2013 IEEE 10th International Conference on e-Business Engineering

An Architecture for Justified Assessments of Service Provider Reputation

Nathan Griffiths
Department of Computer Science
University of Warwick
Coventry, CV4 7AL, UK
Email: nathan.griffiths@warwick.ac.uk

Simon Miles
Department of Informatics
King's College London
London, WCR2 2LS, UK
Email: simon.miles@kcl.ac.uk

***Abstract*—In a service-oriented system, an accurate assessment of reputation is essential for selecting between alternative providers. In many cases, providers have differing characteristics that must be considered alongside reliability, including their cost, experience, quality, and use of sub-providers, etc. Existing methods for reputation assessment are limited in terms of the extent to which the full interaction history and context is considered. While factors such as cost and quality might be considered, the assessment of reputation is typically based only on a combination of direct experience and recommendations from third parties, without considering the wider context. Furthermore, reputation is typically expressed as a simple numerical score or probability estimate with no rationale for the reasoning behind it, and there is no opportunity for the user to interrogate the assessment. Existing approaches exclude from consideration a wide range of information, about the context of providers' previous actions, that could give useful information to a user in selecting a service provider. For example, there may have been mitigating circumstances for past failures, or a provider may have changed their organisational affiliation. In this paper we argue that provenance records are a rich source of information on which a more nuanced reputation mechanism can be based. Specifically, the paper makes two main contributions: (i) we provide an analysis of the challenges and open research questions that must be addressed in achieving a rich provenance-based reputation mechanism, and (ii) we define an architecture in which the results of these challenges fit together with existing technologies to enable provenance-based reputation.**

I. Introduction

Service oriented technologies and cloud computing provide a natural way for small enterprises and individuals to quickly offer new services or content to a global customer base. These services may be purely computational, but will often be a software front-end to a physical service, such as booking a taxi or ordering some equipment. Software can be created and published that combines and supplements services provided by others, including resource services such as storage and hosting, to provide new functionality. By allowing the resources used to adapt as business needs change, enterprises can control their costs to a manageable level.

There is a massive growth in use of open APIs that are the building blocks for services, and the ability to use such technologies is essential for small enterprises as online commerce becomes key to most businesses. However, the potential provided by the technology does not necessarily mean that new services are adopted. Considered from the perspective of a customer, trying a new provider rather than an established name has a higher associated risk of receiving an inadequate service. Even if over time a provider is able to acquire customers who provide online reviews of the service, these can be superficial and a single negative review can disproportionately harm a recently started service. In particular, reviews will not take into account the composite nature of cloud-based services, so that failure of one service can lead to a subsequent loss of business by another service that relies on it, even if the latter stops using the former. Moreover, simply relying on strangers' reviews is inadequate for many customer needs, where failure of an important service has significant negative consequences. There are also issues regarding a lack of information and reviews for new services in a marketplace, meaning that a potential customer has insufficient information to support their decision making. While providing the technology to get started, the current form of cloud provision makes it difficult for small enterprises and individual entrepreneurs to succeed, especially in markets that already contain established providers.

A key challenge to supporting the publication and consumption of services is to define an appropriate reputation mechanism, which takes into account the complexity of real-world interactions. Existing reputation mechanisms assume that service quality, and which providers are responsible for it, are entirely transparent to the consumer on receiving the service, which is untrue where services are compositions. Moreover, existing reputation mechanisms do not account for the relationships between providers and consumers, the circumstances of a particular past success or failure, or organisational structures between providers. As a result, existing reputation mechanisms typically provide a coarse measure of a provider's reputation, often reduced to a numerical valuation, with no justification of *why* a particular reputation assessment is made. Reputation assessments rely on extensive records of distributed interaction, which data provenance and other technologies allow for, but such technologies are typically not yet used for recording business interactions. In this paper we propose a novel architecture for using provenance records as a base for a rich reputation framework, which will provide reputation information about service providers with a much richer history and justification than at present. We present the architecture itself, along with identifying the key research challenges that must be addressed for its realisation.

978-0-7695-5111-1/13 $26.00 © 2013 IEEE
DOI 10.1109/ICEBE.2013.53

II. BACKGROUND

In this section we introduce related work in the areas of service-oriented technologies and marketplaces, reputation, provenance, and reputation assessment based on provenance. Note that in the remainder of this paper we use the term *agent* to refer to individuals, organisations, producers and consumers interchangeably.

A. Service-oriented technologies

Service-Oriented Architectures (SOAs) are a paradigm to allow software developers to focus on the fulfilment of required enterprise functionalities at a conceptual level by providing standardised communication protocols, interfaces, and workflow and service management infrastructures/policies. SOAs enable developers to compose required services from existing ones, without being concerned by the barriers caused by heterogeneous operating and hardware systems, or syntax level differences among different software and locations [1].

The key property of SOAs is to allow services to publish their interfaces and related information, so that they become searchable and can be discovered over the Internet. Those services, having been discovered and selected, can be composed in certain ways to provide desired functions in order to meet specific application requirements [2]. The process of discovery and composition can be carried out at design and/or run time, so that new applications can be formed rapidly. Service discovery, selection and governance become critical processes for producing effective composite services, and so various service discovery and selection methods based on functional and non-functional specifications [3], [4], [5] have been proposed. However, research to include reputation in these methods is still in its infancy [6]. The focus of existing service-oriented governance approaches [7] is to control and implement intra-organisational services rather than inter-organisational ones, and so they lack mechanisms such as reputation assessment.

B. Service-oriented marketplaces and strategies

A service-oriented marketplace can be seen as a dynamic marketplace, where individuals interact to achieve their goals. The need for control mechanisms that ensure correctness and fairness of providers' and consumers' behaviour represents one of the most challenging issues in service-oriented marketplaces. The trust and decision strategies in such systems are widely studied by researchers from many domains covering interaction protocols design, social commitments, resource allocation heuristics and game theory.

Reputation and trust models are usually represented as interaction trust [8], encounter-derived reputation [9], subjective reputation [10] or the individual dimension of trust [11] and experience-based trust [12]. Smith and desJardins [13] created a formal network incorporating aspects of competence and integrity. Their work focuses on applying game-theoretic concepts to model and learn about other individuals, based on previous experiences. Trust and fairness in open distributed systems has been generally analysed in [14] including a description of the most common attacks and a classification of adversarial behaviour.

C. Reputation

Trust and reputation are concepts that originate in the social sciences, and are now commonly applied in a range of online computational systems, to improve the success of interactions by minimising uncertainty when self-interested individuals or organisations interact [15]. Trust is defined as an assessment of the likelihood that an individual or organisation will cooperate and fulfil its commitments [16]. Reputation is complementary to trust, and can be viewed as the public perception of the trustworthiness of a given entity [17]. In a service-oriented system individuals and organisations rely on providers to successfully execute services with an appropriate quality in order to fulfil their own goals, and such reliance implies a degree of risk, as success depends in part upon a third party. Trust and reputation provide an effective way of assessing and managing this risk.

Many computation models of trust and reputation have been developed and applied in a variety of settings including application specific domains such as service-oriented systems, Grid computing and P2P systems, and more generally in multi-agent systems (see [1], [17], [15] for extensive reviews of the main approaches). Most established trust models, such as ReGreT [18], FIRE [8], MDT-R [19] and TRAVOS [20] use a combination of direct experience and third party experiences as the base for assessing trust and reputation, and use numerical or probabilistic representations for trust [21].

The existing approaches to trust and reputation are effective in many situations, but do not function well if there is a lack of evidence to evaluate trust, such as where agents only participate for a short time or have a relatively small number of previous interactions [22]. Existing computation approaches to assessing trust and reputation are typically based on evaluating an individual's direct experiences supplemented by the opinions of others in the form of recommendations. The overall interaction histories of the agents involved are not fully considered, and so potentially relevant information is omitted, such as the reasons for previous failures, past alliances or affiliations, or changes in environmental state.

Relatively little work has considered how to evaluate trust when faced with insufficient evidence for existing methods, however, two approaches have recently been proposed. First, by observing interactions a role taxonomy can be evolved and agents assigned to roles, enabling individuals to estimate the expected behaviour and performance of others [23]. Second, a combination of monitoring and reputational incentives can be used to mitigate against a lack of evidence [22]. While these approaches are promising, they are not a general solution since they rely on observations and domain features that are not readily available in all applications.

In any situation where individuals' private data or experience is used to make assessments and observations, such as trust and reputation, there is a risk that specific information about an individual might be revealed against their wishes. For example, if guest feedback on a particular hotel is used to assess the hotel's reputation, the individuals who have stayed in the hotel may not wish to share this publicly. In the case where an explanation of such assessments is required this problem is exacerbated since it requires revealing details of the information on which the assessment was made. This issue

is largely unconsidered with respect to trust and reputation models (with a small number of exceptions such as [24]). However, the issue of privacy preservation has been explored in the data-mining community, by perturbing data so as to still allow valid patterns to be learnt from the perturbed data while hiding the real data values by reconstructing distributions at an aggregate level [25]. However, it has been shown that (partial) de-anonymisation is often possible when the data consists of multiple correlated attributes [26] or by pulling in other relevant data for instance from social media.

Trust and reputation enable agents to select appropriate service providers to minimise the risk of failure, from the possible providers on whom there is sufficient information. However, trust and reputation are inadequate in cases where there is insufficient information, such as where new providers or services are introduced to the system or where a service under consideration has had little recent use. The architecture proposed in this paper uses provenance records to provide a rich source of information on which to base trust and reputation assessment, and so utilises a broader range of evidence than existing approaches. The use of provenance-based reputation assessment enables a richer level of reasoning regarding reputation, for example through considering mitigating circumstances, changing alliances between agents, and the context of the previous interactions that inform reputation assessment.

D. Provenance

In order to make rich reputation decisions, data about what has occurred in the past is required. This cannot be simply a collection of logged events, but must express the causality between them, otherwise it is impossible to discern which agent's actions led to the success or failure of service provision. We need the facilities to record and later access the provenance of such results, i.e. how they have come into being through processes and interactions. Similarly, for an agent to understand why they have been ascribed a particular reputation, the reasons leading to the decision need to be recorded, stored and accessible. That is, we need the provenance of the decisions.

The amount of literature on provenance technologies has exploded in recent years, including surveys of the field [28], [29], [30]. Key issues regarding provenance of data in distributed systems (such as service marketplaces) include how to model the provenance so it interlinks records of what has occurred in disparate parts of the system, e.g. between the provider of a composite service and the providers of the component services that are used; how to record, store and query the provenance data; and how to adapt systems so that they record the information that will be of practical use later, making those systems *provenance-aware*.

Provenance is strongly related to lineage, audit, traceability, distributed logging, and similar ideas from many disciplines, but with an emphasis on causally connected records dispersed across distributed systems. Over the past decade, attempts have been made to unify the approaches taken in different research communities. The Provenance Challenges were international exercises in which teams from different organisations applied their own techniques to recording the provenance of results within the same case study, then attempted to integrate the resulting records. This led to the development of a common model for provenance, the Open Provenance Model (OPM) [31], already widely used in EU and other projects. The W3C then initiated efforts to produce a standard for provenance modelling and access on the web, drawing on OPM as well as many other relevant initiatives, such as Dublin Core. This effort concluded with a W3C recommendation, PROV [27], which specifies the model, its semantics, and its serialisation for semantic web applications, along with supporting specifications on HTTP access, XML serialisation, and so on.

To illustrate what provenance looks like in practice, consider the following example taken from the introductory primer for PROV [27]. Figure 1 shows PROV data in the form of a graph. Oval nodes denote *entities* that existed, such as items of data; rectangles denote *activities* that took place, using and generating entities; and, pentagons denote *agents* that are responsible for the activities having taken place. The nodes can be annotated with extra information, as is shown for the two agents. The edges denote causal relations pointing from effect to cause: that an activity used an entity in its processing, that an entity was generated by an activity; and that an activity was associated with an agent that was in some way responsible for it occurring. Reading from left to right along the top, the graph describes how a government data set on crime statistics was composed (aggregated) by region, following a list of places by region. This aggregated data was then used to generate a visual chart. An agent, Derek, was responsible for both of the two steps, and acted on behalf of his employer, Chart Generators Inc.

E. Reputation assessment from provenance records

Provenance records capture the information about interactions that is needed to provide a richer nuanced form of reputation, that is able to distinguish confidence and reputation along dimensions such as quality of results, validation, or certification, across complex processes [32]. There has been relatively little work considering the mechanics of how to assess trust based on the information that provenance records provide. Fundamentally, provenance records can be viewed as an extensive store of information that can be queried in order to assess trust. One of the earliest approaches to provenance-based reputation assessment was to use a decision tree, where each node represented a question that has a boolean answer, that was traversed with respect to the provenance records to obtain a trust measure [33], where trust itself takes the form of a standard probabilistic measure. To provide a richer assessment more information can be used by considering the provenance path of information, the trustworthiness of the information itself, and the reliability of the provider to assess reputation [34], [35]. Alternatively, a risk model can be defined that considers the main risk classes and relationships, which can facilitate a detailed risk assessment for an interaction by evaluating the complete provenance path [36]. Other artificial intelligence techniques such as case-based reasoning [37] have also be applied to estimate reputation based on the information contained in provenance records.

Reputation assessment can also be combined with machine learning on provenance records to support decision making where there is a lack of specific information about a particular provider. Any mechanism for assessing reputation requires

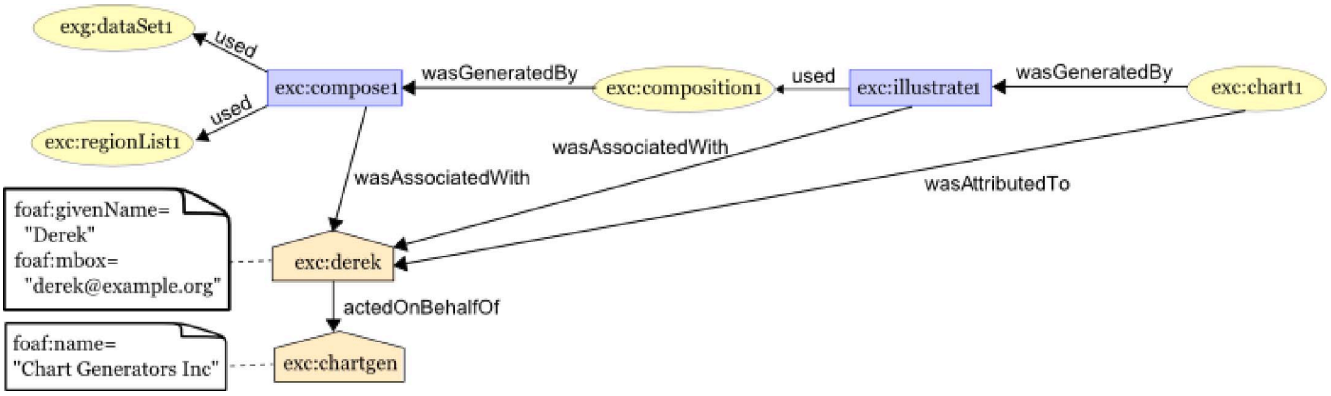

Fig. 1. An example PROV graph describing how some data was generated through steps for which some agents were responsible [27].

an extensive history of interactions, or a set of third party recommendations, on which to perform the assessment. In the absence of direct interactions, reputation assessment relies on third party recommendations, but traditional reputation mechanisms (as introduced above) do not consider more general features of an interaction. Provenance records contain a rich source of information from which features of interactions can be extracted, such as temporal or structural aspects, and these can be used in combination with reputation to train a machine learning algorithm to *predict* reputation based on provenance derived features [38]. In an online video tagging domain, using provenance-based estimates of reputation derived from attributes such as time, day, and typing duration, has been shown to be effective in cases where there was insufficient interaction history to use a traditional reputation mechanism (and in some instances give a small increase in performance) [38].

III. Existing Architecture for Reputation Assessment

Existing approaches to assessing reputation typically rely on the direct experiences of the agent making the assessment and on third party recommendations. Current reputation systems will generally follow the structure shown in Figure 2. A client agent will either request reputation information on a particular agent, or will ask for a recommendation on the most reputable agent with regard to a particular task. The request will be made to an *assessor*, an agent capable of quantifying reputation from the past history of agents. The assessor has access to a store of records of interactions between the agents over which reputation is assessed. This store will have been built up through monitoring of interactions as they occur. Each record provides information relevant to the reputation of all parties to one interaction. Each interaction record somehow indicates success or failure, possibly on a scale, and this is used by the assessor to create a reputation measure. If a recommendation has been requested, then multiple agents are assessed and the agent with highest reputation is recommended. Otherwise, the reputation measure is returned to the client.

Current approaches for assessing reputation have some significant limitations, that arise from the limited range of information that is considered in making an assessment. At a fundamental level, the problem is that there is no holistic view of interactions, and many aspects of the environment are not considered that have the potential to give valuable insight into reputation, such as the structure of organisations involved, the workflow in terms of sub-providers or aggregate services, the context of interactions, or any mitigating circumstances. As a consequence, existing reputation assessment mechanisms are relatively coarse, and do not account for the complex interactions and relationships that exist in real-world environments. By utilising provenance records as the source of information for assessing reputation a richer, more nuanced, view of reputation is possible.

In using an existing reputation assessment mechanism system designers and developers are often required to make assumptions regarding reputation and interactions that are not fully justified. For example, many approaches to reputation assessment consider recent interactions to be more important than older interactions, and their use requires the specification of a window size of past interactions to consider, or a decay function defining how to dampen the effect of older interactions. However, some older interactions may be significant, such as where a similar workflow was used, while recent interactions may be less applicable, for example if an unreliable sub-provider was used that has since been replaced. In a provenance-derived reputation system we can avoid making unjustified assumptions, by instead using provenance query templates to clearly define and extract the information that is relevant to a particular user in a given situation.

IV. Research Challenges

Given the above limitations, we now present a series of challenges we believe have to be met in order to provide richer, better justified assessments of service provider reputation, or recommendations based on such assessments. For each challenge, we provide a brief motivating discussion which indicates our own thinking on how the challenge is best addressed. This is expanded on further in the following section, where we present a revised architecture in which mechanisms to address the challenges are explicitly included.

A. Challenge 1: Basing reputation on rich historical data

The reputation of a provider is a judgement about how well they will provide a service at the present time. It is based on

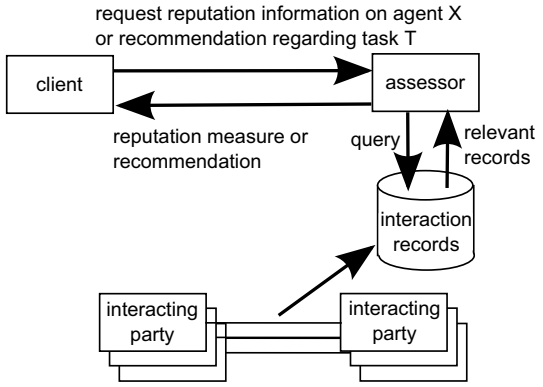

Fig. 2. The current approach to reputation assessment.

historical information, following the heuristic that the past is a good indicator of the present. However, historical data can come in different forms.

A record of interaction between two (or a few) parties may state what data was exchanged between those parties, in what sequence, and possibly what else could be observed in the environment in fulfilling a service. The quality of the service may be judged by indicators within the exchanged data, the timing of events, or in the environment. For example, the messages may indicate the price ultimately paid for the service or the time at which the service was delivered. The interaction record may be more abstract than this, e.g. customer feedback on the product ultimately received, its timeliness and cost.

However, as the challenges below will describe, such records miss much information that is relevant to determining reputation. The first key obstacle, then, is to capture richer records from which to judge reputation. Such records should not assume that every interaction occurs in isolation, e.g. one service provider is often dependent on another, a client may use a particular service because it is compatible with a product it previously bought from another provider, etc.

A provenance model, such as W3C's PROV, assumes interdependence between past events: any activity, entity or agent can be related to another. This allows cause and effect, indirect responsibility, original sources of data, and more to be modelled. These aspects are essential for meeting the other challenges below.

In order to capture provenance data, just as with interaction records, appropriate monitoring has to be applied. There is existing research on adapting service-oriented systems to record provenance [39], and methodologies for determining what in practice should be recorded [40]. However, in order to use these, one needs to know what information the provenance should contain.

The challenge is to adapt service-oriented systems to capture records that account for interdependence between interactions and agents, and so can be used in generating rich reputation assessments.

B. Challenge 2: Situation-based assessments

Some existing reputation models account for different types of service and the situations in which they were provided, to understand which past interactions are relevant to consider. However, the idea of a situation is defined only abstractly and as if each situation is wholly distinct from others.

In practice, a service will be used in situations that are more or less similar to each other in particular ways. For example, an auction service might hold an auction in which many agents participate or only a few. A hosting service may host an application with many dependencies on other services to perform its function, or few or none. Combining the two examples, if an auction service is the application being hosted, then all combinations of the two variations of situation apply. These situations are not distinct, just more or less similar.

If we have a rich description of the context in which services were provided, i.e. a provenance graph in which the service provision is part, then we are in a position to extract this situation description and judge to what degree the situation of a past service provision matches a new situation, and so provide a reputation assessment appropriate to this situation.

The challenge is to allow reputation assessment to take into account the details of the situation in which past interactions took place in providing an assessment appropriate for the current client's situation.

C. Challenge 3: Providing rationales

When a client asks for a reputation assessment or recommendation, they are doing so as part of processes, with particular aims, and in particular contexts. As discussed above, some of these aspects may be made apparent to the assessor, so that it can tailor the reputation assessment to a particular situation. However, much may be infeasible or undesirable to communicate to the assessor.

Instead, it would be preferable for the client to understand why it has received the assessments it has: their *rationales*. In this way, it can judge whether or to what extent the assessment is appropriate to its needs. The rationale may be processed automatically by a sufficiently sophisticated service or else presented to a user via a user interface, but in either case it needs to be structured.

The rationale behind a reputation assessment can be divided into two parts. First, there is the evidence on which the reputation assessment is based, e.g. provenance records. Second, there is the process that the assessor has conducted in order to derive the assessment from the records. A record of the latter process is itself provenance, the provenance of the assessment, and can be expressed using the same provenance model and captured with the same mechanisms.

The challenge is to supplement reputation assessments and recommendations with structured rationales that allow a client to reason about how to make use of the assessments.

D. Challenge 4: Accounting for mitigating circumstances

The successfulness of providing a service is an indicator of how well a provider may perform in general, including in the future, but it is not the same thing. Inferring future success and failure (the value of a reputation assessment) from past success and failure is an approximation that may be misleading in certain circumstances. Sometimes there are specific mitigating

circumstances which explain the failure of service provision and, as such, not all failures should be treated equally when judging reputation.

For example, the failure to deliver some goods on time on a day when an unexpected transport strike occurs may be considered a failure under mitigating circumstances. We may decide that the provider should have found an alternate way to deliver the goods, and so still consider some negative effect on their reputation to be appropriate, but probably not as much of a negative effect as when failing to deliver on time without the strike taking place.

Causes of failure may be direct or indirect, and may or may not be the responsibility of the provider. Comparably, success may sometimes be due more to luck than competence. An adequately rich representation of the past allows a reputation assessor to judge whether mitigating circumstances apply, and so take them into account in the assessment.

The challenge is to take mitigating circumstances affecting the successfulness of service provision, i.e. direct or indirect causes out of the provider's control, into account when assessing reputation.

E. Challenge 5: Accounting for joint responsibility

Following on from the above challenge, another way to consider how success and failure should affect reputation in a nuanced way is to examine which parties were responsible for what outcomes and to what degrees. Where one service provider relies on a second provider to provide part of their composite service, then failures of the latter can cause failures of the former, even if the first provider's provision is excellent in all other regards. If the first provider later abandons the second provider in favour of an alternative, how should the reputation of the first provider now be judged? If an individual working on behalf of an organisation provides a good service, to what degree should this be judged to be a positive reflection on the individual or on the organisation? If the individual moves to another organisation, what now are the reputations of the individual and the original organisation?

Unlike mitigating circumstances, as described above, joint responsibility often occurs over time. With an adequately rich account of past service provision, the joint responsibility should be evident, and we may be able to make some estimate of the degree to which success or failure is due to particular parties.

The challenge is to take account of the joint responsibility of multiple parties in past interactions when assessing reputation.

F. Challenge 6: Tailoring assessment to personal need

Each client has its own preferences, assumptions, and ongoing goals etc. These will affect how they interpret the reputations and rationales and choose which provider to use. As argued above, if clients express the situation in which they require a service, this could be accounted for in the reputation assessment, while providing rationales allow sufficiently sophisticated clients (or their users) to interpret reputation assessments according to these personal requirements. However, both of these methods require clients to be aware of and be able to express (and possibly reason over) their personal requirements. This will be infeasible in some applications, especially when a user would be required to articulate all their preferences in an ambiguous way to their client.

In such cases, we need to automatically capture data from which to personalise reputation assessments. Specifically, we can record what a client does following receipt of a reputation assessment or recommendation; most obviously, what service do they then use. This is a record of past interaction as with any other. The provenance of that service choice is based partly on the reputation assessment that the client received, but also on the client's preferences. Therefore, a record of the client's choices informs the assessor about how the client is interpreting the assessments given and what other factors may be affecting choice.

By taking this into account (learning this aspect of the client's behaviour), the assessor could give assessments that are more useful for the client's needs. For example, if particular features of providers' assessments are observed as having a likely effect on the client's choice, this suggests that these features are important for the client. The reputation of a provider can then be weighted by the fact that such features are more significant than others.

The challenge is to allow the effects of reputation assessment on an individual client to be learnt, so that future assessments can be tailored to those aspects the client treats as most important.

G. Challenge 7: Long term assessment responsibility

Many recommender, reputation assessment and service selection mechanisms assume that, first, the service assessments/selections provided will be used immediately and, second, the execution of the service will be practically instantaneous when compared with the rate of change of the rest of the environment. In practice, neither of these assumptions may hold. In particular, services are sometimes long term jobs or subscriptions, and the fact that a client has started using a particular service does not mean it necessarily needs to continue with that service for the full duration. The earlier it knows that it may be appropriate to abandon its current request or change service provider for a subscription, the better.

An assessment of a service given at one time may not be the same as the assessment given at a later point in time, because the records on which the assessment is based will change over time as the service is used and feedback is accumulated, etc. Therefore, if a reputation assessor aims to give accurate assessments then, for long term services, a single assessment may not be sufficient. Instead, the assessor would ideally update its previously given assessments over time, whenever they change significantly, and inform the relevant clients that the assessment has changed.

The challenge is for assessments to be long term, so matching the use of some services, and be updated as new relevant information is discovered.

V. AN ARCHITECTURE FOR PROVENANCE-BASED REPUTATION ASSESSMENT

To meet the challenges, we propose to extend the structure in Figure 2 to the more extensive architecture shown in Fig-

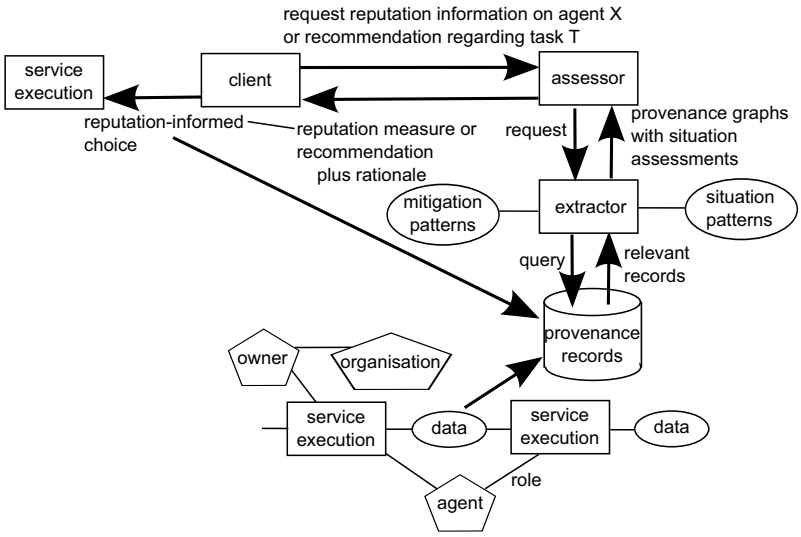

Fig. 3. A proposed architecture for provenance-based service provider reputation, with mechanisms to address each challenge identified in this paper.

ure 3. Here, the client makes the same requests to an assessor for reputation assessment or a recommendation. However, the records the assessor now relies on are rich interconnected provenance graphs rather than a list of independent interactions. The provenance records are recorded as a side-effect of the interaction of agents, by one or multiple parties in the interactions (Challenge 1).

This allows mitigation, situation, indirect responsibility, and other such context to be accounted for, and for the interdependencies of providers to be understood. In practice, it is helpful to distinguish the tasks of the assessor and that of an extractor, which will look for relevant patterns in the provenance that indicate relevant situations (Challenge 2) and mitigating circumstances (Challenge 4). Provenance data is suitable for this because it includes the causal connections between interactions, and so captures the dependencies between agents' actions. It can also include multiple parties to an interaction and their organisational connections, as illustrated by the "owner", "organisation" and "agent" provenance graph nodes in the figure, so allowing joint responsibility to be accounted for (Challenge 5).

The effect of the extractor is to filter the provenance for key subgraphs from which reputation can be directly assessed. The assessor can make assessments as before, looking for success and failure, and adjusting these by mitigation and situation relevance. The subgraphs give a description of what has occurred, allowing the reputation measure or recommendation returned to the client to be accompanied by a rationale (Challenge 3).

Further, we aim to personalise reputation assessments. Clients will decide which service to interact with based on the reputation or recommendation, plus rationale, given, and on their own preferences and reasoning. The interaction due to this choice, and its link back to the assessment provided by the assessor, is itself included as part of the provenance records. This helps the extractor then distinguish which situations are actually relevant to the particular client (Challenge 6). Finally, for a single request from the client, a series of requests are sent by the assessor to the extractor over time, to ensure that reputation assessments are updated based on further data accumulated (Challenge 7). This is not a change in the architecture but rather a change in the assessor's behaviour, so not shown on the figure.

VI. Conclusion

Accurate assessments of reputation are essential for selecting appropriate service providers, and yet existing approaches to reputation are limited in terms of the extent of the information considered, the incorporation of contextual aspects such as mitigating circumstances, and the provision of a rationale for reputation assessment to support interrogation by a user. Provenance provides a potential solution, by enabling a richer, more nuanced, type of reputation assessment to be performed, that considers all relevant information and provides users with a rationale that explains how the assessment was arrived at. The foundational technologies already exist to support provenance-derived reputation, such as the Open Provenance Model (OPM) [31], and the W3C PROV specification [27]. However, there are a number of research challenges that must be addressed before such an approach to reputation can be realised. In this paper, we have identified these challenges and proposed an architecture for provenance-driven reputation that defines how their results interrelate. Specifically, we have (i) provided an analysis of the challenges and open research questions that must be addressed in achieving a rich provenance-based reputation mechanism, and (ii) defined an architecture in which the results of these challenges fit together with existing technologies to enable provenance-based reputation. Our aim is for the architecture and open research questions to inspire researchers in provenance, reputation and service-oriented computing to address the challenges and help realise a rich provenance-based reputation framework.

References

[1] N. Griffiths and K.-M. Chao, Eds., *Agent-Based Service-Oriented Computing*. Springer, 2010.

[2] M. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-oriented computing: State of the art and research challenges," *IEEE Computer*, vol. 40, no. 11, pp. 38–45, 2007.

[3] L. Zeng, B. Benatallah, A. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "QoS-aware middleware for web services composition," *IEEE Transactions on Software Engineering*, vol. 30, no. 5, pp. 311–327, 2004.

[4] W.-L. Lin, C.-C. Lo, K.-M. Chao, and M. Younas, "Consumer-centric QoS-aware selection of web services," *Journal of Computer and System Sciences*, vol. 74, no. 2, pp. 211–231, 2008.

[5] S. Dietze, A. Gugliotta, J. Domingue, H. Yu, and M. Mrissa, "An automated approach to semantic web services mediation," *Service Oriented Computing and Applications*, vol. 4, no. 4, pp. 261–275, 2010.

[6] S. Lim Choi Keung and N. Griffiths, "Trust and reputation," in *Agent-Based Service-Oriented Computing*, N. Griffiths and K.-M. Chao, Eds. Springer, 2010, pp. 189–224.

[7] IBM, "Effective SOA governance: Technical white paper," IBM, Tech. Rep., 2013.

[8] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt, "An integrated trust and reputation model for open multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 13, no. 2, pp. 119–154, 2006.

[9] L. Mui, M. Mohtashemi, and A. Halberstadt, "Notions of reputation in multi-agents systems: a review," in *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, 2002, pp. 280–287.

[10] J. Sabater and C. Sierra, "ReGreT: A reputation model for gregarious societies," in *Proceedings of the 4th Workshop on Deception Fraud and Trust in Agent Societies*, 2001, pp. 61–70.

[11] ——, "Reputation and social network analysis in multi-agent systems," in *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, 2002, pp. 475–482.

[12] K. Fullam and K. Barber, "Learning trust strategies in reputation exchange networks," in *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, 2006, pp. 1241–1248.

[13] M. Smith and M. desJardins, "Learning to trust in the competence and commitment of agents," *Autonomous Agents and Multi-Agent Systems*, vol. 18, no. 1, pp. 7–11, 2009.

[14] A. Wierzbicki, *Trust and Fairness in Open, Distributed Systems*. Springer, 2010.

[15] J. Sabater and C. Sierra, "Review on computations trust and reputation models," *Artificial Intelligence Review*, vol. 24, no. 1, pp. 33–60, 2005.

[16] D. Gambetta, "Can we trust trust?" in *Trust: Making and Breaking Cooperative Relations*, D. Gambetta, Ed., 1988, pp. 213–237.

[17] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.

[18] J. Sabater and C. Sierra, "Evaluating the ReGreT system," *Applied Artificial Intelligence*, vol. 18, no. 9-10, pp. 797–813, 2004.

[19] N. Griffiths, "Enhancing peer-to-peer collaboration using trust," *Expert Systems with Applications*, vol. 31, no. 4, pp. 849–858, 2006.

[20] W. Teacy, J. Patel, N. Jennings, and M. Luck, "TRAVOS: Trust and reputation in the context of inaccurate information sources," *Autonomous Agents and Multi-Agent Systems*, vol. 12, no. 2, pp. 183–198, 2006.

[21] Y. Wang and M. Singh, "Formal trust model for multiagent systems," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007, pp. 1551–1556.

[22] C. Burnett, T. Norman, and K. Sycara, "Trust decision-making in multi-agent systems," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, 2011, pp. 115–120.

[23] R. Hermoso, H. Billhardt, and S. Ossowski, "Role evolution in multi-agent systems as an information source for trust," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, 2010, pp. 217–224.

[24] E. Gudes, N. Gal-Oz, and A. Grubshtein, "Methods for computing trust and reputation while preserving privacy," in *Proceedings of the 23rd Annual IFIP WG 11.3 Working Conference on Data and Applications Security XXIII*, 2009, pp. 291–298.

[25] D. Agrawal and C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," in *Proceedings of the 20th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2001, pp. 247–255.

[26] H. Kargupta and S. Datta, "On the privacy preserving properties of random data perturbation techniques," in *Proceedings of the 3rd IEEE International Conference on Data Mining*, 2003, pp. 66–106.

[27] W3C, "PROV model primer," http://www.w3.org/TR/prov-primer/, 2013.

[28] R. Bose and J. Frew, "Lineage retrieval for scientific data processing: A survey," *ACM Computing Surveys*, vol. 37, no. 1, pp. 1–28, Mar. 2005.

[29] L. Moreau, "The Foundations for Provenance on the Web," *Foundations and Trends in Web Science*, vol. 2, no. 2–3, pp. 99–241, Nov. 2010.

[30] Y. Simmhan, B. Plale, and D. Gannon, "A survey of data provenance in e-science," *SIGMOD Record*, vol. 34, no. 3, pp. 31–36, 2005.

[31] L. Moreau, B. Clifford, J. Freire, J. Futrelle, Y. Gil, P. Groth, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, B. Plale, Y. Simmhan, E. Stephan, and J. Van den Bussche, "The Open Provenance Model core specification (v1.1)," *Future Generation Computer Systems*, vol. 27, no. 6, pp. 743–756, Jun. 2011.

[32] L. Moreau, S. Chapman, A. Schreiber, R. Hempel, O. Rana, L. Varga, U. Cortes, and S. Willmott, "Provenance-based trust for grid computing: Position paper," University of Southampton, Tech. Rep., 2004.

[33] S. Rajbhandari, A. Contes, O. Rana, V. Deora, and I. Wootten, "Trust assessment using provenance in service oriented applications," in *Proceedings of the 10th IEEE International Enterprise Distributed Object Computing Conference Workshops*, 2006, p. 65.

[34] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *Proceedings of the 5th VLDB workshop on Secure Data Management*, 2008, pp. 82–98.

[35] X. Wang, K. Govindan, and P. Mohapatra, "Provenance-based information trustworthiness evaluation in multi-hop networks," in *Proceedings of the IEEE Global Telecommunications Conference*, 2010, pp. 1–5.

[36] P. Townend, D. Webster, C. Venters, V. Dimitrova, K. Djemame, L. Lau, J. Xu, S. Fores, V. Viduto, C. Dibsdale, N. Taylor, J. Austin, J. Mcavoy, and S. Hobson, "Personalised provenance reasoning models and risk assessment in business systems: A case study," in *Proceedings of the 7th IEEE International Symposium on Service Oriented System Engineering*, 2013, pp. 329–334.

[37] Q. Bai, X. Su, Q. Liu, A. Terhorst, M. Zhang, and Y. Mu, "Case-based trust evaluation from provenance information," in *Proceedings of the International Joint Conference on Trust, Security and Privacy in Computing and Communications*, 2011, pp. 336–343.

[38] D. Ceolin, P. Groth, W. van Hage, N. A., and W. Fokkink, "Trust evaluation through user reputation and provenance analysis," in *Proceedings of the 8th Workshop on Uncertainty Reasoning for the Semantic Web*, 2012, pp. 15–26.

[39] P. Groth and L. Moreau, "Recording process documentation for provenance," *IEEE Transactions on Parallel and Distributed Systems*, 2009. [Online]. Available: http://www.ecs.soton.ac.uk/ lavm/papers/tpds09.pdf

[40] S. Miles, P. Groth, S. Munroe, and L. Moreau, "PrIMe: A Methodology for Developing Provenance-Aware Applications," *ACM Transactions on Software Engineering and Methodology*, vol. 20, no. 3, pp. 1–42, Aug. 2011.