC, the alternatives are even less.

2.2 Research Questions

Before carrying out the actual mapping study, we performed a preliminary analysis to search for existing cloud migration methods for legacy systems that take into account security in their proposals. This analysis showed a lack of publications dealing with these issues. Some proposals for cloud migration migrations have been made in [19-21], but few of them deal with “security” or “privacy” or any other related. Furthermore, current migration proposals seldom deal with CC security particularities.

A critical step in a SMS process is to specify the research questions. In our case, the research questions will focus on identifying existing proposals for migrate legacy systems to different CC schemes that deal with the security explicitly and thoroughly. Thence, the scope defined is twofold.

The review is aimed at comprehensive proposals, that is, that deal with all the security aspects that may arise during the migration, from adapting the security requirements already present in the target legacy system to improving the security properties of the target.

The approaches must consider the particular characteristics of CC. Only those studies that deal specifically with the CC model have been taken into account. To attain the scope we have devised four research questions given in table 1.

<table>
<thead>
<tr>
<th>ID</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>What types of CC migration strategies are proposed for legacy systems?</td>
</tr>
<tr>
<td>RQ2</td>
<td>How CC migration strategies deal with the security issues raised from legacy?</td>
</tr>
<tr>
<td>RQ3</td>
<td>How CC migration processes deal with the security issues raised from the CC model?</td>
</tr>
<tr>
<td>RQ4</td>
<td>How CC migration processes are aligned with security standards?</td>
</tr>
</tbody>
</table>
2.3 Review Protocol

From [12] we know that we had to define in advance a detailed review protocol. This is both to allow us to plan the activities and to avoid biased results in the SMS. In this section we present the protocol that we defined and followed in the SMS.

Once the research questions were settled, a set of search terms was extracted from them. These terms, aka keywords, were used in the review to identify all the relevant approaches that were related to the research questions and to attempt to answer them. A precise definition of the keywords is vital if comprehensive results are to be retrieved without neglecting important approaches.

Considering that the proposed research questions include characteristics from different research areas, it was necessary to define the keywords in such a way that none of them was excluded. We also joined some synonyms of the keywords to prevent a proposal being ignored because of alternate vocabulary use. In the end, the proposed keywords that were used combined in different ways in our SMS were the following: Migration/reengineering; Strategy/process; Combined in expressions such as "migration strategy" and "migration process"; Legacy system/software; Cloud Computing; and Security/Privacy Standard/ISO/COBIT.

We tried to keep this set of keywords a bit fuzzy and generic in an attempt to compile all the proposals related to include security requirements in the migration of legacy systems to the CC. Thence; it was expected that the search for these terms should return some additional results that were not strictly related to the review’s goal (for instance, “migration process” may. Selection filters were subsequently applied in order to frame the relevant studies.

Another crucial point when doing a SMS is the selection of the sources. We used search engines from the following sites: Google Scholar, Elsevier, IEEE Xplore, ACM Digital Library and Science Direct.

Because of language limitations, only studies written in English were to be considered in this review. And, because we are dealing with a very recent and rapidly changing research area we also decided to restrict the time scope to those items published from 2008, year in which the term “cloud computing” became widely used within the academia.

From [12] we also know that it's advised to include in the SMS protocol a preliminary search in seek of already published reviews on the subject. This was done by querying for the more general “cloud computing” filtered with the term “literary review” and “mapping study”. In our case, and probably because of the lack of maturity of the research topic, we promptly realised that our available resources did not contain any publications dealing with the proposed research questions. The SMS process consequently should continue with the search for primary studies which contained the aforementioned keywords.

A combination of the previous search terms was filled to the forms of the aforementioned engines. So we got a first approximation to our research question that allowed us to get a list of probably relevant primary studies. Complementary to the results obtained from the search engines, we got some material from experts in the field and other colleagues who had been working in related fields to whom we consulted.

Then, we had to filter the results to extract those pieces of literature that better
satisfied the SMS conditions. We did this through a first scan of the title of the article and the keywords, and later a second pass by browsing the abstracts. These proposals were further narrowed through the definition of a set of inclusion and exclusion criteria, which had to be objective in order to reduce the bias of the results and to ease the repeatability of the SMS process.

One of the objectives pinpointed in this SMS was that of compiling a comprehensive catalogue of migration frameworks from legacy applications to the cloud. Another inclusion criterion was to ensure that we selected proposals dealt with SOA and/or cloud migration with some mention to security aspects. An objective criterion was established by relying on an accepted and widespread security standard: the ISO/IEC 27000 standard family.

This group of criteria was used to narrow down the initiatives obtained by the first search. In most cases, it was sufficient to contrast the title and abstract or executive summary with the proposed criteria to decide whether to include or exclude the proposal. Nevertheless, when in doubt it was, in some cases, necessary to analyze the whole text in order to make the appropriate decision.

The quality assessment of the selection of publications performed was conducted in the aforementioned multistage process. As previously stated, the SMS was executed iteratively, signifying that the results were analyzed after each cycle to confirm whether we were heading in the right direction. The final list of papers obtained is given in Table 2.

<table>
<thead>
<tr>
<th>ID</th>
<th>Paper</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>[RR5]</td>
<td>Cloudward Bound: Planning for Beneficial Migration of Enterprise Applications to the Cloud</td>
<td>[23]</td>
</tr>
<tr>
<td>[RR6]</td>
<td>REMICS-REuse and Migration of Legacy Applications to Interoperable Cloud Services</td>
<td>[24]</td>
</tr>
<tr>
<td>[RR8]</td>
<td>Service Migration in a Cloud Architecture</td>
<td>[26]</td>
</tr>
<tr>
<td>[RR9]</td>
<td>Dynamic Service and Data Migration in the Clouds</td>
<td>[27]</td>
</tr>
<tr>
<td>[RR10]</td>
<td>Automatic conformance checking for migrating software systems to cloud infrastructures and platforms</td>
<td>[28]</td>
</tr>
</tbody>
</table>

### 3 Data Extraction and Comparison

This section contains the summary of the information extracted from each of the pieces of literature found by our SMS process. A brief description of each is provided
along with the most relevant aspects related to the comparative criteria previously defined.

3.1 An Analysis of Security and Privacy Issues in Smart Grid Software Architectures on Clouds [RR1]

This paper presents some security issues arisen when dealing with a cloud implementation for managing a grid of devices connected to meters (smart grid) that includes some legacy applications and hardware in their edge devices.

The paper is mainly a taxonomical analysis, and does not propose a detailed method or procedure for integrating or migrating legacy apps and or data on the edge, though it gives a possible approach that consist in virtualization. Otherwise the paper only points out the compatibility issue with both existing and new apps. Mostly, the proposal for dealing with legacy applications and or hardware is the coexistence of legacy platforms with the parts of the system that has been migrated to the cloud, as well as virtualization of legacy applications into environments “with identical configuration as the legacy” and running all the software stack on cloud infrastructure.

This paper also points out the necessity of negotiating and monitoring the enforcement of SLA with some security standards related to web services like WS-Policy, WS-Security and XACML.

3.2 Digital Stakes at Risk! Modernizing Legacy Systems [RR2]

The report presents quite a lot of information about studies done on legacy systems in the public sector in USA, their characterization and drivers to its modernization/migration. It also presents seven surveys on modernization/migration of legacy systems. The 4th survey is specific on security and enterprise risks.

The report does not presents in detail any specific method or procedure for migrate and or integrate legacy applications, but it points out to the main different approaches that the interviewees have applied: wrapping, automated migration, data conversion, extension, rehosting/replatforming, reengineering and replacing with commercial-of-the-self, re-architecting/ renovation, SOA integration, Enterprise Application Integration (EAI), virtualization/emulation (VM) and others.

Some of the strategies presented cannot be regarded as migrations dealing with software engineering (e.g. virtualization), but give us an idea about the trends and options when modernizing legacy systems and the success rate that can be expected with each strategy.

Though the report deals specifically with security and risks, there is no link between the methods pointed out and the risks associated to each method. Actually, most of the risk that are mentioned in the report are not related to information security, but to enterprise risks.
3.3 The CloudMIG Approach: Model-based Migration of Software Systems to Cloud-optimized Applications [RR3]

This presents a specific model for migrating legacy systems into the cloud called CloudMIG. The model presents six activities that aim to transform the legacy system into a cloud ready one; namely: extraction, selection, generation, adaptation, evaluation and transformation.

The work does not deal with security issues, though the 3rd activity (Generation) provides a model with the target architecture violations of the cloud environment constraints (CEC). But they are generic constraints, the paper is not specific about security constraints neither of the legacy nor the target, and though we assume that security constraints could be modelled as CEC there is no clue about how to do it.

It does not relate the model to any security standard nor any process aimed to ensure the security requirements of the target or the artefacts produced, though the Knowledge Discovery Metamodel (KDM) is proposed in the first activity as a suitable model for building the original legacy software architecture.

3.4 Migrating Legacy Applications to the Service Cloud [RR4]

This paper presents a process for migrating legacy software into software-as-a-service architecture in seven steps. It relies on model driven architecture (MDA) transformations and the Software Engineering Institute horse model [16]. The seven steps are the following: architectural representation of the legacy, redesign of the architecture, MDA transformation, web service generation, web service based invocation of legacy functionalities, selection of the CC platform and web service deployment. They apply the model to a case of study for migrating a legacy application in the area of oil risk analysis.

The paper does not deal with security issues but in the last step (migration to the cloud). And there only mention security in a general non specific way along with scalability and networking. Nor it seems to make detailed questioning about security constraints of the legacy. The paper does not mention any security standard nor any activity nor task nor process related to integration or review of the security constraints of the target.

3.5 Cloudward Bound: Planning for Beneficial Migration of Enterprise Applications to the Cloud [RR5]

This paper presents a model to deploy enterprise applications in a hybrid cloud environment; where applications are partly hosted within the organization premises. It gives an in detail model to evaluate which components of a legacy application can be migrated to the cloud, and which components must stay. The authors illustrate the model with a real case use at a large university involving the migration of their Enterprise Resource Planning (ERP) system.

The paper deals specifically with the migration of access control lists and their definition in the firewalls of the organization. Nevertheless, it does not deal with
security requirements of the application, but the security requirements of the communications (firewall contexts and ACL rules).

The model is not a software engineering one, but a systems oriented one. Thence, the requirements of security or the migration from the application perspective is not addressed. The paper does not mention any security standard.

3.6 REMICS-reuse and Migration of Legacy Applications to Interoperable Cloud Services [RR6]

This report presents a tool-supported MDA methodology for migrating legacy applications to cloud. The model presents four activities: recover, migrate, model driven interoperability, validate control and supervise and forward MDA through cloud.

The paper does not deal specifically with security, nor risks, nor any standard of security. They refer to the REMICS project web page for further publications and the state of the art, but we have not been able to trace any further progress.

3.7 Decision Support Tools for Cloud Migration in the Enterprise [RR7]

This paper presents in further detail the decision support system mentioned in the previous paper. It proposes a model for risks assessment when migrating legacy applications to the cloud and tries to identify a template for risk evaluation and mitigation approaches.

The paper does not propose a detailed legacy application migration process, nor deals with security constraints of the legacy apps, nor mention any standard of security. Nevertheless, they give five risks associated with security according to a survey that the authors have made within 50 academic papers.

3.8 Service Migration in a Cloud Architecture [RR8]

This paper present several security and integration issues related to the migration of existing applications within three support areas: acquisition, implementation and security. In section 4.3 presents several concerns on security that organizations that plan to migrate must take into account.

It states that, “presently, cloud computing cannot support users who cannot switch from legacy applications because equivalent cloud applications do not exist”.

In the security section it gives questions that the practitioner should ask himself or herself like the fact that “data must often be retained locally to satisfy regulatory requirements” or regulatory and legal issues related to the security of data and/or its availability. It does not present a technical process for migration nor align any activity with the security process or with any security standard.

3.9 Dynamic Service and Data Migration in the Clouds [RR9]

This paper presents a framework to ease the migration of services to the cloud and
gives a decision model to select the services that can be moved. It does so from the point of view of the supporting SOA services, where the general computing platform (GCP) is also regarded as a service that uses VM technologies.

The paper states that a GCP that uses VM solves some of the security problems but that there still remain some that must be specifically address through an authentication framework and the use of certificates and certificate authorities (CA). All the security model proposed relies on this issue. It does not align the use of the framework with any security requirement specific to the legacy application migrated, nor gives any process to integrate security into the target. It does not link security with any security standard or model.

### 3.10 An Extensible Architecture for Detecting Violations of a Cloud Environment's Constraints during Legacy Software System Migration [RR10]

This paper presents a detailed description of CloudMIG and explains more thoroughly than the one in [RR3] the mechanisms to deal with the target architecture violations of the cloud environment constraints. It also gives a metamodel that should help to semiautomatically migrate legacy systems to the cloud.

The process deals with risks related to the migration process, namely, the runtime constraints, but not specifically to the security constraints of the legacy or the target platform. It does not relate the model to any security standard nor any process aimed to ensure the security requirements of the target or the artefacts produced.

### 3.11 Literature Comparison

In the paper reviewed we can largely find two main groups of research interest. One of the groups deals with the research question about specific strategies to migrate to cloud computing (RQ1) and the other group deals with issues on security general (RQ2), security specific to cloud (RQ3) and linked to standards (RQ4). Table 3 shows the mapping of the papers reviewed and how they relate to the research questions.

As we can see, most of the proposals only deal with the general strategies, and the ones that deal with the security issues don't propose any strategy; perhaps with the exception of RR5 which is focused only on hybrid CC proposals. The only one that names (though does not develop) some security standards does not give any strategy to migrate legacy, but deals with security from a descriptive perspective.

Although there are seven publications that mention security aspects to take into account in the migration (RQ2), none of them presents an approach indicating which are the most important issues to consider, how to perform the migration of these aspects of security, what set of security requirements have to consider, which are the most appropriate mechanisms used to implement certain security services for the Cloud, what security standards are more appropriate taking into account different standards for areas such as healthcare (e.g., Health Insurance Portability and Accountability Act (HIPAA)), finance (e.g., Payment Card Industry Data Security...
Standard (PCI DSS)), security (e.g., ISO 27001, ITIL, COBIT), and audit (e.g., Standards for Attestation Engagements (SSAE) No. 16), and so on. That is, a migration process to guide and indicate to us the steps, tasks, recommendations, mechanisms, standards, and decisions to follow with the main objective of migrating security aspects and services to the Cloud.

Table 3. Literature mapped to research questions

<table>
<thead>
<tr>
<th>ID</th>
<th>RQ1</th>
<th>RQ2</th>
<th>RQ3</th>
<th>RQ4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR1</td>
<td>Smart Grids</td>
<td>Smart Grids</td>
<td>Only names (WS oriented)</td>
<td></td>
</tr>
<tr>
<td>RR2</td>
<td>Non CC specific</td>
<td>Partly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR4</td>
<td>SOA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR5</td>
<td>Hybrid CC</td>
<td>ACL</td>
<td>Only hybrid CC</td>
<td></td>
</tr>
<tr>
<td>RR6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR7</td>
<td>Only DSS</td>
<td>List risk</td>
<td>List risk</td>
<td></td>
</tr>
<tr>
<td>RR8</td>
<td>No</td>
<td>Only questions</td>
<td>Only questions</td>
<td></td>
</tr>
<tr>
<td>RR9</td>
<td>GCP via VM</td>
<td>VM &amp; CA</td>
<td>VM &amp; CA</td>
<td></td>
</tr>
<tr>
<td>RR10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RQ1: What types of CC migration strategies are proposed for legacy systems?
RQ2: How CC migration strategies deal with the security issues raised from legacy?
RQ3: How CC migration processes deal with the security issues raised from the CC model?
RQ4: How CC migration processes are aligned with security standards?


4 Conclusions and Further Research

In the previous sections we have seen, the results of following a structured process to try to answer a set of research questions relating security in the migration of legacy systems to the cloud computing paradigm.

As we said in section 1 and 2, we decided to follow a systematic approach to avoid biases and to make the mapping study as repeatable as possible. That is the purpose of using a process as the ones proposed by Petersen et al in [13] and Kitchenham in [14].

From our systematic approach we can state that there is a lack of research in the field and that this lack of research points out to a gap that should be filled if cloud computing services are meant to be incorporated into the large corporate businesses. As we have justified, to replace the legacy cloud proposals must address the issue of security in a structured way and, up to date, there is no overall approximation strategy to the cloud migration that deals with security in a straightforward manner.

Given that there are no initiatives where a migration process is proposed for security aspects, something that, as we have seen, managers regard as one of the most important issues for a core application that has to be migrated to the Cloud, we feel...
that there is an urgent need to provide methodologies, techniques and tools not only for giving access to the cloud for the data and services which are locked in these closed legacy parts of the core information system, while maintaining the security standards, but also to provide a strategy that ease the certification of this security and process.

Our next step will be identifying existing migration processes that, though focused on other technologies, would address the security of legacy integrated into the process so that we could develop from there a common strategy and adapt it to the cloud specificities. Other point of interest of our research will be making up a catalogue of security standards related to web services and IT in general so that we take them into account when developing our strategy. And a final issue that will concern us is how to adapt the certification processes for software engineering with the details of the cloud computing delivering model.

Acknowledgements

This research is part of the following projects: GEODAS (TIN2012-37493-C03-01) and SIGMA-CC (TIN2012-36904) financed by the “Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional FEDER” (Spain).

References