Information and Software Technology 146 (2022) 106874

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

A Delphi study to recognize and assess systems of systems vulnerabilities

Miguel A. Olivero [a,*], Antonia Bertolino [b], Francisco José Dominguez-Mayo [a], Ilaria Matteucci [c], María José Escalona [a]

[a] University of Seville, Seville, Spain
[b] Istituto di Scienza e Tecnologie dell'Informazione, Pisa, Italy
[c] Istituto di Informatica e Telematica in Consiglio Nazionalle delle Ricerche, Pisa, Italy

ARTICLE INFO

Keywords:
Delphi
Expert judgment
Security
Systems of systems

ABSTRACT

Context: System of Systems (SoS) is an emerging paradigm by which independent systems collaborate by sharing resources and processes to achieve objectives that they could not achieve on their own. In this context, a number of emergent behaviors may arise that can undermine the security of the constituent systems.
Objective: We apply the Delphi method with the aims to improve our understanding of SoS security and related problems, and to investigate their possible causes and remedies.
Method: Experts on SoS expressed their opinions and reached consensus in a series of rounds by following a structured questionnaire.
Results: The results show that the experts found more consensus in disagreement than in agreement about some SoS characteristics, and on how SoS vulnerabilities could be identified and prevented.
Conclusions: From this study we learn that more work is needed to reach a shared understanding of SoS vulnerabilities, and we leverage expert feedback to outline some future research directions.

1. Introduction

A System of Systems, abbreviated as SoS, refers to a set of constituent systems that cooperate by sharing resources and processes to achieve a common goal, or mission, that they could not achieve individually. Although the origins of the discipline could be dated back to the 1950's [1], it is Maier's work in 1998 [2] that is commonly cited as the SoS seminal work. Maier introduces a SoS as an assemblage of components that possess operational and managerial independence and concludes his description of SoS characteristics by predicting that SoSs "*will exist more widely in the future […] with the ubiquity of smart systems operated and managed independently*".

Although it took several years, his prediction eventually came true, and today research on SoS has gained momentum. As evidence of the growing relevance of the topic, the curve in Fig. 1 depicts the yearly distribution of the over one thousand citations reported in Scopus for the cited Maier's work: we clearly see the citations reach a peak in the last five years. Furthermore, a tertiary study [3] of the SoS literature has been completed in 2020, in which Cadavid and coauthors collect and review a set of 19 secondary studies published on the topic since 2013, which is quite remarkable. However, what emerges from this recent review of the literature in [3] is that, in contrast to some specific areas such as SoS architecture, modeling, or design, which have received greater attention, there are other important aspects of SoS that have been disregarded (e.g., security, flexibility, evolution). In another recent study [4], Teixeira and coauthors have focused on quality attributes for a specific class of SoS oriented to business processes they call "System-s-of-Information Systems". The authors conclude that relevant SoS quality attributes (e.g., reliability, security, and testability) are scarcely covered or even not mentioned.

The inherent composite nature of a SoS produces quite complex scenarios, such as examining the quality attributes of which is a non-trivial task. In particular, the non-composability of attributes such as security or reliability individually exposed by the constituent systems hinders their generalization over the resulting SoS. Actually, despite the fact that the constituent systems were secure when working on their own, the joint work and shared resources might produce new vulnerabilities that affect various constituent systems to arise. Furthermore, the concepts of 'secure' or 'security' are subjective, as it is the importance of the constituent systems for certain assets. In fact, something that could

* Corresponding author.
E-mail addresses: molivero@us.es (M.A. Olivero), antonia.bertolino@isti.cnr.it (A. Bertolino), fjdominguez@us.es (F.J. Dominguez-Mayo), ilaria.matteucci@iit.cnr.it (I. Matteucci), mjescalona@us.es (M.J. Escalona).

https://doi.org/10.1016/j.infsof.2022.106874
Received 21 July 2021; Received in revised form 4 February 2022; Accepted 10 February 2022
Available online 12 February 2022

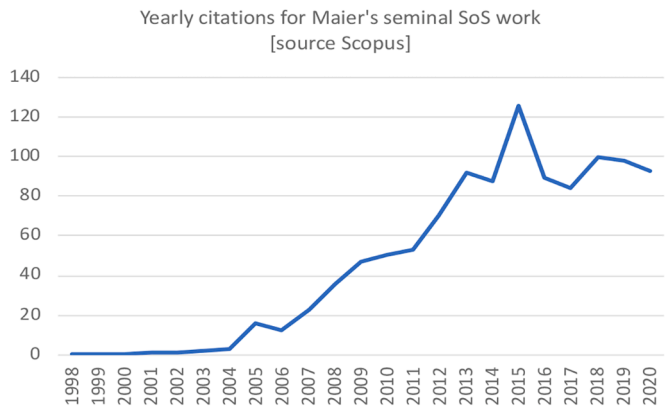

Fig. 1. Citations to Maier's work per year.

be critical to secure for a constituent system could be redundant for another.

The evolving and dynamic behavior of SoS makes it difficult to recognize the new vulnerabilities that emerge in a SoS. This is because of the dynamic cooperation among the constituent systems, which produces emergent behaviors. The vulnerabilities that might be introduced along with these unexpected behaviors cannot be easily addressed or detected in advance. As an example, in [6] we present how, by leveraging unforeseen matches among personal data belonging to different own digital identities, new security risks emerge for the digital persona (which is an instance of a SoS) that did not exist for each individual identity. Considering this, in previous studies we have investigated approaches to assess SoS security [5].

In our own recent systematic literature study on SoS security [7] we could ascertain that while several works exist that address the engineering of security in SoS, vulnerabilities emerging at global SoS level are not well understood yet and are not given adequate consideration. Therefore, in this work we present a study aiming at identifying the relevance of the issue: precisely, we use the Delphi method [8] to survey a group of experts to determine the potential existence of security issues due to the composition of shared resources on SoSs, their nature, and extent. Delphi studies allows for the validation of hypotheses in scenarios not yet mature, by leveraging experts' judgment.

The results of this Delphi study allowed us to identify the interests of the academia and the industry regarding SoS security, discover the main gaps and challenges to be addressed, and draw the requirements for a possible solution. The study has also been useful to determine the relevance of this research line for both academia and industry.

The remaining of this paper is organized as follows: Section 2 introduces the background and summarizes some related work. Section 3 presents the Delphi method. Section 4 describes our execution of the Delphi study. This section has been organized in three subsections corresponding to the three stages: A) planning, B) conducting and reporting, and C) results. Section 5 compares the current state of the art with the results obtained. Finally, Section 6 draws conclusions and suggests some research lines for future work.

2. Background and related work

In 2005, Maier highlighted the existing research challenges to be addressed in the SoS context [9] that mostly differed from the handling of more conventional systems. Among others, he listed the need to account for social and technical aspects when designing SoS; the problem of defining adequate architectures in view of evolution; the lack of adequate methods for describing and analyzing upper layers of the SoS (e.g., concerning security, or reliability properties); the complexity of decision-making concerning management of resources and missions.

Considering the future work directions pointed out by Maier, many authors have produced scientific contributions according to these challenges. According to a 2013's systematic review [10], SoS architecture has been the most researched challenge so far. Regarding the studies analyzed, only 7.2% (14/194) focused on SoS security aspects. The authors emphasized that the maturation rhythm of the SoS area of research was slower than that of others.

During 2015, three relevant literature reviews were published. One of them is a Systematic Mapping Study (SMS) which structured the scientific developments in SoS [11]. The authors concluded that until 2015 the most researched areas were architecture, modeling, and simulation. They also highlighted the immaturity of this area, the lack of citations of existing works, and the predominant participation of US researchers in this topic. This SMS studied nearly 3000 works, but the list of selected studies is not available.

Another Systematic Literature Review (SLR) conducted in 2015 analyzed SoS according to their architecture [12]. Despite the fact that SoS architecture had been the most researched topic, the authors concluded that there was a lack of consensus on how SoS architectures are described. In addition to that, they found that security was barely mentioned.

The third review conducted in the same year examined the quality attributes of SoS [13]. The authors highlighted that the three most relevant quality attributes were security, performance, and interoperability, with 14 selected primary studies referring to each of them. This work also highlighted the increasing difficulty in ensuring security in SoS given the dependencies, trade-offs, and relationships that existed between the different quality attributes.

A year later, another systematic review analyzed the software engineering methods that can assist in the integration of constituent systems [14]. The authors noticed that most studies were carried out to tackle specific issues, making it difficult to generalize the proposed approaches to broader SoS contexts.

In 2018, Daneva and Lazarov [15] present the results of a SLR focusing on the SoS requirements of smart cities. Their results showed that among the 32 selected studies, architecture requirements were still the most discussed topic and little was mentioned about security or privacy challenges. Similarly, in an SMS focusing on a specific class of SoS (i.e., Systems of Information Systems, SoIS) [4] the authors found out that only 2 of the 25 selected primary studies included security as a quality attribute in their research. Besides, the authors concluded that there is no specific modeling language that supports all the particularities of a SoIS, including the quality attributes.

According to the overviewed literature reviews, SoS security continues to be an area that is still immature as relatively few scientific contributions have been found. In a recent study [6] the security of Virtual SoSs are examined, but no clear guidelines or approaches were found that define how to identify if shared resources among the constituent systems were creating the vulnerabilities that affect the whole composition.

Experts' judgements techniques are often used for planning purposes as well as a support for taking decisions. They involve a set of experts on a topic to provide their opinion or criterion on their area of expertise. The purpose of using an experts' judgment technique in this study is to improve our understandings of the SoSs concept and more specifically to determine the interest and awareness of the academia and the industry regarding security of SoS. In particular, we focus on issues related to security vulnerabilities seen as emergent behaviors. Additionally, this survey would help identifying whether the slow pace of maturation in SoS Security is motivated by the lack of interest of researchers or its novelty.

According to the experts' judgment technique purpose and the resources available, there are different alternatives that can be used. Some of the best-known techniques are: (i) *Brainstorming* [16,17], (ii) *Delphi Method* [18], (iii) *Didactic Interaction* and (iv) *Nominal Group Technique* [19].

Considering our goal, we require to use an alternative that

guarantees the experts' freedom to clearly describe and communicate their knowledge and opinion. Such alternative should be compatible with a limited number of participants because of the novelty of the SoS area of research.

Delphi is an experts' judgment method on which a group of well-recognized experts in a certain field express their opinions in a series of rounds by following a structured questionnaire. Delphi method is not designed to be conducted face to face, nor to be run synchronously. Also, Delphi method preserves the anonymity of the participants, which guarantees complete freedom of expression, and the questionnaire design is not restricted just to binary responses.

A recent study has surveyed the use of the Delphi method in information systems area [20]. That survey highlights that the feedback received from the experts contributes to evaluate the studies research questions. Therefore, we found the Delphi technique as the experts' judgment technique that fits the best to conduct this survey.

The main purpose of a Delphi questionnaire [21] is to reach consensus in the answers among the expert panel participants. The degree of consensus on a survey round can be measured by means of several approaches. A literature review conducted by Heiko [22] summarizes some of them. Two main methods are identified from this survey: *(i) Descriptive statistics.* This method uses a subjective analysis of the results to detect the level of consensus among the participants. This analysis is usually done by examining the mean, median and standard deviation of experts' responses. *(ii) Inferential statistics.* This method covers statistical methods used to assess the consensus using a defined scale. Among these methods there are: *Cronbach's Alpha* [23,24], *Cohen's Kappa* [25] or *Kendall's W* [26], or *Simple Correspondence Analysis* [27] among others.

Also, beyond the quantitative results, a descriptive analysis of the justifications provided by the experts help in understanding their perspectives. This analysis enriches the results of the survey as it allows to detect the degree of interest of the topic and the areas of improvements.

The Delphi method has been previously used as a reference method to assist in a common knowledge understanding in other areas as in [28]. This method has also been widely used in fields as social or medical sciences and in the software development field [29–31].

This method has been also applied in the Computer Science area, for example to validate the use of methodological proposals of frameworks [32], or to characterize the reversibility of Cloud computing decisions [33].

We find that the Delphi method well supports our research because it allows to assess the opinion of the panel and to understand the experts' reasoning and disagreements. Using a Delphi survey, we can determine if the experts agree that SoS emergent behaviors can be a source of security vulnerabilities, and what could be a potential solution to avoid the identified vulnerabilities being exploited.

3. Research method

The Delphi method is usually structured in subsequent stages that orchestrate the process of defining the survey, conducting it, and processing the results. Some studies have defined five main activities: Selecting the subject, Selecting experts, First round, Further rounds and Conclusions [34].

In this study we have organized these activities into three main stages (*Planning, Conducting and Reporting, Results*) as summarized in Fig. 2. And we have also made explicit two additional activities *Statistical processing* and *Generating statistical data* on which the results for each round are calculated. This structure is closer to researchers' workflow as it is similar to Systematic Research' stages common structure [35]. The first two stages are, in turn, structured into three steps each, as we describe in the following.

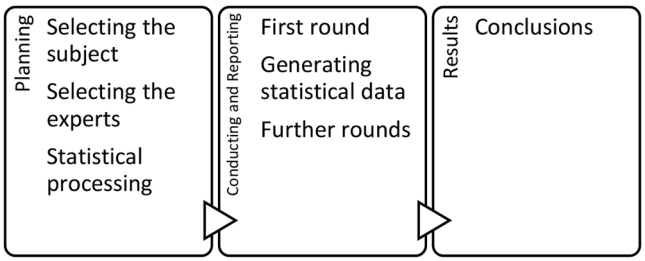

Fig. 2. Delphi process.

3.1. Stage 1: planning

Selecting the subject. This activity focuses on choosing the topic of the survey to which the experts will participate. A proper definition of the subject impacts on the quality of the results and the requirements to select the experts. Indeed, this activity is the most important one as the clearer the topic is identified and the more specific each question is, the more relevant the information that can be extracted from the survey. The outcomes of this activity are the research questions that define the purpose of the questionnaire, and the questionnaire itself that helps in gathering the required information to give an answer to these research questions.

Selecting the experts. This activity focuses on choosing the experts that will participate in the survey. A good selection improves the quality of the results. There are not clear guidelines for selecting the participants; their knowledge on the subject being studied and their willingness on participating in the survey are the key factors. The authors of [36] state that a number from 15 to 20 participants is an acceptable number to balance results and effort.

Statistical processing. During this activity the statistical process for analyzing the experts' responses is defined. Defining the statistical process before launching the survey prevents authors' bias to determine the results. Stopping conditions are also defined at this stage to determine when it is no longer necessary to launch additional rounds. Such stopping conditions can be reached by consensus, by a stability level, or when enough rounds have been conducted, among others.

3.2. Stage 2: conducting and reporting

First round. The first round introduces the questionnaire to the participants. There are two approaches to conduct this stage, either by using a well-structured questionnaire or by using a preliminary version and this first round as a test round to validate the questionnaire. In any case, the experts respond to this questionnaire and provide their opinion to the Delphi organizers, who analyze all the responses before launching further rounds. We chose the first alternative.

*Generating statistical data***.** This activity is conducted by the Delphi organizers. In this activity, the answers of the experts are analyzed and interpreted by means of previously defined statistical processing (e.g. descriptive analysis, homogeneity, concordance). The organizers shall produce an anonymous summary of the experts' panel opinion as a round outcome. This summary is used to determine if Delphi method has reached a stop condition, or it should be sent to the experts before the next round if needed.

Further rounds. When other rounds are launched, the Delphi organizers send to the experts a summary of the previous round including an assessment of the values when possible. Such assessment includes statistical data with the participants comments for each question in the survey. Based on this, the experts analyze the general results on the panel, compare it with their responses and they can modify their previous answer or decide to confirm their previous argument. Experience with Delphi surveys shows that three iterations are usually enough to gather the information about the panel opinion [32].

3.3. Stage 3: results

Conclusions. Once a stopping condition (one among those defined in *Stage1, Planning, Activity 3, Statistical Processing*) is met, a report is produced containing an analysis of the gathered information.

4. SoS security study

In this section we report about the conduction of the three stages of our survey.

4.1. Stage 1: planning

As said, Stage 1 establishes the purpose of the questionnaire and describes how the study has been planned.

1. Selecting the subject

Both the subject definition and the questionnaire design impact the conduction of the survey. In fact, the Delphi questionnaire begins with a well-defined purpose that impacts on the questionnaire design [34]. Despite the questionnaire has been defined before launching the survey, this questionnaire suffered small modifications along the rounds according to experts' comments.

This survey aims to validate a research hypothesis that arises from our previous studies. The first one refers to an SMS [7] that looks forward to identifying the state of the art of security in the SoS context. The second one [37] is a study that identified which are the main security challenges a SoS faces according to its architecture.

Considering the insights gained in these preceding works, the research hypothesis that leads this survey is "Emergent behaviors of the SoS generate security vulnerabilities". In other words, "SoS emergent behaviors produce unexpected flaws that could be exploited by attackers.". When doing joint work, systems accountants may have a different concept of what "*security*" is, what do they need to protect and what they should protect against. Such heterogeneity might be used by an attacker to scale privileges from one system to another up to gaining access to the whole SoS in a sort of cascade attack.

Three research questions have been identified to help validating such hypothesis and achieving this goal:

RQ 1. *Do experts interpret SoS in the same way?* This question helps in identifying if experts understand the SoS in the same way according to SoS main characteristics defined in the literature. This RQ allows to determine if the cases are detailing a SoS, and what experts consider as a SoS.

RQ 2. *Do experts agree on what are the causes of the vulnerabilities?* This question focuses on the security aspects, and how undesired emergent behaviors could impact on the security. It helps in identifying if experts do agree on a same opinion when determining which SoS characteristics are causing the exploited vulnerabilities, which ones are affected by the consequences, and how these vulnerabilities could be prevented. This is the main RQ for this study. From this RQ we can identify if experts do have the same understanding of SoS, Security and Good Practices. This RQ may highlight if Security on SoS is a context of interest and worth of research.

RQ 3. *What do experts identify as a potential solution?* The third research question aims to identify how a solution should be designed so to help in reducing the impact of vulnerabilities that arise as emergent behaviors. This RQ is useful to help at identifying future work and challenges to be addressed.

Previous studies have postulated towards modeling a SoS to examine its vulnerabilities and potential countermeasures by means of test cases [5]. However, not having found similar works, doubt arose whether the research proposal is quite novel or not of interest to the SoS community. These three questions would help providing the general opinion of experts and determine whether the topic is relevant enough to invest further effort researching it, as well as how experts imagine a potential solution, or to identify and understand the reasons of such a lack of scientific contributions regarding in this area. The comments received from the experts would allow us to define the future work according to the suggestions and clarifications.

Each one of the research questions is directly related to one dimension to be analyzed. Each dimension is independent from the others, meaning that experts may agree or disagree for each one of these dimensions individually. For each dimension we develop one or more survey questions as described below. In total there are six different questions (see Tables 1–6) These questions are composed of a *Statement*, which is the question itself, a set of *Alternatives,* which can be either a set of different alternatives which Experts can choose or a Liker scale, and a *Motivation.*

The survey aims to discover what the experts think regarding the *characteristics of described SoS*, the *causes of the security vulnerability* and the *nature of a solution*.

(i) Characteristic of described SoS. This dimension helps in identifying how the scenarios are understood by the experts. We refer to the SoS characteristics identified in other literature studies [2]. The purpose is to respond to RQ1 and determine if all experts do interpret the SoS in the same way. The question designed to analyze this dimension (*Question1*) is described in Table 1.

(ii) Causes of the security vulnerability. This dimension explores two main areas and is covered by three Delphi questions. Its purpose is to provide answer to Research Question 2, (i.e., understand what are the causes of the vulnerabilities). These three questions allow us to understand if experts consider that these vulnerabilities emerge from the constituent systems or if the vulnerabilities are the result of an unplanned composition [37,38]. The responses help in identifying which is the nature of the threat for the vulnerabilities according to experts. On one side it studies the origin of the exploited vulnerability. A question (*Question 2*) asks about what is being affected because of the vulnerability, namely the consequences of the vulnerabilities (described in Table 2). Another question (*Question* 3) asks for those elements that could be analyzed to help in detecting the vulnerability before it is exploited. These elements represent the causes of the vulnerabilities (see Table 3). On the other side, a question (*Question 4*) asks specifically about the nature of the vulnerability to know if the experts consider the vulnerability as an emergent behavior (detailed in Table 4). This question highlights the nature of the vulnerability. Were the nature of the vulnerabilities not to belong to the SoS, this research area would be not of interest to promote the security on SoS.

(iii) Identify the nature of a solution. This third dimension is related to the third Research Question, aiming to identify a potential solution. Two Delphi questions are presented to determine the nature of a potential solution to address the security issue described in the scenarios. These questions would provide knowledge regarding the requirements of a solution and can help enhancing approaches that aim at assessing the security of SoS [5]. A question (*Question 5*) asks for the need of homogeneous security standard (Table 5) and helps in understanding if a

Table 1
Delphi question 1.

Question 1	
Statement	Among the following characteristics, which one would apply to this scenario?
Alternatives	A1. Operational Independence - A2. Managerial Independence - A3. Geographical distribution - A4. Emergent Behaviors - A5. Evolutionary Development Processes - A6. Heterogeneity of constituent Systems
Motivation	This question aims to determine the characteristics of described SoS. More precisely if these alternatives are present on the different scenarios. The alternatives are the characteristics belonging to a SoS. Each alternative has two options: *Yes* or *No*.

Table 2
Delphi question 2.

Question 2	
Statement	The vulnerability is affecting the…
Alternatives	A1. Data from a vulnerable constituent A2. Data from the whole set of systems involved A3. Functionalities from a vulnerable constituent A4. Functionalities from the whole set of systems involved A5. The reason why a vulnerable system is doing joint work A6. The reason why the whole set of systems involved are working together
Motivation	This question aims to determine the causes of the security vulnerability described. More precisely what is being affected because of the attack. The alternatives are the shared resources (Data and Functionalities) and the mission for the SoS. Each alternative has two options: *Yes* or *No*.

Table 3
Delphi question 3.

Question 3	
Statement	This vulnerability could be early detected by analyzing the...
Alternatives	A1. Data from a vulnerable constituent A2. Data from the whole set of systems involved A3. Functionalities from a vulnerable constituent A4. Functionalities from the whole set of systems involved A5. The reason why a vulnerable system is doing joint work A6. The reason why the whole set of systems involved are working together
Motivation	This question aims to determine the causes of the security vulnerability described. More precisely what could help in identifying the attack. The alternatives are the shared resources (Data and Functionalities) and the mission for the SoS. Each alternative has two options: *Yes* or *No*.

Table 4
Delphi question 4.

Question 4	
Statement	The vulnerability is an emergent behavior of this particular combination of systems.
Motivation	This question aims to determine the causes of the security vulnerability described. More precisely to determine if the attack is produced because of an unexpected behavior (emergent behavior). This question is ranked in a 1 to 5 Likert-scale.

Table 5
Delphi question 5.

Question 5	
Statement	The vulnerability is caused by a lack of agreement in homogeneous security levels
Motivation	This question aims to determine the nature of a solution. More precisely to identify the need of a homogeneous security level to prevent the vulnerability. This question is ranked in a 1 to 5 Likert scale.

homogeneous security is feasible and effective. A second question (*Question 6*) focuses on the need of guidelines to mitigate such SoS vulnerabilities (Table 6).

The questionnaire presents three different cases describing each one an attack on a different SoS. Such attacks are caused because of the shared resources among the constituent systems. The questionnaire asks for these three dimensions for each one of these three cases. The purpose of these cases is to get information about how the experts understand the emergent behaviors in SoS and generate knowledge from this. We have introduced three different cases on this survey to reduce the effects of single scenario misinterpretation when discovering the experts' point of view. These cases are not fully described due to the difficulty when

Table 6
Delphi question 6.

Question 6	
Statement	What is the minimum coordination that may be required to address this issue?
Alternatives	Work on demand Each work is planned and controlled without considering global guidelines A standard procedure, each work is planned and controlled uniformly among the parties
Motivation	This question aims to determine the nature of a solution. More precisely to identify the coordination level required to address the issue among the different SoS accountants. These alternatives are interpreted as follows: Low maturity level: *Work on demand.* Moderated maturity level: *Each work is planned and controlled without considering global guidelines.* High maturity level: *A standard procedure, each work is planned and controlled uniformly among the parties.* This question is ranked in a 1 to 3 Likert scale.

looking for formal definitions of SoS in the scientific literature. The cases described have been inspired by real attacks which are described in the web [39], and in Darktrace reports [40]. These cases belong to the health care domain (see Table 7), the social networks domain (see Table 8), and the nuclear plants domain (see Table 9). These domains were chosen to offer a diversity of cases.

A pilot round was conducted internally before launching the first round. In this preliminary round we detected spelling errors and design issues, which were addressed before launching the first round. Later, in the first round, the questionnaire included some additional questions to allow characterizing the participants. The only authentication required was a token that was sent to each participant individually, so that they can use it to continue the survey when desired, allowing to pause the questionnaire if needed, and avoiding registering participants' personal data. To provide the required background, the survey was designed with a brief introduction describing the purpose of the Delphi questionnaire, and descriptions about SoS and security.

After each round the participants received an anonymized report including a statistical analysis of the results and the comments provided by other experts to allow its analysis from their side before the next round begins.

2. Selecting the experts

There are no common guidelines in the literature to define a recommended number of participants or criteria to set up the fittest participant profile on the surveys. It depends on the analyzed topic.

This survey browses the effects of security vulnerabilities in a SoS, and the nature of a solution approach. Thus, the panel of experts should be composed of experts regarding *SoS, Security, Privacy, Trust*, and *Good*

Table 7
Security vulnerabilities. Case 1.

Case 1
ACME Health is a health care company. This clinic works in collaboration with a hospital. This clinic uses a central system that manages all the electronic healthcare records of their patients. In the clinic each doctor owns a computer in the office that they use for daily working. Doctors need to use biometric identification to unlock the office computer to avoid externals access patient's data. A company hired by the hospital, has installed new systems in the waiting room to manage the patient's queue with the aim of provide a better experience to their patients. This system is connected to the ACME Health main system among others to manage patients' queue. The purpose of this new system is to allow patients retrieve their data and prioritize them in the queue according to their history and give instruction to what office shall they go. An attacker exploited this new system and was able to gain access to the whole Electronic Health-Record every patient data in the waiting list for ACME Health revealing confidential data.

Table 8
Security vulnerabilities. Case 2.

Case 2
Honan is a journalist who was victim of an attacker that gathered information available on various systems he used. Honan is a very active internet user. At the time of this attack he used several services on different systems in Internet as Twitter, Amazon, Facebook, iCloud, Google, among others. Due to a controversial message in a social network, an offended attacker decided to destroy Mr. Honan digital life as a revenge. By combining information from one system after another, they could generate a complete profile of his life, use recovery password assistance, thus they were able to exploit the security on the systems and impersonate Mr. Honan's identity. Consequently, the attacker gained access to Mr. Honan digital life and was able to delete his entire digital existence.

Table 9
Security Vulnerabilities. Case 3.

Case 3
Systems in a nuclear power plant are being audited. The experts are analyzing Geiger counters that workers use to detect leaks and launch contingency plans if needed. During the audit, the experts detected that the Geiger counters being used had a security issue on its design. A vulnerability could be exploited on these counters because of their firmware, and the unencrypted data they send. An attacker may use this vulnerability to trick the power plant control room and send fake signals of radiation leaks, or to hide real radiation leaks warnings. Despite in the nuclear plants the Geiger counter systems are not directly joint working with reactor and control systems, it may become a real-life risk if leaks are not detected. Auditors concluded that other systems and sensors worked as expected.

practices. Among the invited participants there are active members in the SoS community, members of the program committee, and organizers of SoS related conferences such as *International Conference on System of Systems Engineering (SoSE)* and *International Workshop on Software Engineering for Systems-of-Systems (SESoS)*. The other participants who were not actively researching in SoS were examined on how their area of expertise would help this survey. The community of researchers in SoS is not as large as the communities in other more mature areas such as Testing, MDE, or Security. This fact makes it difficult to find experts to invite them to participate in this survey. For this reason, theoretical knowledge is a required quality and practical experience is only desired. Initially, 37 participants were invited to collaborate on this study; 18 of them accepted to participate, and 15 of them actually participated.

Table 10 summarizes the decision of the experts to participate. Table 11 summarizes the type of organization in which participants have identified their work profile. Among them, all participants have worked in academia for some time. However, four of them also worked for industrial companies, and they have been correspondingly identified. Finally, Table 12 shows the years of experience of the panel in their respective areas.

Regarding the geographical distribution of the experts, they are from 7 different countries (described in Table 13). This geographical distribution enhances the general responses of the questionnaire since cultural bias is reduced, providing different perspectives to the survey.

After reviewing the nature of the participants, a balanced group among diverse knowledge areas has been reached, which allowed us to conduct a Delphi study able to produce results by considering different perspectives.

Table 10
Responses to participation invitation.

Participants responses	Number	Percentage
Total experts invited	37	100%
Formally Refused	1	2.70%
Did not respond	16	43.24%
Agreed to participate	18	48.64%
Participated	15	40.54%

Table 11
Participants' organizations.

Organization	Number	Percentage
Academy (exclusively)	11	73.33%
Academy + Industry	4	26.66%

Table 12
Experts' experience.

Years of experience	< 3 Years	>= 3 and < 5 Years	> 5 Years
Experts in Good practices	0 (0%)	1 (6.66%)	14 (93.33%)
Experts in Security, Privacy, Trust	10 (66.66%)	1 (6.66%)	4 (26.66%)
Experts in SoS	4 (26.66%)	3 (20.00%)	8 (53.33%)

Table 13
Participants' geographical distribution.

Country	Number	Percentage
Argentina	2	13.33%
Belgium	1	6.66%
Brazil	2	13.33%
Germany	3	20.00%
Italy	5	33.33%
Netherlands	1	6.66%
Spain	1	6.66%

3. Statistical processing

A statistical processing of the results for each round is used to identify the general opinion of the panel. These results are also used to identify stability on the answers among rounds and to determine if experts have reached consensus on their answers.

The survey questions were measured using Likert scales. The alternatives chosen by each expert have been converted to a numeric value (as represented in Tables 14 and 15) so that the data for the statistical processing can be calculated for each one of the statements.

The consensus on a survey round can be measured by means of several approaches. A literature review conducted by Heiko [22] summarizes some of them. The analysis of the experts' responses on this questionnaire is made by using three different statistical techniques, each one with a different purpose. The responses are processed with a Descriptive Analysis to generalize the results, Cronbach Alpha to measure the reliability of the questionnaire and Kendall's W to measure the stability of the results among rounds.

Descriptive analysis. The questions in the survey were measured using a Likert scale. Notwithstanding, to produce a statistical study of the results these Likert values have been converted to a numeric value. Therefore, median and mode are calculated for each statement as rated by the experts.

To visualize the panel opinion, two keywords are defined to describe

Table 14
Likert scale interpretation.

5-points Likert Statement	Binary Likert Statement	Numeric value
Strong Disagreement	No	1
Disagree		2
Neutral		3
Agree		4
Strong Agreement	Yes	5

Table 15
Maturity level interpretation.

Question 6 statements	Numeric value
Low maturity level	1
Moderated maturity level	2
High maturity level	3

and summarize the experts' decision as *agreement* and *disagreement*. The *agreement* of the panel is measured by calculating the percentage of experts who chose *Yes, Agree,* or *Strong Agreement* on each one of their responses. Likewise, *disagreement* is measured through the percentage of experts who chose *No, Disagree,* or *Strong Disagreement*. For the sake of consistency when summarizing the results, Question 6 results have been categorized as *agreement* when experts chose a high maturity level and *disagreement* when experts chose a low maturity level.

These values are used to determine whether a question has reached consensus or not either for an *agreement* or *disagreement* cluster by highlighting the common opinion for each statement.

We require at least a 60% of the participants answering the same to determine that consensus has been reached either for *agreement* or *disagreement*. On the other hand, if more than 80% of the experts agree on the same response it is understood as a solid consensus. Table 16 summarizes these values.

In the survey we show three different cases of security vulnerabilities. However, the results of each round are analyzed by studying each one of these three cases individually. In this way, after analyzing the results of the three cases, we compare them and evaluate the general rating for each question. Such general rating can adopt four different values: (i) Consensus: Agreement and (ii) Consensus: Disagreement for those cases on which the experts provided a common response in at least two of three cases. (iii) No consensus reached determines these experts' choices where consensus has been not achieved for at least two cases. (iv) Depend on the case results determine these answers that reached a different consensus depending on the case.

Cronbach Alpha. The Cronbach Alpha [23,24] is used to measure the reliability of the metric used in each question to measure the value provided to the experts. The values provided by the Cronbach Alpha are interpreted according to the values described in [41] and summarized in Table 17.

Kendall's W. The Kendall's' W [26], also known as coefficient of concordance, is a statistic algorithm used to measure the agreement among raters assessing a certain number of items [26]. This algorithm is used to assess the similarity of two ordered sets. The more similar is the order of the two sets of items, the closer is the W of Kendall to 1. The more different ordered are the set of items, the closer is the Kendall's W to 0.

In the analysis of the results the Kendall's W is used to sort the alternatives' average agreement on each round and evaluate how similar the rank of responses agreement is, and then the similarities between rounds to assess the evolution of the responses.

4.2. Stage 2: conducting and reporting

The Delphi survey has been conducted with the help of 15 experts belonging to different research areas providing their opinion. The communication with these participants has been done by email, sending them individual and personalized messages for each one. The study started its first round on February 4th, 2020 and took three rounds to reach consensus. The study finished after the third round on July 10th, 2020, lasting 157 days in total. To participate, the experts received questionnaires containing all the defined questions by means of LimeSurvey tool [42].

Table 16
Consensus interpretation rates.

Assessment percentage	Interpretation
>80% on agreement or disagreement	Solid consensus
>60% on agreement or disagreement	Consensus
<60% on agreement or disagreement	Consensus not reached

Table 17
Cronbach's Alpha interpretation.

Cronbach Alpha Range	Interpretation
$1.0 > \alpha > 0.9$	Excellent
$0.9 > \alpha > 0.8$	Good
$0.8 > \alpha > 0.7$	Acceptable
$0.7 > \alpha > 0.6$	Questionable
$0.6 > \alpha > 0.5$	Poor
$0.5 > \alpha > 0.0$	Unacceptable

The execution of the Delphi method is described with details round by round, detailing the results of the experts' panel for each one. Readers looking for the outcome of this survey could jump to Section 4.3. Stage 3. where the results are described and discussed.

1. First Round

First Delphi Round started on February 4th, 2020 and was receiving answers from the experts until March 12th, 2020 (37 days). The results of this round are studied by using the median, and the mode, and homogeneity and concordance are described.

Descriptive analysis

The descriptive analysis for the rounds has been organized individually by cases. Therefore, the results of the questionnaire are grouped by each one of the cases.

As a general overview, this round does not show a well-defined trend. The values each question got from each case is detailed next.

Case 1. The results of the first round for Questions 1–6 over Case 1 are reported below in Tables 18–23, respectively.

Case 2. The results of the first round for Questions 1–6 over Case 2 are reported below in Tables 24–29, respectively.

Case 3. The results of the first round for Questions 1–6 over Case 3 are reported below in Tables 30–35, respectively.

Homogeneity and concordance analysis

The consensus for some alternatives is observed after analyzing the results from the three cases together. Table 36 summarizes the consensus for the first round and Table 37 outlines the statistical results.

Consensus: agreement: There is a consensus for agreement in at least two cases, and there is no case with disagreement consensus. The general result for each question or alternative is considered as consensus if the question has a consensus for the same hub in the three cases, or the question has a consensus for the same hub in two cases and the third one is without consensus. Consensus: Agreement is achieved for Q2 A1

Consensus: disagreement: There is a consensus for disagreement in at least two cases, and there is no case with agreement consensus. General disagreement is achieved for Q1 in A3 and A5; Q2 A6, and Q3 A2, A5

Table 18
Round 1 - case 1 - question 1.

Question	Alternative	Agreement	Disagreement
Q1	A1	47%	53%
	A2	80%	20%
	A3	20%	80%
	A4	53%	47%
	A5	7%	93%
	A6	47%	53%

The consensus was reached in 3/6 alternatives. A high number of the experts said that this case has *Managerial Independence* and the case lacks *Geographical distribution* and *Evolutionary Development Processes*. The remaining 3/6 alternatives (*Operational Independence, Heterogeneity of constituent Systems and Emergent Behaviors*) have not a clear position.

Table 19
Round 1 - case 1 - question 2.

Question	Alternative	Agreement	Disagreement
Q2	A1	60%	40%
	A2	60%	40%
	A3	33%	67%
	A4	7%	93%
	A5	20%	80%
	A6	13%	87%

There is a strong consensus in 3/6 alternatives: A2, A5, and A6. The remaining 3/6 alternatives have not a well-defined position.

Table 20
Round 1 - case 1 - question 3.

Question	Alternative	Agreement	Disagreement
Q3	A1	27%	73%
	A2	40%	60%
	A3	53%	47%
	A4	47%	53%
	A5	33%	67%
	A6	27%	73%

There is not an agreement among the experts for any of the alternatives for this question. A high number of experts share the same opinion in disagreeing with A1, A2, A5 and A6.

Table 21
Round 1 - case 1 - question 4.

Question	Mode	Median	Agreement	Disagreement
Q4	1	3	40%	40%

For this question there is not a decided position as agreement and agreement were chosen the same (40%).

Table 22
Round 1 - case 1 - question 5.

Question	Mode	Median	Agreement	Disagreement
Q5	4	4	53%	10%

For this question there is not a clear decision for one position. On one side the 53% of the participant agree with this sentence, 10% were disagree and the remaining 37% did not vote.

Table 23
Round 1 - case 1 - question 6.

Question	Mode	Median	High maturity level	Low maturity level
Q6	3	3	73%	13%

73% of the participants chose the third alternative for this question. Other two alternatives remained with a 13% of agreement each.

and A6.

Consensus not reached: No consensus reached for at least two cases. Other alternatives have not reached consensus in at least two of the cases. This lack of consensus is achieved in A4 for Q1, Q2 and Q3, Q4 and Q5.

Depends on the case: There is a difference of opinion among agreement and disagreement for the cases. The remaining of the questionnaire continued without a clear consensus whether agreement or disagreement. The more variable responses among the cases are for Q1 A2, Q1 A6, Q2 A2, and Q3 A3.

The Cronbach Alpha is calculated individually for each question, using the responses of the three cases. The Cronbach Alpha, as illustrated in Table 38, shows that the metric used to assess the experts' responses is excellent in this round.

Table 24
Round 1 - case 2 - question 1.

Question	Alternative	Agreement	Disagreement
Q1	A1	47%	53%
	A2	27%	73%
	A3	40%	60%
	A4	53%	47%
	A5	27%	73%
	A6	60%	40%

For this question, a strong disagreement has been reached in 2/6 alternatives: A2 and A5. A moderate consensus is reached in 2/6, being an agreement and a disagreement in A3 and A6. The remaining 2 alternatives have not a clear position.

Table 25
Round 1 - case 2 - question 2.

Question	Alternative	Agreement	Disagreement
Q2	A1	60%	40%
	A2	60%	40%
	A3	20%	80%
	A4	47%	53%
	A5	0%	100%
	A6	20%	80%

There is strong consensus in disagreeing with 3/6 alternatives for this question: A3, A5 and A6. Moderate agreement is achieved in two alternatives: A1 and A2.

Table 26
Round 1 - case 2 - question 3.

Question	Alternative	Agreement	Disagreement
Q3	A1	27%	73%
	A2	47%	53%
	A3	33%	67%
	A4	47%	53%
	A5	0%	100%
	A6	20%	80%

For Question 3, a Strong disagreement is reached in 4/6 alternatives: A1, A3, A5 and A6.

Table 27
Round 1 - case 2 - question 4.

Question	Mode	Median	Agreement	Disagreement
Q4	1	2,5	40%	47%

Experts are not decided whether this vulnerability is an emergent behavior of the SoS or not, with half of the panel deciding differently.

Table 28
Round 1 - case 2 - question 5.

Question	Mode	Median	Agreement	Disagreement
Q5	1	2,5	33%	47%

As it happened for Question 4 the experts were not able to reach a consensus for this one.

Table 29
Round 1 - case 2 - question 6.

Question	Mode	Median	High maturity level	Low maturity level
Q6	3	3	53%	7%

The panelists agreed that a higher coordination maturity level is required to address the identified issue.

Table 30
Round 1 - case 3 - question 1.

Question	Alternative	Agreement	Disagreement
Q1	A1	67%	33%
	A2	33%	67%
	A3	33%	67%
	A4	33%	67%
	A5	33%	67%
	A6	40%	60%

Experts reached an agreement in all the alternatives. Experts only agreed in the A1, reaching a disagreement in the remaining 5/6 SoS properties.

Table 31
Round 1 - case 3 - question 2.

Question	Alternative	Agreement	Disagreement
Q2	A1	67%	33%
	A2	27%	73%
	A3	67%	33%
	A4	47%	53%
	A5	40%	60%
	A6	40%	60%

Experts reached consensus in 5/6 alternatives. 2/5 are consensus for agreement (A1, A3), and the remaining 3/5 are consensus for disagreements (A2, A5, A6). The only alternative on which the experts showed not a clear position A4.

Table 32
Round 1 - case 3 - question 3.

Question	Alternative	Agreement	Disagreement
Q3	A1	60%	40%
	A2	7%	93%
	A3	87%	13%
	A4	20%	80%
	A5	27%	73%
	A6	13%	87%

This question reached consensus in all its alternatives, being 2/6 agreements (A1, A3) and 4/6 disagreements (A2, A4, A5 and A6).

Table 33
Round 1 - case 3 - question 4.

Question	Mode	Median	Agreement	Disagreement
Q4	4	3	40%	40%

Experts were not able to reach a consensus regarding if the vulnerability is an emergent behavior.

Table 34
Round 1 - case 3 - question 5.

Question	Mode	Median	Agreement	Disagreement
Q5	4	4	60%	27%

There is a moderate consensus in agreeing that the vulnerability is caused by a lack of a homogeneous security.

Table 35
Round 1 - case 3 - question 6.

Question	Mode	Median	High maturity level	Low maturity level
Q6	3	3	73%	20%

A high number of experts think that a higher coordination maturity level is recommendable to address the described issue.

Table 36
Round 1 - aggregated results.

Question	Alternative	Overall result
Q1	A1	Depends on the case
	A2	Depends on the case
	A3	Consensus: Disagreement
	A4	Consensus not reached
	A5	Consensus: Disagreement
	A6	Depends on the case
Q2	A1	Consensus: Agreement
	A2	Depends on the case
	A3	Depends on the case
	A4	Consensus not reached
	A5	Consensus: Disagreement
	A6	Consensus: Disagreement
Q3	A1	Depends on the case
	A2	Consensus: Disagreement
	A3	Depends on the case
	A4	Consensus not reached
	A5	Consensus: Disagreement
	A6	Consensus: Disagreement
Q4		Consensus not reached
Q5		Consensus not reached
Q6		Consensus: Agreement

Table 37
Round 1 - statistical results.

Result	Number
Consensus: Disagreement	28.57% (6)
Consensus: Agreement	4.76% (1)
Not consensus reached	23.80% (5)
Depends on the case	42.85% (9)

Table 38
Round 1 Cronbach Alpha values.

	Q1	Q2	Q3	Q4	Q5	Q6
Cronbach Alpha	0.98	0.97	0.97	0.99	0.95	0.96

Round results

The first round provided the experts' first impression, which are examined by the whole panel in further rounds.

As identified through the results analysis, the experts' responses have not provided a well-defined consensus within the responses. It is evidenced when observing the comments provided by the experts, which, eventually are contradictory. This is quite remarkable for question 4. Some of them are shown in Tables 39–41.

Additionally, experts provided some description of a potential solution, listed in Table 42, which allow to define how future work should be considered.

Also, as this is the first round, there is no chance of reaching stability. Thus, a second round is proposed to the participants. Experts were provided with an anonymized version of the results, and the survey responses to allow them to reconsider their previous vote or to reinforce their previous decision.

Table 39
Round 1. Contradiction A.

Case 1. Question 4.
The attacker only gained access to the patients' data thanks to a communication between the ACME Health, clinics software and queue software. That is, the vulnerability emerged from the combination of different systems.
I would consider something "emerging" where there is no formalized interaction, an evolution on the usages of the platform, users or natural evolution of the aggregation of the systems. In this case, the situation is quite standard and predictable.

Table 40
Round 1. Contradiction B.

Case 2. Question 4.
I don't think the vulnerability is due to the combination of systems, rather to the use of the different data inside each one of them.
Each system handles its own security, but the SoS system that emerges when using them together is more vulnerable than any single system stat-alone.

Table 41
Round 1. Contradiction C.

Case 3. Question 4.
This vulnerability can result in not handling a leak or treating a leak that does not exist. These are undesired emergent behaviors.
The security issue specifically concerns a vulnerable constituent.

Table 42
Potential solutions description.

All cases. Question 6
A standard procedure could help
Again, coordinating the security levels end-to-end seems to be the better strategy
As explained in the text on the link "The very four digits that Amazon considers unimportant enough to display in the clear on the Web are precisely the same ones that Apple considers secure enough to perform identity verification". So, if there is a standard procedure, maybe this vulnerability could be avoided.
It is important to consider global guidelines

The comments from the experts also let us to know that there are still room for improvement in the questionnaire design, which is improved to better reflect the experts' opinion in the further rounds.

2. Second round

The second round was launched on April 8th, 2020 and was receiving answers until May 25th, 2020 (47 days). In this round all the participants agreed to continue with the survey. Experts received a report summarizing an anonymous version of aggregated results on the first round and the comments provided by the experts. After reading considering this information four experts decided to keep their previous answers, and the remaining eleven participants changed their mind.

Questionnaire redesign

The questionnaire in this round suffered a minor revision to agree with the experts' comments. Experts agreed that Question 1, Question 2, and Question 3 were difficult to rate only by using binary values. Experts claimed that an absolute Yes or absolute No were not always compliant with their thinking. Experts argued that there were some conditions in the described cases that could be met occasionally, so just a Yes would not represent their real meaning.

To mitigate such flaw a wider range of responses was provided for Question 1, Question 2 and Question 3. Two more values are provided for this assessment. The result for these questions is that the possible answers are: "No", "Rarely", "Might be", and "Yes".

A numeric value was given to each value to measure the variation on the responses for these questions as shown in Table 43. These values override the values previously assigned to "*No*" and "*Yes*" in round 1 to allow calculating the variation on the responses.

Table 43
4-Point scale interpretation.

Likert statement	Clustered as	Numeric value
No	Disagreement	1
Rarely	Disagreement	2
Might be	Agreement	3
Yes	Agreement	4

The remaining of the questionnaire (Descriptive texts, and Questions 4, 5 and 6) remained without any changes.

Descriptive analysis

The results of the second round are described by analyzing each case individually as it was done for the first round.

Five participants did not modify their previous answers and decided not to modify any of their previous responses or justifications. The number of comments has not been affected in this second round. Only one of the participants did not provide details about their responses. As it has been done for the first round, the results for this second round are discussed grouped by cases.

Case 1. The results of the second round for Questions 1–6 over Case 1 are reported below in Tables 44–49, respectively.

Case 2. The results of the second round for Questions 1–6 over Case 2 are reported below in Tables 50–55, respectively.

Case 3. The results of the second round for Questions 1–6 over Case 3 are reported below in Tables 56–61, respectively.

Results discussion

Provided results were justified by the experts' as it was done in the previous round. These justifications help to discover the reasoning of the disagreements. The general understanding has been studied after analyzing the experts' answers according to the results obtained in each question.

The main contradictions highlighted when analyzing the results of Round 1 are still present on this round. Experts were not able to convince each other about their thinking. This could be motivated due to a lack of a common understanding on the SoS context. The definition of Emergent Behavior is the one that presents the major issue. Experts were unable to reach consensus when considering if the cases were describing an attack on a SoS or over a single system. See Tables 62 and 63.

The Cronbach Alpha is calculated individually for each question, using the responses of the three cases. The Cronbach Alpha shows in Table 64 that the metric used to assess the experts' responses is excellent in this round.

Kendall's W is used to measure the stability among rounds. The agreement on the responses for all the cases are ordered. The Kendall's W measures the variation in set of responses sorted by agreement percentage. Using this method, the Kendall's W presents a value of 0.61. This value represents a moderate consensus, still with room for improvement in further rounds.

Round results

The second round provides the experts' impression after analyzing the results provided in the first round.

The experts' analyzed the comments and the general results. Then the experts take one among three alternatives. To retain their previous results, to change their previous opinion, or to reinforce their position.

As identified through the results analysis (summarized in Tables 65 and 66), the experts' responses have increased their consensus, reducing the difference in the responses depending on the case. The number of questions without a clear consensus have been also increased because of reducing the difference of response among the questions.

The changes in the responses showed that stability is moderated but a third round is needed to confirm the stability on the answers. Thus, a

Table 44
Round 2 - case 1 - question 1.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q1	A1	1	2	40% (−7%)	60% (+7%)
	A2	4	4	87% (+7%)	13% (−7%)
	A3	1	1	33% (+13%)	67% (−13%)
	A4	4	4	53% (-)	33% (−14%)
	A5	2	1	13% (+6%)	80% (−13%)
	A6	4	4	80% (+33%)	20% (−33%)

Question 1 suffered a very relevant modification in Alternative 6. Consequently, this alternative passed from a not consensus to a consensus for the agreement. Other questions remained without relevant modifications.

Table 45
Round 2 - case 1 - question 2.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q2	A1	4	4	80% (+20%)	13% (−27%)
	A2	4	4	67% (+7%)	33% (−7%)
	A3	1	2	47% (+14%)	47% (−20%)
	A4	1	1	20% (+13%)	73% (−20%)
	A5	1	1	27% (+7%)	60% (−20%)
	A6	1	1	13% (-)	73% (−14%)

Question 2 experienced relevant changes in Alternative 1. The modification did not affect the assessment for question, which remained the same.

Table 46
Round 2 - case 1 - question 3.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q3	A1	4	1	33% (+6%)	67% (−6%)
	A2	1, 4	1	40% (-)	60% (-)
	A3	3, 4	3	53% (-)	40% (−7%)
	A4	2	2	40% (−7%)	53% (-)
	A5	1, 4	2	40% (+7%)	53% (−14%)
	A6	1, 2, 4	1,5	27% (-)	67% (−6%)

Question 3 got light variations in the assessment, with no relevant modifications.

Table 47
Round 2 - case 1 - question 4.

Question	Mode	Median	Agreement	Disagreement
Q4	1	2	27% (−13%)	53% (+13%)

Opinion for Question 4 moved to the disagreement. There is still not consensus reached, but the experts' tendency is to disagree in for this issue.

Table 48
Round 2 - case 1 - question 5.

Question	Mode	Median	Agreement	Disagreement
Q5	4	4	53% (-)	20% (+10%)

Question 5 was unaffected in this second round.

Table 49
Round 2 - case 1 - question 6.

Question	Mode	Median	High maturity level	Low maturity level
Q6	3	3	73% (-)	20% (+7%)

Question 6 was unaffected in this second round.

Table 50
Round 2 - case 2 - question 1.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q1	A1	4	3	60% (+13%)	40% (−13%)
	A2	4	4	53% (+26%)	47% (−26%)
	A3	1, 4	1	40% (-)	60% (-)
	A4	4	3.5	60% (+7%)	33% (−14%)
	A5	1	1	20% (−7%)	73% (-)
	A6	4	4	67% (+7%)	33% (−7%)

Question 1 experimented light modification in the values. These modifications have produced two new agreements, for alternative 1 and alternative 4. Previous consensus that disagreed alternative 3 (*Geographical distribution*) is no longer a consensus, as the 26% of the votes moved from disagreement to agreement.

third round is proposed to the participants. As it was done when starting round 2, the experts were provided with an anonymized version of the results and the survey responses to allow them to reconsider their previous vote or to reinforce their previous decision.

Table 51
Round 2 - case 2 - question 2.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q2	A1	4	4	80% (+20%)	20% (−20%)
	A2	4	4	67% (+7%)	33% (−7%)
	A3	1	1	33% (+13%)	67% (−13%)
	A4	4	1	40% (−7%)	60% (−7%)
	A5	1	1	13% (+13%)	80% (−20%)
	A6	1	1	27% (+7%)	67% (−13%)

Consensus for Question 2 improved, having all its alternatives with a consensus reached.

Table 52
Round 2 - case 2 - question 3.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q3	A1	1	1	33% (+6%)	60% (−13%)
	A2	1	2	40% (−7%)	53% (-)
	A3	1	3	53% (+20%)	47% (−20%)
	A4	3, 4	3	47% (-)	53% (-)
	A5	1	1	20% (+20%)	73% (−27%)
	A6	1	1	33% (+13%)	60% (−20%)

Question 3 had light and moderated modifications along the alternatives. A previous consensus for disagreeing alternative 3 has been lost, as 20% of the experts changed their mind to agree with this alternative.

Table 53
Round 2 - case 2 - question 4.

Question	Mode	Median	Agreement	Disagreement
Q4	1, 4	2	33% (−7%)	53% (+7%)

Question 4 suffered light modifications in the assessment but remain without consensus.

Table 54
Round 2 - case 2 - question 5.

Question	Mode	Median	Agreement	Disagreement
Q5	4	3	47% (+14%)	40% (−7%)

Question 5 suffered light modifications in the assessment but remain without consensus.

Table 55
Round 2 - case 2 - question 6.

Question	Mode	Median	High maturity level	Low maturity level
Q6	3	3	60% (+7%)	7% (-)

A 7% or participants changed their vote for a high maturity level in Question 6, making it a consensus for this option.

Table 56
Round 2 - case 3 - question 1.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q1	A1	4	4	73% (+6%)	27% (−6%)
	A2	4	2	47% (+14%)	53% (−14%)
	A3	1	2	40% (+7%)	60% (−7%)
	A4	4	1	40% (+7%)	53% (−14%)
	A5	1	1	20% (−13%)	73% (+6%)
	A6	4	3	47% (+7%)	40% (−20%)

Question 1 got no relevant modification it the assessments. These modifications produced three alternatives (2, 4, 6) to lose the consensus. The remaining alternatives remained without changes in the consensus.

Table 57
Round 2 - case 3 - question 2.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q2	A1	4	4	73% (+6%)	27% (−6%)
	A2	1	1	40% (+13%)	60% (−13%)
	A3	4	4	53% (−14%)	40% (+7%)
	A4	1, 2, 4	2	33% (−14%)	60% (+7%)
	A5	1	1	27% (−13%)	67% (+7%)
	A6	1	1	27% (−13%)	67% (+7%)

Question 2 suffered few modifications in the assessment on its alternatives. It removed an agreement consensus for alternative 3 and produced a disagreement consensus for alternative 4. The remaining alternatives remained without changes in the consensus.

Table 58
Round 2 - case 3 - question 3.

Question	Alt.	Mode	Median	Agreement	Disagreement
Q3	A1	4	4	60% (-)	27% (−13%)
	A2	1	1	7% (-)	87% (−6%)
	A3	4	4	80% (−7%)	7% (−6%)
	A4	1	1	13% (−7%)	73% (−7%)
	A5	1, 4	1	20% (−7%)	67% (−6%)
	A6	1, 4	1	33% (+20%)	53% (−34%)

Relevant changes in alternative 6 for Question 3 cleared a consensus for disagreement. The remaining alternatives remained without changes in the consensus.

Table 59
Round 2 - case 3 - question 4.

Question	Mode	Median	Agreement	Disagreement
Q4	1, 4	2	47% (+7%)	53% (+13%)

Question 4 did not experience relevant modification and still remains without consensus despite in this round the 100% of the experts chose an option for this question.

Table 60
Round 2 - case 3 - question 5.

Question	Mode	Median	Agreement	Disagreement
Q5	4	4	73% (+13%)	13% (−14%)

Question 5 remained without changes in the consensus.

Table 61
Round 2 - case 3 - question 6.

Question	Mode	Median	High maturity level	Low maturity level
Q6	3	3	80% (+7%)	7% (−13%)

Question 6 remained without changes in the consensus.

Table 62
Round 2. Contradiction A.

Case 2. Question 4.
Each system handles its own security, but the SoS system that emerges when using them together is more vulnerable than any single system stand-alone.
Emergent behavior (= defined with "New functionalities") is not the reason, because there are no new functionalities that have emerged in the SoS.

3. Third Round

The third round was launched on June 12th, 2020 and was receiving answers until July 10th, 2020 (28 days). As in previous rounds, all the fifteen experts agreed to continue with this round. A report containing a summary from the second round was provided to the experts.

Table 63
Round 2. Contradiction B.

Case 3. Question 4.
The vulnerability is an emergent behavior since the problem happens when leakage sensors interact with the other constituent systems. Other systems and sensors work as expected.
I don't see here any of the element of emergent behavior.

Table 64
Round 2 Cronbach Alpha values.

	Q1	Q2	Q3	Q4	Q5	Q6
Cronbach Alpha	0.98	0.97	0.97	0.99	0.95	0.96

Table 65
Round 2 - statistical results.

Result	Number	Difference with round 1
Consensus: Disagreement	33.33% (7)	+1
Consensus: Agreement	19.06% (4)	+3
Not consensus reached	33.33% (7)	+2
Depends on the case	14.27% (3)	−6

Table 66
Round 2 - aggregated results.

Question	Alternative	Overall result	Previous result
Q1	A1	Depends on the case	Same
	A2	Consensus: Agreement	Depends on the case
	A3	Consensus: Disagreement	Same
	A4	Consensus not reached	Same
	A5	Consensus: Disagreement	Same
	A6	Consensus: Agreement	Depends on the case
Q2	A1	Consensus: Agreement	Same
	A2	Depends on the case	Same
	A3	Consensus not reached	Depends on the case
	A4	Consensus: Disagreement	Consensus not reached
	A5	Consensus: Disagreement	Same
	A6	Consensus: Disagreement	Same
Q3	A1	Depends on the case	Same
	A2	Consensus: Disagreement	Same
	A3	Consensus not reached	Depends on the case
	A4	Consensus not reached	Same
	A5	Consensus not reached	Consensus: Disagreement
	A6	Consensus: Disagreement	Same
Q4		Consensus not reached	Same
Q5		Consensus not reached	Same
Q6		Consensus: Agreement	Same

The questionnaire for this round was the same as for the previous one. According to the feedback given by the experts the questionnaire was able to capture their knowledge and they were able to provide a proper answer to the questions.

All experts ratified their previous assessment, therefore there are no changes on the results already described in round 2. The analysis of such data is omitted in this manuscript for this round to avoid unnecessarily repeating the same results. In this meaning, the conclusion for this round remains the same as the previous round, without any modifications.

Regarding stability, as no changes occur in this round the Kendall's W is 1. As stability is reached, the Delphi method can be considered as concluded with a consensus reached on 14 of 21 issues (66%), being 3 of them (14%) a consensus that depends on the case being studied.

4.3. Stage 3: results

After conducting three rounds using the Delphi method, we have

been able to discover what are the experts' opinion for each one of the three dimensions being studied in this study. In this section we have summarized the experts' agreements and disagreements along the Delphi rounds by using tables that describe the evolution of each question.

The results obtained after applying the Delphi method show that the experts did not reach consensus on all the questions for all the cases. In fact, a wide diversity of opinions was found depending on the case. This difference of opinions between one case and another might be caused by the complexity of the topic, as there is not a single point of view. The reasoning is that the way of dealing with the SoS security is different depending on the SoS and the vulnerabilities under consideration. On the one hand, experts answered with neutral marks when they did not have evidence enough, which made not reaching consensus on some questions. On the other hand, when experts reached consensus, they tended more frequently to disagree than to agree.

As a final remark, some experts suggested that a better opinion would be provided if well-defined cases describing real examples plenty of details were used, so that they would analyze the SoS better.

Dimension 1. Characteristics of described SoS.

The first dimension aims to discover if experts interpret SoS in the same way. Question 1 asked about the characteristics that define a SoS according to the literature for each case. At the end of the three Delphi rounds the experts reached a consensus by describing the scenarios with following features:

- A1: Some cases have operational independence.
- A2: Only one of the cases clearly has shown Managerial Independence.
- A3: Do not have geographical distribution.
- A5: Do not have evolutionary development processes.
- A6: Have heterogeneity of constituent systems.

On one hand, experts agreed that described SoS had *Operational Independence* and *Heterogeneity of constituent Systems*. On the other hand, experts converged on disagreeing about the *Geographical distribution* or *Evolutionary Development Processes*.

The results of this dimension show that experts could not reach consensus when determining if the SoS presented in the cases had *Managerial Independence* or to discern the *Emergent Behaviors*. Having such a variety of opinions from the experts has made it explicit that identifying whether a SoS does or does not possess some characteristics is far from simple, even though these characteristics have remained the same for more than two decades. These results show that, given the current definitions, there is still margin for misunderstanding. Table 67 summarizes the evolution of experts' agreement for this question.

Dimension 2. Causes of the security vulnerabilities.

The second dimension aims to determine how the presented vulnerabilities were originated and how it could have been prevented.

Question 2 in the Delphi survey asked for what items were affected by the vulnerability. According to this question, experts agreed that data from all the systems have been affected on all cases. Experts also agreed that joint work on the SoS are not affected on these cases. The results of the questionnaire is summarized in Table 68, where experts agreed on the following for the presented cases:

- A1: Vulnerabilities do affect data from a vulnerable constituent.
- A2: Vulnerabilities could affect data from the whole set of systems involved.
- A4: Vulnerabilities do not affect functionalities from the whole set of systems involved.
- A5: Vulnerabilities do not affect the reason why a vulnerable system is doing joint work.
- A6: Vulnerabilities do not affect the reason why the whole set of systems involved are working together.

Regarding Question 3, it asked for ways of preventing such vulnerabilities from being exploited. This question did not have consensus among the experts. In fact, experts only agreed partially on two statements. The lack of consensus is an indication that there are not bases on which base a common opinion. Agreement for this question has been summarized in Table 69.

For the described cases, experts reached consensus to disagree that:

- A1: Vulnerabilities could be early potentially detected by analyzing the data from a vulnerable constituent.
- A5: Vulnerabilities could be early detected by analyzing the reason why the whole set of systems involved are working together.

Question 4 identifies the vulnerability as an Emergent Behavior on the SoS. The lack of consensus in Question 4 (either agreement and disagreement as summarized in Table 70) is produced because some experts' think that the case description depicts a vulnerability generated by a single system and not because of the shared resources on the SoS. So, these experts consider that these vulnerabilities should not be considered as vulnerabilities generated because of an SoS emergent behavior.

Vulnerabilities in SoS have been proven to have an impact on the reliability of SoS, and therefore vulnerabilities resonate with the result of the shared goal among the constituent systems. As it has been identified in a recent SMS [43], malicious actions that cause malfunctions

Table 67
Question 1 summary.

Alt.	Case	Round 1		Round 2 & Round 3	
		Agreement	Disagreement	Agreement	Disagreement
Alt. 1	Case 1	47%	53%	40% (−7%)	60% (+7%)
	Case 2	47%	53%	60% (+13%)	40% (−13%)
	Case 3	67%	33%	73% (+6%)	27% (−6%)
Alt. 2	Case 1	80%	20%	87% (+7%)	13% (−7%)
	Case 2	27%	73%	53% (+26%)	47% (−26%)
	Case 3	33%	67%	47% (+14%)	53% (−14%)
Alt. 3	Case 1	20%	80%	33% (+13%)	67% (−13%)
	Case 2	40%	60%	40% (-)	60% (-)
	Case 3	33%	67%	40% (+7%)	60% (−7%)
Alt. 4	Case 1	53%	47%	53% (-)	33% (−14%)
	Case 2	53%	47%	60% (+7%)	33% (−14%)
	Case 3	33%	67%	40% (+7%)	53% (−14%)
Alt. 5	Case 1	7%	93%	13% (+6%)	80% (−13%)
	Case 2	27%	73%	20% (−7%)	73% (-)
	Case 3	33%	67%	20% (−13%)	73% (+6%)
Alt. 6	Case 1	47%	53%	80% (+33%)	20% (−33%)
	Case 2	60%	40%	67% (+7%)	33% (−7%)
	Case 3	40%	60%	47% (+7%)	40% (−20%)

Table 68
Question 2 summary.

Alt.	Case	Round 1		Round 2 & Round 3	
		Agreement	Disagreement	Agreement	Disagreement
Alt. 1	Case 1	60%	40%	80% (+20%)	13% (−27%)
	Case 2	60%	40%	80% (+20%)	20% (−20%)
	Case 3	67%	33%	73% (+6%)	27% (−6%)
Alt. 2	Case 1	60%	40%	67% (+7%)	33% (−7%)
	Case 2	60%	40%	67% (+7%)	33% (−7%)
	Case 3	27%	73%	40% (+13%)	60% (−13%)
Alt. 3	Case 1	33%	67%	47% (+14%)	47% (−20%)
	Case 2	20%	80%	33% (+13%)	67% (−13%)
	Case 3	67%	33%	53% (−14%)	40% (+7%)
Alt. 4	Case 1	7%	93%	20% (+13%)	73% (−20%)
	Case 2	47%	53%	40% (−7%)	60% (−7%)
	Case 3	47%	53%	33% (−14%)	60% (+7%)
Alt. 5	Case 1	20%	80%	27% (+7%)	60% (−20%)
	Case 2	0%	100%	13% (+13%)	80% (−20%)
	Case 3	40%	60%	27% (−13%)	67% (+7%)
Alt. 6	Case 1	13%	87%	13% (-)	73% (−14%)
	Case 2	20%	80%	27% (+7%)	67% (−13%)
	Case 3	40%	60%	27% (−13%)	67% (+7%)

Table 69
Question 3 summary.

Alt.	Case	Round 1		Round 2 & Round 3	
		Agreement	Disagreement	Agreement	Disagreement
Alt. 1	Case 1	27%	73%	33% (+6%)	67% (−6%)
	Case 2	27%	73%	33% (+6%)	60% (−13%)
	Case 3	60%	40%	60% (-)	27% (−13%)
Alt. 2	Case 1	40%	60%	40% (-)	60% (-)
	Case 2	47%	53%	40% (−7%)	53% (-)
	Case 3	7%	93%	7% (-)	87% (−6%)
Alt. 3	Case 1	53%	47%	53% (-)	40% (−7%)
	Case 2	33%	67%	53% (+20%)	47% (−20%)
	Case 3	87%	13%	80% (−7%)	7% (−6%)
Alt. 4	Case 1	47%	53%	40% (−7%)	53% (-)
	Case 2	47%	53%	47% (-)	53% (-)
	Case 3	20%	80%	13% (−7%)	73% (−7%)
Alt 5	Case 1	33%	67%	40% (+7%)	53% (−14%)
	Case 2	0%	100%	20% (+20%)	73% (−27%)
	Case 3	27%	73%	13% (−7%)	73% (−7%)
Alt. 6	Case 1	27%	73%	20% (−7%)	67% (−6%)
	Case 2	20%	80%	33% (+13%)	60% (−20%)
	Case 3	13%	87%	33% (+20%)	53% (−34%)

Table 70
Question 4 summary.

Case	Round 1		Round 2 & Round 3	
	Agreement	Disagreement	Agreement	Disagreement
Case 1	40%	40%	27% (−13%)	53% (+13%)
Case 2	40%	47%	33% (−7%)	53% (+7%)
Case 3	40%	40%	47% (+7%)	53% (+13%)

are one of the main factors that can jeopardize a SoS. Moreover, if the human factor is considered a constituent system within the SoS, it would be the one providing more vulnerabilities to the whole SoS as the human factor is prone to social engineering attacks [44]. Not having a standard concept of whether emergent behaviors do or do not involve the existence of vulnerabilities may well be a challenge when designing a solution to improve the security of SoS.

Dimension 3. Identify the nature of a solution.

Research Question 3 aims to identify the nature of a potential solution for SoS vulnerabilities.

Question 5 identifies if the attack is originated because of a lack of homogeneity in the security. This question reached consensus only in one of the cases, as summarized in Table 71. Nevertheless, experts did agree in their justification that a homogeneous security level improves the security. The reason behind this apparent contradiction is that some experts also thinks that the attackers could keep exploiting such homogeneous security level even if homogeneous security level was used.

Therefore, experts could not reach consensus to determine what is causing the vulnerabilities. They did agree on some of the alternatives of what could be affected, and how a solution would be. However, they were not able to identify if emergent behaviors are involved in the presented cases, or if these attacks were produced because a lack of a standard security among the constituent systems.

Question 6 asked about the structure and behavior of a solution to prevent emergent behaviors as vulnerabilities. Experts agreed that the minimum coordination that may be required to address this issue is planning and controlling individual works without considering global

Table 71
Question 5 summary.

Case	Round 1		Round 2 & Round 3	
	Agreement	Disagreement	Agreement	Disagreement
Case 1	53%	10%	53% (-)	20% (+10%)
Case 2	33%	47%	47% (+14%)	40% (−7%)
Case 3	60%	27%	73% (+13%)	13% (−14%)

Table 72
Question 6 summary.

Case	Round 1		Round 2 & Round 3	
	High mat. level	Low mat. level	High mat. level	Low mat. level
Case 1	73%	13%	73% (-)	20% (+7%)
Case 2	53%	7%	60% (+7%)	7% (-)
Case 3	73%	20%	80% (+7%)	7% (−13%)

guidelines. This is evidenced with a general agreement on all the three cases, as summarized in Table 72.

Notwithstanding, it has been identified that existing SoS documentation is not often shared among the SoS accountants. Hence, there is a limited shared knowledge about the structures and behaviors of constituent systems [45]. Though, enhancing the security of a SoS requires detecting and correcting deviations in the behavior, which in turn requires from a visualization of constituent systems' interactions and their erroneous behavior. Furthermore, the human factor has an unpredictable behavior. Therefore, the need of detecting and correcting the erroneous behavior is a must if the human factor acts as a constituent system.

5. Discussion

Experts have discussed about the origin of vulnerabilities as emergent behaviors in SoS, and they have also described the nature of a potential solution to prevent them. Considering the background and the results obtained in this survey, we observe that there is interest in this area of research. This conclusion comes from the observation of the many manuscripts published in this context, and from the interest shown by the experts on this topic in their responses. The general opinion of the experts was that this topic is relevant for both academia and industry. However, experts seemed not to share the same understanding of the concepts in their responses, and the lack of homogeneous standards to define SoS has been highlighted.

In order to diagnose the causes of a security problem in a SoS, it seems that it is essential to clearly define the characteristics of a SoS. Due to the difficulty in understanding the characteristics of a SoS, more specifically concerning *Management independence* and *Emergent Behavior*, experts could not reach a consensus to determine the cause the vulnerabilities. Also, they were not able to identify whether emerging behaviors are involved in the presented cases, or these attacks were caused by the lack of a homogeneous security among the constituent systems.

So far, the level of information available to assess and evaluate SoS security and the capability to coordinate such information may considerably vary according to the architecture used to compose the SoS. These challenges have been identified in a previous study [37]. Such a study examines the differences when comparing security properties in different architectures. For instance, when comparing Directed and Virtual architectures, the existence of a central entity coordinating the SoS goal achievement, and the existence of interaction guidelines may help in determining policies and expected behavior for those constituent systems within the SoS, whereas in a virtual SoS it is quite difficult to control or predict the behavior of a constituent system due to its nature. Therefore, considering the current state of the art it would be easier to predict and prevent emergent behavior-based vulnerabilities in a Directed or Collaborative SoS rather than in an Acknowledged or Virtual one.

Previous research has focused on modeling the functionalities of SoSs by considering the available information. On the one hand, there are some approaches to model specific SoS features such as mKAOS [46] (a model used to describe the goal, or *mission*, of the joint work) or SoSADL [47] (a language used to describe the architectural composition). On the other hand, there is an approach aiming to describe the composition of constituent systems and includes relevant information to better understand the SoS business logic within the SoIS context [48] (a model that

involves social and economic systems as constituent systems). Also, there is an approach that considers modeling SoS emergent behaviors [49]. Notwithstanding, even though these artifacts are useful to provide information and to describe SoSs regarding a specific context, there has not been defined a standard language to represent simultaneously the SoS components and the responsibilities for each of the constituent systems. Such language should consider depicting what each system provides and requires, what each system is accountant of, and the humans that would interact with it. Additionally, further efforts would be required to support Virtual SoS formal descriptions.

On the one hand, the lack of a common vocabulary is evidence of an emerging area of knowledge. Therefore, we can conclude that the SoS vocabulary has not yet a quite well-defined meaning. The absence of a homogeneous vocabulary hampers the development of this research area, as researchers cannot easily share or compare information. A controlled vocabulary would help avoid misunderstandings and standardize the meaning of each terminology or concept. This is in fact one of the limitations of this study, as the descriptions of the cases in the Delphi survey have been interpreted in different ways by the experts.

On the other hand, current artifacts designed to describe SoSs have not been considered to describe the *'role'* or *'responsibility'* each constituent system takes when participating in a SoS. Consequently, there is not a mechanism to determine which constituent system may be required to develop further countermeasures to protect the SoS, or which one should oversee the expected behavior among the constituent systems. Therefore, it is not simple to determine the value each one provides to the SoS, or the security requirements each constituent system could call for when doing the joint work.

6. Conclusions

Previous studies analyzed the research challenges of security on SoS [5,37] and identified the existence of an issue originated on the collaboration of constituent systems on a SoS. The definition of such issue, however, might not be relevant if the issue itself is not of interest for academia or industry. The main goal for this study is to validate the existence of such security problem that arise as a combination of shared resources in the SoS and evaluate the interest of researchers and practitioners on this topic. That goal is pursued by using the Delphi Method.

The three research questions that led this study were translated into a survey of six questions over three cases that have been used along the different rounds of the Delphi method. The main challenge of surveys of this kind is the low participation ratio. It is difficult to find participants that can devote the significant time and effort that such surveys would require. The Delphi method was conducted with 15 participants and required 157 days to be completed.

The questionnaire covered three different areas regarding security and SoS, namely: *SoS characteristics, causes of security vulnerabilities* and *identifying the nature of a solution*. Experts helped in retrieving useful perspectives from the academia and the industry on areas such as *SoS, Security* and *Good Practices*. On the one hand the statistical values used to identify consensus and stability from experts' responses proved to be useful to interpret the general results and how experts perceived this topic. On the other hand, the textual feedback received as comments or justifications from the experts in this study helped in knowing and understanding about their ideas and their way of thinking.

After analyzing the results, three gaps were identified as some concepts (e.g., systems responsibilities, purpose of participating in a SoS, emergent behaviors) were not interpreted the same among the experts.

Gap 1. Dimension 1 showed that experts could not reach consensus when determining if the SoS presented some characteristics or not. Some of them are difficult of interpret, as it is not clear how far the systems need to be to determine if the *Geographical Distribution* characteristic applies. Furthermore, other characteristics with a well-defined meaning as *Managerial Independence*, or *Emergent Behaviors* did not reached consensus either agreeing or disagreeing. This fact emphasizes the difficulty of understanding the characteristics of a SoS. Not having the same conceptualization of the characteristics of a SoS is evidence of the lack of a common vocabulary. These differences might be hampering the definition of case studies or even industrial scenarios that would enable the transference of SoS research to the industry.

Gap 2. Considering Dimension 2, answers to Question 3 evidenced how the lack of a common understanding can affect security. In Question 2 experts could determine what is being affected on each one of the three cases. However, Question 3, which asked about what could be examined to prevent the vulnerabilities, did not reach consensus in Alternative 3 "*Functionalities from a vulnerable constituent*" and Alternative 4 "*Functionalities from the whole set of systems involved*".

The lack of consensus to determine either studying functionalities is useful or not to prevent vulnerabilities, seems to be related with the lack of consensus in the *Emergent Behavior* characteristic. Experts were not able to determine if the functionalities enabling the attacks to succeed are considered to belong to the "*vulnerable constituent*" or to the "*whole set of systems involved*". This fact highlights the importance of a detailed definition of which functionalities are used in the joint work within the SoS. Such a definition would be easier to determine by establishing a shared document of shared resources among the constituent systems. Moreover, including the capabilities of humans who interact with the constituent systems, or their shared resources would allow to detect misbehavior and defend from social engineering attacks as well.

Gap 3. Question 4 and Question 5 from the survey did not reach consensus either on agreement or on disagreement because experts were considering different perspectives for *Emergent Behaviors* and *Homogeneity of Security*. Some experts considered that the vulnerabilities on the SoS produced from one single system should not be considered as an Emergent Behavior, whereas other experts claimed that such vulnerabilities belong to the SoS as it would not have such vulnerability without that constituent system. In a similar way, experts were not able to define if lack of homogeneous security is the source of these vulnerabilities. This gap arises as there is not well-defined criteria to determine the responsibilities each constituent system takes.

Challenges. Future work should consider improving the definition and understanding of SoS concepts to define a common shared interpretation of these concepts (e.g., SoS characteristics, the origination and effects of emergent behaviors or SoS accountability) and explore the impact on SoS security when considering the human factor as an ordinary constituent system. Also, considering the *Homogeneity of Security*, the definition of a common definition of "Security" for all constituent systems within a SoS might be required, as it would improve the definition of defensive mechanisms to coordinately protect the shared resources. Additionally, a backup plan to guarantee the common goal achievement in the event of a constituent system failure would be desirable.

Regarding the threats of this survey some experts missed more technical details to better identify how the vulnerability arose and how it could have been prevented. The results of this study could be improved by describing a fully developed case study, which could help in a deeper analysis of the analysis of data from a vulnerable constituent, the opinion of experts.

The results obtained from the Delphi questionnaire also depicted a solution approach. Therefore, an extension of [5] is scheduled as a framework. Such framework would detail a set of stages and activities to assess the security in the context of SoS.

CRediT authorship contribution statement

Miguel A. Olivero: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Antonia Bertolino:** Conceptualization, Methodology, Validation, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition. **Francisco José Dominguez-Mayo:** Conceptualization, Methodology, Writing – original draft. **Ilaria Matteucci:**

Conceptualization, Writing – original draft. **María José Escalona:** Conceptualization, Validation, Supervision, Writing – original draft, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been partially supported by the MIUR PRIN 2017 Project: SISMA (Contract 201752ENYB). and by the Spanish Ministry of Science, Innovation and Universities (NICO, PID2019-105455GB-C31). The authors would like to thank the experts participating in the survey as without their generous availability and effort this work would not have been possible. Likewise, the authors thank the anonymous reviewers who provided valuable suggestions to improve the quality of this manuscript.

References

[1] K.E. Boulding, General systems theory - the skeleton of science, Manage. Sci. 2 (3) (1956) 197–208.
[2] M.W. Maier, Architecting principles for systems-of-systems, Syst. Eng. 1 (4) (1998) 267–284, doi: 10.1002/(SICI)1520-6858(1998)1:4%3C267::AID-SYS3%3E3.0.CO;2-D.
[3] H. Cadavid, V. Andrikopoulos, P. Avgeriou, Architecting systems of systems: a tertiary study, Inf. Softw. Technol. (2020), https://doi.org/10.1016/j.infsof.2019.106202.
[4] P.G. Teixeira, V.H. Lazaro Lopes, R. Pereira Dos Santos, M. Kassab, V.V. Graciano Neto, The status quo of systems-of-information systems, in: Proceedings - 2019 IEEE/ACM 7th International Workshop on Software Engineering for Systems-of-Systems and 13th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems, SESoS-WDES 2019, 2019, pp. 34–41, https://doi.org/10.1109/SESoS/WDES.2019.00013, no. i.
[5] M.A. Olivero, A. Bertolino, F.J. Dominguez-mayo, M.J. Escalona, I. Matteucci, Security assessment of systems of systems, in: Proceedings - 2019 IEEE/ACM 7th International Workshop on Software Engineering for Systems-of-Systems and 13th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems, SESoS-WDES 2019, May 2019, pp. 62–65, https://doi.org/10.1109/SESoS/WDES.2019.00017.
[6] M.A. Olivero, A. Bertolino, F.J. Domínguez-Mayo, M.J. Escalona, I. Matteucci, Digital persona portrayal: identifying pluridentity vulnerabilities in digital life, J. Inf. Secur. Appl. 52 (2020), 102492, https://doi.org/10.1016/j.jisa.2020.102492. Jun.
[7] M.Á. Olivero, A. Bertolino, F.J. Dominguez-Mayo, M.J. Escalona, and I. Matteucci, "A systematic mapping study on security in systems of systems," Unpublished manuscript.
[8] N.C. Dalkey, The Delphi method: an experimental study of group opinion. Studies in the Quality of Life: Delphi and Decision-Making, 1972.
[9] M.W. Maier, Research Challenges for Systems-of-Systems Context : Collaborative Systems, Aerospace Corporation, 2005, pp. 1–6.
[10] J. Klein, H. Van Vliet, A systematic review of system-of-systems architecture, in: QoSA 2013 - Proceedings of the 9th International ACM Sigsoft Conference on the Quality of Software Architectures, 2013, pp. 13–21, https://doi.org/10.1145/2465478.2465490.
[11] J. Axelsson, A systematic mapping of the research literature on system-of-systems engineering, in: 2015 10th System of Systems Engineering Conference (SoSE), 2015, pp. 18–23.
[12] M. Guessi, V.V.G. Neto, T. Bianchi, K.R. Felizardo, F. Oquendo, E.Y. Nakagawa, A systematic literature review on the description of software architectures for systems of systems, in: Proceedings of the ACM Symposium on Applied Computing, 2015, pp. 1433–1440, https://doi.org/10.1145/2695664.2695795, vol. 13-17-Apri, no. v.
[13] T. Bianchi, D.S. Santos, K.R. Felizardo, Quality attributes of systems-of-systems: a systematic literature review, in: Proceedings - 3rd International Workshop on Software Engineering for Systems-of-Systems, SESoS 2015, 2015, pp. 23–30, https://doi.org/10.1109/SESoS.2015.12.
[14] I.G. Vargas, T. Gottardi, R. Teresinha, V. Braga, Approaches for integration in system of systems: a systematic review, in: Proceedings - 4th International Workshop on Software Engineering for Systems-of-Systems, SESoS 2016, 2016, pp. 32–38, https://doi.org/10.1145/2897829.2897835.
[15] M. Daneva, B. Lazarov, Requirements for smart cities: results from a systematic review of literature, in: 2018 12th International Conference on Research Challenges in Information Science (RCIS), 2018, pp. 1–6, https://doi.org/10.1109/RCIS.2018.8406655. , May 2018vol.-May.
[16] M. Diehl, W. Stroebe, Productivity loss in brainstorming groups: toward the solution of a riddle, J. Pers. Soc. Psychol. (1987), https://doi.org/10.1037/0022-3514.53.3.497.
[17] R.B. Wentworth, A.F. Osborn, Applied Imagination, J. Mark. (1955), https://doi.org/10.2307/1248180.
[18] H.A. von der Gracht, Consensus measurement in Delphi studies, Technol. Forecast. Soc. Change (2012), https://doi.org/10.1016/j.techfore.2012.04.013.
[19] O. Carney, J. McIntosh, A. Worth, The use of the nominal group technique in research with community nurses, J. Adv. Nurs. (1996), https://doi.org/10.1046/j.1365-2648.1996.09623.x.
[20] A. Alarabiat, I. Ramos, The delphi method in information systems research (2004-2017), Electron. J. Bus. Res. Methods (2019), https://doi.org/10.34190/JBRM.17.2.04.
[21] V. Mahajan, H.A. Linstone, M. Turoff, The Delphi method: techniques and applications, J. Mark. Res. (1976), https://doi.org/10.2307/3150755.
[22] Joseph A. Gliem, Rosemary R. Gliem, Calculating, interpreting, and reporting Cronbach's Alpha reliability coefficient for likert-type scales, in: 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education, 2003 Conference (Columbus, Ohio : Ohio State University), 2003, https://doi.org/10.1109/PROC.1975.9792.
[23] L.J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika (1951), https://doi.org/10.1007/BF02310555.
[24] N. Schmitt, Uses and abuses of coefficient alpha, Psychol. Assess. (1996), https://doi.org/10.1037//1040-3590.8.4.350.
[25] J. Cohen, Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit, Psychol. Bull. (1968), https://doi.org/10.1037/h0026256.
[26] P. Legendre, Species associations: the Kendall coefficient of concordance revisited, J. Agric. Biol. Environ. Stat. 10 (2) (2005) 226–245, doi: 10.1198/108571105 × 46642.
[27] H.O. Hirschfeld, A connection between correlation and contingency, Math. Proc. Camb. Philos. Soc. (1935), https://doi.org/10.1017/S0305004100013517.
[28] M.D. Dawson, P.S. Brucker, The utility of the Delphi method in MFT research, Am. J. Fam. Ther. (2001), https://doi.org/10.1080/01926180152026115.
[29] R.T. Nakatsu, C.L. Iacovou, A comparative study of important risk factors involved in offshore and domestic outsourcing of software development projects: a two-panel Delphi study, Inf. Manage. (2009), https://doi.org/10.1016/j.im.2008.11.005.
[30] R. Schmidt, K. Lyytinen, M. Keil, P. Cule, Identifying software project risks: an international Delphi study, J. Manage. Inf. Syst. (2001), https://doi.org/10.1080/07421222.2001.11045662.
[31] N. Dalkey, O. Helmer, An experimental application of the Delphi method to the use of experts, Manage. Sci. (1963), https://doi.org/10.1287/mnsc.9.3.458.
[32] C.J. Torrecilla-Salinas, O. De Troyer, M.J. Escalona, M. Mejías, A Delphi-based expert judgment method applied to the validation of a mature Agile framework for Web development projects, Inf. Technol. Manage. (2019), https://doi.org/10.1007/s10799-018-0290-7.
[33] W. Bouaynaya, Characterization of cloud computing reversibility as explored by the Delphi method, Inf. Syst. Front. 22 (6) (2020) 1505–1518, https://doi.org/10.1007/s10796-019-09947-5.
[34] C.C. Hsu, B.A. Sandford, The Delphi technique: making sense of consensus, Pract. Assess. Res. Eval. 12 (10) (2007).
[35] B. Kitchenham, Systematic literature reviews in software engineering, Inf. Softw. Technol. 51 (1) (2009) 7–15, https://doi.org/10.1016/j.infsof.2008.09.009.
[36] B. Ludwig, Predicting the future: have you considered using the Delphi methodology? J. Ext. 35 (5) (1997).
[37] M.A. Olivero, A. Bertolino, F.J. Dominguez-Mayo, M.J. Escalona, I. Matteucci, Addressing security properties in systems of systems: challenges and ideas, Lect. Notes Comput. Sci. (2019) 138–146, https://doi.org/10.1007/978-3-030-30856-8_10, vol. 11732 LNCS, no. WebistMay.
[38] C. Guariniello, D. DeLaurentis, Communications, information, and cyber security in systems-of-systems: assessing the impact of attacks through interdependency analysis, Proc. Comput. Sci. 28 (2014) 720–727, https://doi.org/10.1016/j.procs.2014.03.086, no. Cser.
[39] "How Apple and Amazon security flaws led to my epic hacking." https://www.wired.com/2012/08/apple-amazon-mat-honan-hacking/.
[40] Darktrace, "Global threat report - selected case studies," 2017.
[41] M. Darren, G. Paul, SPSS for Windows Step by Step: a Simple Guide and Reference, Pearson Education, Inc, Boston, 1999.
[42] LimeSurvey, "LimeSurvey." https://www.limesurvey.org/.
[43] F.H. Ferreira, E.Y. Nakagawa, R.P. dos Santos, Reliability in software-intensive systems: challenges, solutions, and future perspectives, in: 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2021, pp. 54–61, https://doi.org/10.1109/SEAA53835.2021.00016. Sep.
[44] K. Tsilipanos, I. Neokosmidis, D. Varoutas, A system of systems framework for the reliability assessment of telecommunications networks, IEEE Syst. J. 7 (1) (2013) 114–124, https://doi.org/10.1109/JSYST.2012.2207274.
[45] O.S. Ligaarden, K. Stølen, Analyzing security risks in critical infrastructures embedded in systems of systems: how to capture the impact of interdependencies, in: ESREL, Rhodes, Greece, CRC Press, 2010, pp. 347–353.
[46] E. Silva, T. Batista, Formal modeling systems-of-systems missions with mKAOS, in: Proceedings of the ACM Symposium on Applied Computing, 2018, pp. 1674–1679, https://doi.org/10.1145/3167132.3167311. Apr.

[47] F. Oquendo, "Formally describing the software architecture of systems-of-systems with SosADL," Aug. 2016. doi: 10.1109/SYSOSE.2016.7542926.
[48] V.V. Graciano Neto, R.P. dos Santos, D. Viana, R. Araujo, Towards a conceptual model to understand software ecosystems emerging from systems-of-information systems, in, Commun. Comput. Inf. Sci. 1081 (2020), https://doi.org/10.1007/978-3-030-46130-0_1. CCIS.
[49] F. Oquendo, Architecturally describing the emergent behavior of software-intensive system-of-systems with SosADL, in: 2017 12th System of Systems Engineering Conference (SoSE, 2017, pp. 1–6.