Systematic Evaluation of Hyperscale Cloud Providers Trustworthiness for Hosting Telco Workloads

Rambod Pakrooh*, Carol Fung*, Luis Suárez†, Abdollah Jabbari*
*CIISE Department, Concordia University, Montreal, Canada
rambod.pakrooh@mail.concordia.ca, {carol.fung, abdollah.jabbari}@concordia.ca
†Ericsson Research Security, Montreal, Canada
luis.suarez@ericsson.com

***Abstract*—Telecommunications companies (telcos) are currently transitioning critical workloads from private infrastructure to Hyperscale Cloud Providers (HCPs). Leveraging managed environments from providers such as Amazon, Google, and Microsoft enables substantial reductions in both CapEx and OpEx. However, making this decision is challenging as telcos find it difficult to trust these providers in hosting and protecting their valuable assets. Despite the extensive research on cloud providers trustworthiness, there are still some major challenges that need to be addressed.**

In this paper, we propose a novel framework to help telcos analyze the trustworthiness of different providers systematically and make informed decisions. This framework relies on public information to assess the trustworthiness of providers before making contracts considering the trustworthiness of information sources. We use Dirichlet distribution to model our belief regarding the value of each metric in the absence of fully trusted sources. This framework can be easily extended to include evidence from direct measurement, after making contracts, to update the metric values. Furthermore, we propose an adapted algorithm to update sources trustworthiness over time based on their provided pieces of evidence. Metrics expected values along with their confidence scores are combined using weighted geometric mean in each dimension and across dimensions to get the total trust score for each HCP.

We implement a testbed using Python on top of a virtual network created in the OpenStack cloud computing infrastructure. The experimental results demonstrate the effectiveness of the proposed framework in comparing major HCPs. The resulting trust scores are influenced by both the set of evidence sources involved and the rate at which source trustworthiness evolves over time. To the best of our knowledge, this work is the first granular framework for assessing the trustworthiness of HCPs based on public information.

***Index Terms*—5G, telecommunication, security, trust, cloud providers**

I. Introduction

Driven by advancements in 5G and beyond, telecommunication companies (telcos) increasingly rely on IaaS and CaaS providers to deploy their Virtualized and Containerized Network Functions (VNFs/CNFs). Hyperscale Cloud Providers (HCPs), such as Amazon AWS, Microsoft Azure, and Google Cloud Platform (GCP), are the main players in offering IaaS/CaaS at large scales. These HCPs began their journey as traditional cloud service providers, offering economies of scale and elasticity to the enterprise market. However, their focus is shifting increasingly toward providing faster services, higher capacities, and massive scalability which are of high importance to mission-critical service providers such as telcos [1]. Possessing geographically-distributed large data centers with an abundant amount of compute, storage, and network resources, these providers can host various telco-grade workloads with differing levels of resource requirements.

NFVI (Network Function Virtualization Infrastructure) was originally defined as the set of physical compute, storage, and network resources, as well as the virtualization software (hypervisor) which is used to create virtual environments to host VNFs [2]. In recent years, with the gradual emergence of novel virtualization technologies, NFVI has been expanded to support these new forms of virtualization, specifically containerization [3]. In this sense, HCPs can act as NFVI providers helping telcos in reducing their OpEx (Operating Expenses) and CapEx (Capital Expenditure) by allowing them to avoid the complexities of setting up and managing private clouds. The challenge lies in how telcos can trust these providers in hosting and protecting the security of their data and workloads on the cloud.

Trust is a general concept which lays the foundation for decision making in various domains of human knowledge. According to [4], "trust is a relationship in which an entity, often called the trustor, relies on someone or something, called the trustee, based on a given criterion". According to ETSI [5], "Trust is defined as confidence in the integrity of an entity for reliance on that entity to fulfil specific responsibilities"; they point out the dynamicity of trust and how different measures including identity, attribution, attestation, and non-repudiation can assist in deciding when and how to rely on a relationship or transaction.

Before telcos deploy mission-critical functions on an HCP's

This work is supported by Ericsson and National Cybersecurity Consortium (NCC).

ISBN 978-3-903176-82-9

NFVI, the provider must demonstrate its reliability and trustworthiness. Specifically, the HCP must prove it can secure these assets against threats like malicious hypervisors and co-resident VNFs [2]. Multiple research works in the literature [1], [6]–[22] have proposed approaches for assessing the trustworthiness of cloud providers. These approaches primarily rely on one or more of the following sources: cloud providers' responses to CSA (Cloud Security Alliance) Consensus Assessment Initiative Questionnaire (CAIQ), SLA compliance monitoring, QoS measurements, and users reviews. Most existing solutions either assume a certain level of cooperation from cloud providers in disclosing relevant information or perform trust assessment only after a contractual relationship has been established.

Despite the extensive research on cloud providers trustworthiness [1], [6]–[22], to the best of our knowledge, none of the existing approaches are suitable for granular trust assessment of HCPs before entering binding contracts. To achieve a practical framework for this purpose, we need to address some major challenges. First, due to privacy concerns, HCPs may choose not to share details about their infrastructure to their potential customers. Existing approaches are either assume direct cooperation from cloud providers or calculate trustworthiness only based on reputation. Second, there are always sources that intentionally/unintentionally provide false information. A practical framework should take into account each source's trustworthiness. Third, each telco has its own local policy which may result in different requirements and preferences that should be met by a potential provider. A configurable framework which allows telcos to enforce their policy in the assessment process would be preferable.

To overcome the aforementioned challenges, in this paper, we propose a novel configurable framework to assess the trustworthiness of HCPs based on publicly accessible data sources such as HCPs documentation, CSA STAR registry [23], and news websites such as The Register [24]. In this framework, the Dirichlet distribution is used to model our belief regarding the value of each metric. Each metric's associated distribution is updated upon receiving evidence from public sources, taking into account the trustworthiness of those sources. The total trust score for each HCP is calculated by combining the expected values and confidence scores calculated for all metrics according to their posterior distributions. Furthermore, sources' trust values are updated according to their historical behavior in telling the truth. Simulation results demonstrate the influence of the involved sources and the rate at which their trustworthiness changes on the final HCPs trust scores. Our contributions are summarized as follows:

1) We design a configurable framework for evaluating the trustworthiness of candidate HCPs based on publicly available information. This framework allows telcos to do a fine-grained analysis following their local policies.
2) We use Bayesian updating based on Dirichlet distribution to update our belief regarding the metrics values upon receiving new evidence. Evidence trustworthiness is taken into account during the whole process. Furthermore, each metric is characterized by an expected value along with a confidence score calculated based on the posterior Dirichlet distribution. An approach is proposed to integrate metrics expected values and their confidence score in each dimension and across dimensions to come up with a trust value for the HCP.
3) Finally, we evaluate the feasibility of the proposed solution by implementing it on a testbed built on top of the OpenStack platform.

The rest of this paper is organized as follows; in Section II, related works are summarized; in Section III, our proposed solution is explained in detail; in Section IV, the results from feasibility check and implementation are elaborated; finally, in Section V, concluding remarks are provided.

II. Related Works

In the past decade, there have been multiple research projects on assessing the trustworthiness of cloud providers.

Habib et al. [6] present a trust management framework having the multi-faceted nature of trust and user preferences in mind. The authors are mainly focused on CSA CAIQ as the source of trust information. A mediation-based trust assessment framework for assessing the trustworthiness of infrastructure providers is presented in [7]. The authors argue that a trust model should be contextualized and introduce a cloud broker acting at the mediation layer in different modes according to the deployment context. Manuel [8] presents a very simple trust model based on only four QoS parameters considering provider's past performance and current capabilities.

In an interesting study, Fan et al. [9] present a multi-dimensional trust assessment of cloud providers based on other users' ratings and direct measurements of an active user. SelCSP [10] is a pre-outsourcing framework that estimates the interaction risk of selecting a cloud provider. It combines reputation-based trust with an SLA-transparency assessment of provider competence to rank providers for a given interaction context. A middleware for trust evaluation of cloud providers is designed by Tang et al. [11]. This middleware integrates objective trust through QoS monitoring and subjective trust based on users' ratings, weighted by confidence scores and a tunable parameter λ, to obtain a final trust value. In another project [12], the authors implement a trust label system as a medium for cloud providers to communicate transparently some trust information, including service runtime status, contract adherence, and data current location, with their users in order to foster trust. Gonzales et al. [13] propose a reference model for cloud infrastructure as well as a security assessment framework named Cloud-Trust.

Algamdi et al. [14] design a trust management system using CSA's CTP protocol (Cloud Trust Protocol), CAIQ, and users' feedback as primary sources of trust information. In another study, Sidhu and Singh [15] present a trust evaluation framework based on SLA compliance monitoring performed by a trustable cloud auditor. In [16], the authors examine the effect of Cloud Trust Labels (CTLs) on users perception of

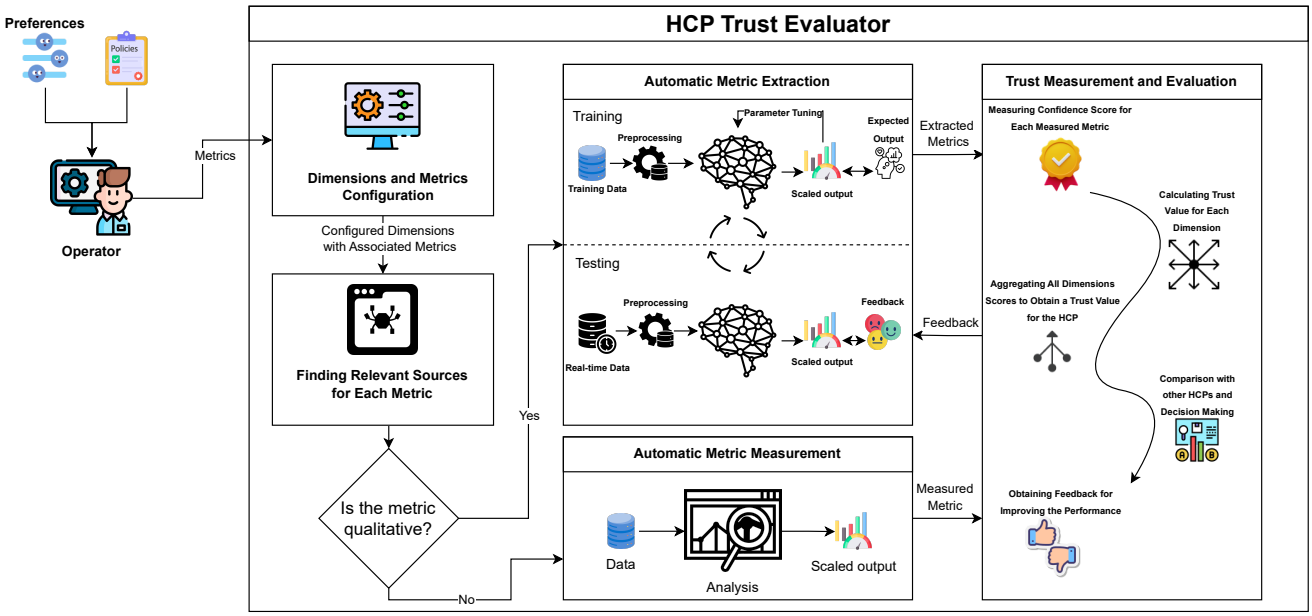


Fig. 1. Proposed Framework Architecture.

cloud service as well as cloud provider trustworthiness. A cloud trust evaluation system based on fuzzy logic is designed by Rizvi et al. [17]. This system takes as input users subjective opinions regarding high-level security factors, and gives a quantitative security index as output.

The are several works on cloud trustworthiness published in the past few years. Brumă [18] presents a trust evaluation framework based on standard tools offered by CSA i.e. CCM (Cloud Control Matrix) and CAIQ-Lite, and AICPA's (the American Institute of Certified Public Accountants) Trust Services Criteria (TSC). Junejo et al. [19] propose a framework to measure trustworthiness by combining multi-dimensional objective evidence (runtime QoS monitoring) and subjective evidence (users' feedback). In another study, Balcão-Filho et al. [20] design a consumer-centric framework for trust evaluation without requiring deep security expertise. Furthermore, John and Singh K [21] propose a framework based on digital twin technology and fuzzy inference method to derive trust scores from operational and security parameters.

There are few works addressing trust in cloud providers in the 5G context in some way. Jorquera Valero et al. [22] propose a trust and reputation management framework for distributed 5G marketplaces. This framework measures trust based on satisfaction, credibility, transaction context, and community feedback and integrates an SLA-based reward and punishment mechanism to update the trust value based on breach predictions, breach detections, and SLA violations. In another interesting study, Bouakkaz et al. [1] come up with an intelligent and comprehensive trust management framework to promote secure and trustworthy relationship among various 5G stakeholders including HCPs, Mobile Network Operators (MNOs), and Communication Service Providers (CSPs). This framework is composed of four modules: a fuzzy logic-based Trustworthiness Assessment module to gather and analyze trust evidence, a Trust Assessment module based on Reinforcement Learning (RL) and game theory to dynamically predict and adapt stakeholder trustworthiness, a Trust Lifecycle Management module to establish, maintain, and revoke trust relationships, and finally, a Trust Decision-Making module to update trust according to positive and negative outcomes.

As far as we know, there is no proposed approach in the literature for granular trust assessment of HCPs based on public information. These works either assume direct cooperation from HCPs or calculate trust based on reputation and user reviews. Furthermore, relying primarily on CAIQ limits applicability when providers do not disclose sufficient questionnaire evidence. A practical framework should rely on as much information as available over time to measure and update trust. Moreover, source trustworthiness in telling the truth should be considered while trust values are updated to prevent malicious sources from affecting the final trust scores. Furthermore, each user may have their own set of policies and requirements for trusting a provider. An open framework allowing users to enforce their local policies in the trust evaluation process is still missing in the literature.

In our proposed framework, trust is evaluated using metrics defined across multiple dimensions, relying solely on publicly accessible information. Furthermore, we account for the trustworthiness of evidence sources when updating our belief regarding each metric value. The trust value of each source is dynamic and continuously updated based on its historical behavior. In the next section, we elaborate on the proposed framework in greater detail.

III. Trust Evaluation Framework

In this section, we propose a trust evaluation framework to assist telcos in assessing the trustworthiness of HCPs, using publicly available data. By using this framework, we overcome the limitations mentioned in the previous section.

A. Architecture Overview

The architecture of the proposed framework is shown in Fig. 1. As demonstrated in this figure, this framework should be accompanied by a web crawler and an LLM model. The web crawler is responsible for finding relevant sources, while the LLM model is used to extract metric values from unstructured data provided by these sources. For the sake of simplicity, we use the available AI products e.g. OpenAI's ChatGPT and Google's Gemini. Each product is already integrated with a web crawler and an LLM engine that can be used to extract metric values from public sources.

B. Trust Dimensions and Metrics

First, we need to identify the metrics that we can measure before making a contract. We have identified eight dimensions to measure the trustworthiness from different perspectives (Table I). Although these dimensions are mostly related to security, they can be extended to include other domains such as financial and environmental domains. We believe telcos should be in charge of dimensions and metrics configuration to feel completely in control.

C. Trust Model

Before diving into details, we need to make some assumptions to simplify the problem:

1) Telcos are responsible for defining the trust dimensions and associated metrics according to their interests and risk appetite.
2) Telcos assign a weight to each dimension/metric according to their preferences and local policies. The weights assigned to each metric in each dimension should sum to 1. Moreover, the sum of dimensions weights is also equal to 1.
3) Metrics are measured, on a scale from 0 to 1, based on evidence from different, independent sources such as third-party registries, CVE databases, and provider's documentation.
4) Each piece of evidence is as trustworthy as its source provider.

Since before making a contract, there is no direct access to the infrastructure, most of the metrics that can be measured are qualitative in nature. Although some telcos may prefer to include some quantitative metrics based on indirect measurement, in this paper, we decided to focus on qualitative metrics as they are more challenging to include in numerical trust calculation. These qualitative metrics are used to characterize the HCP, based on which we can evaluate the trustworthiness. A telco should define two or more possible values for each metric and map these values to unique scores in the range [0,1]. Therefore, each metric m is associated with a set of possible values/outcomes $O^{(m)}$; the value of k may change from a metric to another.

$$O^{(m)} = \{o_1^{(m)}, o_2^{(m)}, ..., o_k^{(m)}\} \\ \underset{1 \le i \in Z^+ \le k}{o_i^{(m)}} \in [0,1]; \quad o_i^{(m)} \neq \underset{i \neq j}{o_j^{(m)}} \tag{1}$$

If we receive evidence from a fully trusted source, we use the value extracted from this piece of evidence as the final value for the metric with a confidence score (trust in the correctness of value) of 1 ($c_m = 1$); otherwise, we resort to the Dirichlet distribution to calculate the metric value along with its confidence score. Fung et al. [25] were among the first to use the Dirichlet distribution along with Bayesian statistics to deal with uncertainty in trust management. The multivariate probability vector associated with an experiment with k possible outcomes can be modeled using a Dirichlet distribution with k parameters (2). These parameters are sometimes called pseudocounts as they are proportionate to the number of observations associated with each outcome. As a rule of thumb, each parameter equals the number of observations plus 1. A Dirichlet distribution with all parameters set to 1 (no evidence) is equivalent to a uniform distribution where all outcomes are equally likely.

$$P = (p_1, p_2, ..., p_k) \sim Dir(\alpha_1, \alpha_2, ..., \alpha_k);\ \alpha_i \geq 1 \\ \alpha_0 = \sum_{i=1}^{k} \alpha_i \qquad E[p_i] = \frac{\alpha_i}{\alpha_0} \\ Var(p_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)} \qquad \underset{i \neq j}{Cov(p_i, p_j)} = \frac{-\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)} \tag{2}$$

In our use case, we measure metrics one by one (experiments) and we expect one of the k possible outcome values. Before receiving any evidence, all outcomes are equally likely (a Dirichlet distribution with all parameters set to 1). Upon receiving a piece of evidence, we update our belief using the Bayesian updating method. Interestingly, the updated belief also follows a Dirichlet distribution with the corresponding parameter to the observed outcome being incremented by 1 (3).

$$Pr(P|E) \propto Pr(E|P) \times Pr(P) \\ P|(E = o_n) \sim Dir(\alpha_1, ..., \alpha_n + 1, ..., \alpha_k) \tag{3}$$

However, it should be noted that not all pieces of evidence are completely reliable and we have to take into account our confidence in the correctness of evidence. Based on the idea of using a learning rate in the Bayesian updating formula introduced in [26], we can incorporate the evidence trustworthiness ($t(e) \in [0,1]$) in (3) as shown in (4).

$$Pr(P|E) \propto Pr(E|P)^{t(e)} \times Pr(P) \\ P|(E = o_n) \sim Dir(\alpha_1, ..., \alpha_n + t(e), ..., \alpha_k) \tag{4}$$

But before that, we have to find a way to calculate $t(e)$. We use an adapted version of the algorithm presented in [27] to calculate the evidence trustworthiness, as follows:

1) Assign an initial trust value to each source based on what we know of them; accredited sources are assigned a

TABLE I
DIMENSIONS AND METRICS

No.	Dimensions	Metrics
1	Identity and Access Management (IAM)	Privileged MFA enforcement, IAM lifecycle maturity, Conditional access support, etc.
2	Compute Isolation & Confidential Computing	Confidential computing support, Measured/verified boot, VM/TEE attestation, etc.
3	Vulnerability & Patch Management	Vulnerability Disclosure Policy (VDP), Security update communication, Bug bounty program, etc.
4	Security Monitoring & Incident Response	24/7 Security operations, Dedicated incident response team, Threat intel integration, etc.
5	Reliability & Resilience	Fault isolation architecture, Reliability SLAs, Disaster recovery capability, etc.
6	Data Governance, Privacy, and Residency	Privacy program maturity, Regional data residency enforcement, Internal data access governance, etc.
7	Supply Chain & Hardware Provenance	Supplier security governance, Responsible sourcing & diversity, Secure hardware decommissioning, etc.
8	Network Security and Segmentation	Micro-segmentation capability, Tenant isolation strength, External network segregation, etc.

fixed trust value of 1; unaccredited sources are assigned a trust value based on their reputation.

2) If there is a piece of evidence from an accredited source, this piece of evidence and all its supporting pieces of evidence will be assigned a trust value of 1; all conflicting pieces of evidence will be assigned a trust value of 0. Finally, move to step 4.
3) If there is not any piece of evidence from an accredited source, the trust value of each received piece of evidence e is calculated as follow:

$$\begin{aligned} sup &= 1 - \prod_{e' \in SE} (1 - t(src_{e'})) \\ con &= 1 - \prod_{e' \in CE} (1 - t(src_{e'})), \text{ if } CE \neq \phi \\ t(e) &= \begin{cases} sup & \text{if } CE = \phi \\ \frac{sup}{sup+con} & \text{if } CE \neq \phi \end{cases} \end{aligned} \tag{5}$$

Where SE is a set of supporting pieces of evidence, CE is a set of conflicting pieces of evidence, and src_e is the source provider of a piece of evidence e.

4) Trust value for each unaccredited source is updated using an exponential moving average on the trust value of its provided piece of evidence, as follows:

$$t_{new}(src_e) = (1-\rho).t_{old}(src_e)+\rho.t(e);\ \rho \in (0,1) \tag{6}$$

The higher the value of ρ is, the more the focus moves toward the new trust value.

5) Repeat steps 2-4 upon receiving a new set of evidence.

After gathering all evidence for a metric, we reach a posterior Dirichlet distribution based on which we can calculate the expected value of the metric ($E[S_m]$) as shown in (7).

$$\begin{aligned} S_m &= \sum_{i=1}^{k} p_i o_i \\ E[S_m] &= \sum_{i=1}^{k} E[p_i] o_i \end{aligned} \tag{7}$$

The more evidence we get for a metric, the more confident we become that the metric value is close to its expected value [28]. Because more evidence leads to a lower variance in the posterior Dirichlet distribution. In other words, the distribution becomes more concentrated about the expected value. Therefore, we can measure a confidence score for the expected value based on the variance. Variance of the metric value can be calculated using (8).

$$\begin{aligned} Var(S_m) &= Var(\sum_{i=1}^{k} p_i o_i) \\ &= \sum_{i=1}^{k} o_i^2 Var(p_i) + 2 \sum_{i<j\leq k} o_i o_j Cov(p_i, p_j) \\ &= \frac{\sum_{i=1}^{k} E[p_i] o_i^2 - (\sum_{i=1}^{k} E[p_i] o_i)^2}{\alpha_0 + 1} \end{aligned} \tag{8}$$

According to the Bhatia-Davis inequality [29] on the variance of any bounded probability distribution with an expected value of μ, we have:

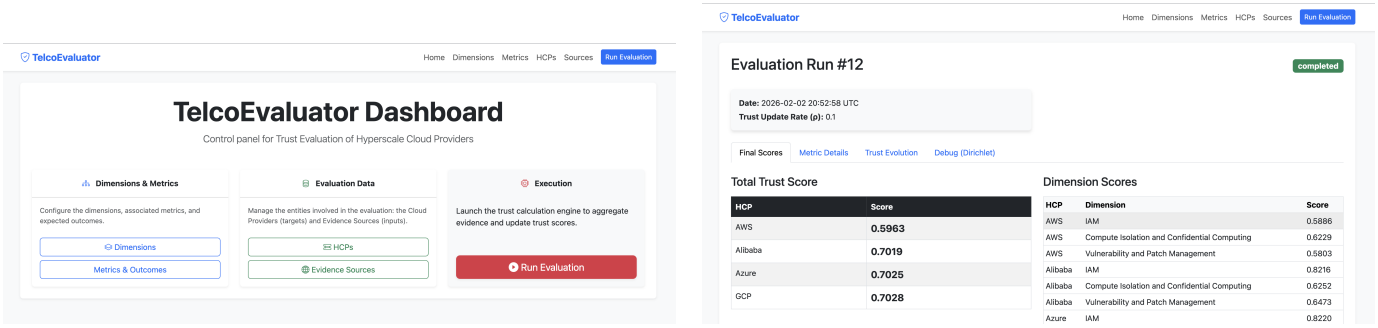
$$\sigma^2 \leq (M - \mu)(\mu - m) \tag{9}$$

In (9), M is the maximum value and m is the minimum value that the random variable can take. Therefore, the maximum variance of the score with an expected value of $\mu = E[S_m]$ can be calculated as follows:

$$MaxVar(S_m) = \mu(1 - \mu) \tag{10}$$

Finally, we can use (11) to calculate our confidence in the metric expected value. As $Var(S_m)$ approaches 0, our confidence approaches 1; the is due to the fact that the probability mass will be concentrated around the expected value. On the other hand, as $Var(S_m)$ approaches $MaxVar(S_m)$, our confidence approaches 0.

$$c_m = 1 - \sqrt{\frac{Var(S_m)}{MaxVar(S_m)}} \tag{11}$$



(a) Home page

(b) An evaluation run results

Fig. 2. Screenshots of the developed application interface.

Standard deviation is used in final calculation as it is more interpretable and intuitive than variance; because it is in the same scale as the original metric.

We use weighted geometric mean to calculate a dimension score (S_d) based on its associated metrics' expected values and confidence scores. By using weighted geometric mean, a metric with a very low expected value exerts a strong multiplicative penalty, proportional to its assigned weight, on the total dimension score. Consequently, metrics with high expected values can only partially compensate for low-valued metrics, ensuring sensitivity to poor security performance. If M_d is a set of all metrics defined in a dimension d, we have:

$$S_d = \prod_{m \in M_d} (c_m E[S_m])^{w_m}; w_m \geq 0; \sum_{m \in M_d} w_m = 1 \quad (12)$$

Finally, we use weighted geometric mean, for the same reason, to calculate the total trust score. If D is the set of all defined dimensions, we have:

$$S_{tot} = \prod_{d \in D} S_d^{w_d}; w_d \geq 0; \sum_{d \in D} w_d = 1 \quad (13)$$

IV. Experimental Results

A. Testbed Implementation

In order to evaluate the feasibility and effectiveness of our proposed framework, we developed a testbed on top of the OpenStack [30] infrastructure. We created multiple compute instances (virtual servers) to represent a telco, four well-known HCPs (namely, AWS, Azure, GCP–Google Cloud Platform, and Alibaba cloud), an accredited auditor who issues CSA STAR Level 2 certificates based on answers to CSA CAIQs (Consensus Assessment Initiative Questionnaire), and some simulated sources of information. All these instances can communicate with each other through a local network to simulate the real-world interaction through the Internet. We used Python to develop a distributed web-based application on top of this virtual infrastructure. In our developed application, a telco client communicates with different sources using RESTful API to gather evidence and calculate trust values following our proposed framework. Fig. 2 shows some screenshots of the developed application interface.

In Fig. 2(a) the home page is shown where you can see some links to configuration pages where the user can define dimensions, metrics, expected outcomes, target HCPs, and potential sources. On the other hand, Fig. 2(b) shows the results page of an evaluation run; as you can see, there are several tabs showing the total trust scores, metric values, trust evolution of used sources, and the current Dirichlet states.

Since the evidence provided by sources is usually unstructured text, we need an LLM (Large Language Model) to extract the metric values. For simplicity, we used two famous LLM-integrated chatbots, namely OpenAI's ChatGPT and Google's Gemini to extract the values of 9 metrics as shown in Table II; for example, our conversation with Google Gemini Deep Research agent is available at https://gemini.google.com/share/fb18f7e0b962. We selected a limited number of metrics to be able to manually verify the correctness of the extracted values. We double-checked the results by following the links provided by the agents in the final reports in order to verify that there were no hallucinations. As shown in table II, all HCPs, except AWS, are assigned the same values for all the selected metrics; only the values of three metrics for AWS, and four metrics for the rest could be confirmed based on CSA STAR certifications; the remaining metrics are out of scope of the CAIQ. We assigned weights (1–5) to each dimension and metric based on their importance in trust evaluation. These weights will be normalized before being used in calculations. Although, as mentioned before, these weights are configurable by telcos based on their own preferences.

We populated the database on each HCP virtual server with metric values inferred from their corresponding documentation, and the database on the accredited auditor virtual server with metric values inferred from CAIQs. Additionally, we created 5 simulated sources: 4 fanatic sources and a fair source. Each fanatic source supports an HCP by confirming the values inferred from their public documentation while telling lies about other HCPS; for example, "AWS Fanatic", fanatically supports AWS by confirming values published by AWS, and telling lies about Alibaba, Azure, and GCP. "Fair Source", on the other hand, supports fairly all HCPs. Using these simulated sources, we would like to show how our framework reacts to

TABLE II
DIMENSIONS, METRICS, WEIGHTS AND VALUES FOR AWS, AZURE, GCP, AND ALIBABA CLOUD

Dimension	Metric	Wt.	AWS		Azure		Google Cloud		Alibaba Cloud	
			Doc	CAIQ	Doc	CAIQ	Doc	CAIQ	Doc	CAIQ
IAM (Weight: 5)	Privileged MFA	5	Enforced	N/A	Enforced	Enforced	Enforced	Enforced	Enforced	Enforced
	IAM Lifecycle Maturity	4	Fully automated	Fully automated	Fully automated	Fully automated	Fully automated	Fully automated	Fully automated	Fully automated
	Conditional Access	3	Built-in / Default	N/A	Built-in / Default	N/A	Built-in / Default	N/A	Built-in / Default	N/A
Compute Isolation (Weight: 4)	Confidential Computing	4	Generally available	Generally available	Generally available	Generally available	Generally available	Generally available	Generally available	Generally available
	Measured Boot	3	Generally available	N/A	Generally available	N/A	Generally available	N/A	Generally available	N/A
	VM Attestation	3	Generally available	N/A	Generally available	N/A	Generally available	N/A	Generally available	N/A
Vuln. & Patch Mgmt (Weight: 4)	VDP	3	Formal public VDP	N/A	Formal public VDP	N/A	Formal public VDP	N/A	Formal public VDP	N/A
	Security Updates	4	Dedicated portal	Dedicated portal	Dedicated portal	Dedicated portal	Dedicated portal	Dedicated portal	Dedicated portal	Dedicated portal
	Bug Bounty	2	Limited / Inactive	N/A	Active w/ scope	N/A	Active w/ scope	N/A	Active w/ scope	N/A

Note: "Doc" refers to Public Documentation; "Wt." refers to Metric Weight (1-5).

Fig. 3. HCPs Total Trust Scores based on Different Combinations of Sources Used ($\rho = 0.1$).

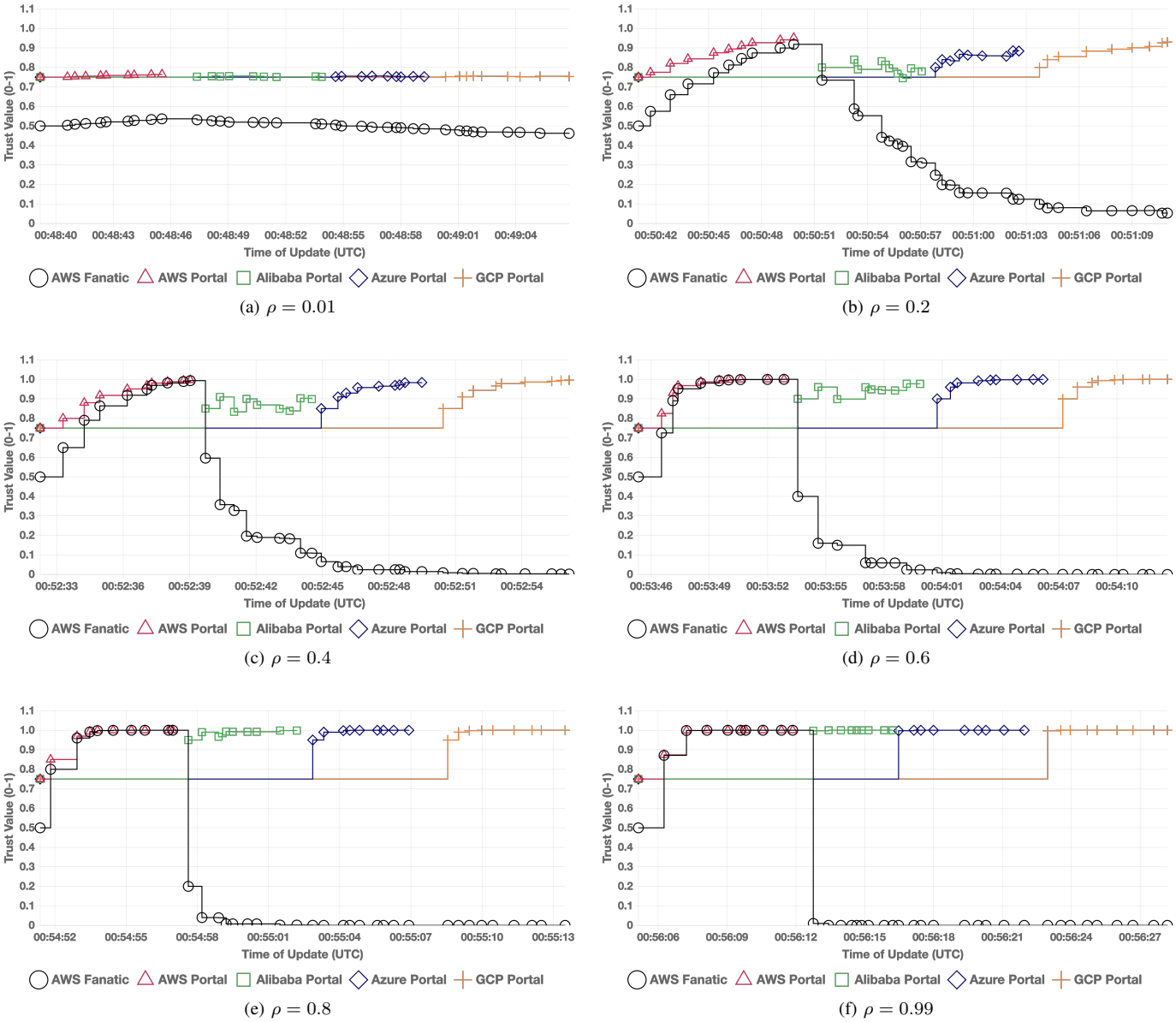
different sources. Regarding the trustworthiness of sources in telling the truth, we assigned a trust value of 0.75 to all HCPs documentation; the accredited source is fully trusted; other unknown simulated sources are assigned an initial trust value of 0.5.

B. Evaluation Runs and Results

1) The Effect of Involved Sources on HCPs Trust Scores: We used different combinations of sources and a fixed value of $\rho = 0.1$ to run our experiments. The total trust score calculated for all HCPs in each experiment is shown in Fig. 3. AWS has the lowest score in all cases; because it has a lower value assigned to its "Bug Bounty" metric. When we rely solely on public documentation, all HCPs are more or less equal with respect to their trust score; this is due to the fact that "Bug Bounty" metric is assigned a low weight and the confidence in all metric values is low. By adding CSA STAR, the trust values for all HCPs improve considerably, and the difference between AWS and the rest becomes noticeable; it follows the fact that only three metric values for AWS are confirmed based on CAIQ, whereas four metric values for each of the other HCPs are confirmed based on CAIQ. Confirmation based on CAIQ improves the confidence in each approved metric value and thus raises the final trust score.

Introducing the "AWS Fanatic" source initially has negative impact. However, as this source trustworthiness declines due to providing false information, its influence progressively diminishes. In our experiments, we evaluate HCPs in the following order: AWS, Alibaba Cloud, Azure, GCP. In the beginning, "AWS Fanatic" has enough power to boost the trustworthiness of AWS; although it is not as powerful as CSA STAR and therefore it cannot boost the AWS trust value beyond the trust values of the remaining HCPs. For GCP, it has almost no influence at all. When we include all fanatic sources, they cancel out each others influence, and the difference between trust values backs to normal. Adding the "Fair Source", enhances the confidence in all metric values and therefor increases the trust scores of all HCPs almost equally.

2) The Effect of ρ on HCPs Trust Scores: Selecting the value of ρ is challenging; low values result in slow changes in the sources trustworthiness, whereas, high values cause sudden, big jumps and/or drops in the sources trustworthiness. In order to demonstrate the effect of ρ on total trust scores we ran 6 experiments using sources AWS, Alibaba, Azure, GCP, CSA STAR, and AWS Fanatic, with 6 different values of $\rho \in \{0.01, 0.2, 0.4, 0.6, 0.8, 0.99\}$. The results are shown in



(a) $\rho = 0.01$ (b) $\rho = 0.2$

(c) $\rho = 0.4$ (d) $\rho = 0.6$

(e) $\rho = 0.8$ (f) $\rho = 0.99$

Fig. 4. Changes in sources trustworthiness over time under different values of ρ (included sources: {AWS, Alibaba, Azure, GCP, CSA STAR, AWS Fanatic}).

Fig. 4–5; Fig. 4 shows the changes in sources trustworthiness over time, while Fig. 5 compares the total scores calculated for each HCP in each run. It can be inferred that, although lower values of ρ reduce the influence of the fanatic source on the total trust scores, they slow down the speed of sources in building trust and therefore the final confidence score for each metric will be low, which leads to low trust scores calculated in general. On the other hand, high values of ρ lead to exaggerated belief in the trustworthiness of sources which leads to overestimation of trust values.

C. Discussion

Based on the results presented in this section, we believe the effectiveness of our proposed framework is mainly reliant on the sources involved and the value of ρ. In fact, the trustworthiness of each piece of evidence depends on the outcome of battle between truthful sources and malicious sources. Selecting a good value of ρ depends on how sensitive we are in rewarding/punishing sources; low values are less rewarding while high values result in abrupt drops in trustworthiness even when a source unintentionally makes a mistake. Furthermore, high values would noticeably reward malicious sources for telling the truth, which could help them exert malicious influence in the future.

V. Conclusion

In this paper, we proposed a framework for trust evaluation of HCPs in hosting and protecting telco workloads in the 5G and beyond era. In this framework, a fine-grained analysis is performed by measuring trust based on multiple metrics

Fig. 5. HCPs Total Trust Scores based on Different Values of ρ (Sources: {AWS, Alibaba, Azure, GCP, CSA STAR, AWS Fanatic}).

from several dimensions. Our belief regarding the value of each metric is modeled using the Dirichlet distribution with Bayesian Updating. Total trust scores are calculated by combining metrics' expected values and confidence scores derived from their distributions. Experimental results demonstrate the effectiveness of our proposed framework in comparing four major HCPs. The effectiveness of this framework relies mainly on the involved sources and the value of ρ. There are some limitations in the proposed framework that we are planning to investigate in our future works. First, evidence freshness should also be considered as a factor to determine evidence trustworthiness. Second, independence of sources should be taken into consideration to avoid inflated confidence based on evidence received from two or more correlated sources. Third, training a specialized LLM and improving its performance based on user feedback helps in enhancing the extraction accuracy. Finally, making the value of ρ adaptable to the context will greatly improve the system's performance in combating malicious sources. We hope that this research lays the foundation for an active, prosperous future research in this direction.

REFERENCES

[1] S. Bouakkaz, L. Suárez, N. Cuppens, and F. Cuppens, "Design of an intelligent trust management architecture for 5g service deployment," in *Proceedings of the 11th International Conference on Information Systems Security and Privacy - Volume 1: ICISSP*, INSTICC. SciTePress, 2025, pp. 310–317.

[2] C. Benzaïd, T. Taleb, and M. Z. Farooqi, "Trust in 5g and beyond networks," *IEEE Network*, vol. 35, no. 3, pp. 212–222, 2021.

[3] ETSI GR NFV-IFA 054 V6.1.1, "Network Functions Virtualisation (NFV) Release 6; Architecture; Report on architectural support for NFV evolution," Feb. 2025.

[4] J.-H. Cho, K. Chan, and S. Adali, "A survey on trust modeling," *ACM Computing Surveys (CSUR)*, vol. 48, no. 2, pp. 1–40, 2015.

[5] ETSI GS NFV-SEC 003 V1.1.1, "Network Functions Virtualisation (NFV); NFV Security; Security and Trust Guidance," Dec. 2014.

[6] S. M. Habib, S. Ries, M. Mühlhäuser, and P. Varikkattu, "Towards a trust management system for cloud computing marketplaces: using caiq as a trust information source," *Security and Communication Networks*, vol. 7, no. 11, pp. 2185–2200, 2014.

[7] P. S. Pawar, M. Rajarajan, T. Dimitrakos, and A. Zisman, "Trust assessment using cloud broker," in *IFIP International Conference on Trust Management*. Springer, 2014, pp. 237–244.

[8] P. Manuel, "A trust model of cloud computing based on quality of service," *Annals of Operations Research*, vol. 233, no. 1, pp. 281–292, 2015.

[9] W.-J. Fan, S.-L. Yang, H. Perros, and J. Pei, "A multi-dimensional trust-aware cloud service selection mechanism based on evidential reasoning approach," *International Journal of Automation and Computing*, vol. 12, no. 2, pp. 208–219, 2015.

[10] N. Ghosh, S. K. Ghosh, and S. K. Das, "Selcsp: A framework to facilitate selection of cloud service providers," *IEEE Transactions on Cloud Computing*, vol. 3, no. 1, pp. 66–79, 2015.

[11] M. Tang, X. Dai, J. Liu, and J. Chen, "Towards a trust evaluation middleware for cloud service selection," *Future Generation Computer Systems*, vol. 74, pp. 302–312, 2017.

[12] V. C. Emeakaroha, K. Fatema, L. van der Werff, P. Healy, T. Lynn, and J. P. Morrison, "A trust label system for communicating trust in cloud services," *IEEE Transactions on Services Computing*, vol. 10, no. 5, pp. 689–700, 2016.

[13] D. Gonzales, J. M. Kaplan, E. Saltzman, Z. Winkelman, and D. Woods, "Cloud-trust—a security assessment model for infrastructure as a service (iaas) clouds," *IEEE Transactions on Cloud Computing*, vol. 5, no. 3, pp. 523–536, 2015.

[14] A. Algamdi, F. Coenen, and A. Lisitsa, "A trust evaluation method based on the distributed cloud trust protocol (ctp) and opinion sharing," *Provider*, vol. 5, no. 17, p. 18, 2017.

[15] J. Sidhu and S. Singh, "Improved topsis method based trust evaluation framework for determining trustworthiness of cloud service providers," *Journal of Grid Computing*, vol. 15, no. 1, pp. 81–105, 2017.

[16] L. Van Der Werff, G. Fox, I. Masevic, V. C. Emeakaroha, J. P. Morrison, and T. Lynn, "Building consumer trust in the cloud: an experimental analysis of the cloud trust label approach," *Journal of Cloud Computing*, vol. 8, no. 1, p. 6, 2019.

[17] S. Rizvi, J. Mitchell, A. Razaque, M. R. Rizvi, and I. Williams, "A fuzzy inference system (fis) to evaluate the security readiness of cloud service providers," *Journal of cloud computing*, vol. 9, no. 1, p. 42, 2020.

[18] L. Maria Bruma, "Using cloud control matrix to evaluate trust in cloud providers," in *Proceedings of the 2021 10th International Conference on Software and Computer Applications*, 2021, pp. 273–278.

[19] A. K. Junejo, I. A. Jokhio, and T. Jan, "A multi-dimensional and multi-factor trust computation framework for cloud services," *Electronics*, vol. 11, no. 13, p. 1932, 2022.

[20] A. Balcão-Filho, N. Ruiz, F. de Franco Rosa, R. Bonacin, and M. Jino, "Applying a consumer-centric framework for trust assessment of cloud computing service providers," *IEEE Transactions on Services Computing*, vol. 16, no. 1, pp. 95–107, 2021.

[21] J. John and J. S. K, "Predictive digital twin driven trust model for cloud service providers with fuzzy inferred trust score calculation," *Journal of Cloud Computing*, vol. 13, no. 1, p. 134, 2024.

[22] J. M. J. Valero, V. Theodorou, M. G. Pérez, and G. M. Pérez, "Sla-driven trust and reputation management framework for 5g distributed service marketplaces," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 4, pp. 1863–1875, 2023.

[23] "CSA STAR registry," https://cloudsecurityalliance.org/star/registry, Cloud Security Alliance.

[24] "The register: Enterprise technology news and analysis," https://www.theregister.com/, Situation Publishing.

[25] C. J. Fung, J. Zhang, I. Aib, and R. Boutaba, "Dirichlet-based trust management for effective collaborative intrusion detection networks," *IEEE Transactions on Network and Service Management*, vol. 8, no. 2, pp. 79–91, 2011.

[26] P. Grünwald and T. van Ommen, "Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It," *Bayesian Analysis*, vol. 12, no. 4, pp. 1069 – 1103, 2017. [Online]. Available: https://doi.org/10.1214/17-BA1085

[27] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *Workshop on Secure Data Management*. Springer, 2008, pp. 82–98.

[28] J. Huang, "Maximum likelihood estimation of dirichlet distribution parameters," *CMU Technique report*, vol. 76, 2005.

[29] R. Bhatia and C. Davis, "A better bound on the variance," *The American Mathematical Monthly*, vol. 107, no. 4, pp. 353–357, 2000. [Online]. Available: https://doi.org/10.1080/00029890.2000.12005203

[30] "Openstack: Open source cloud computing infrastructure," https://www.openstack.org/, OpenInfra Foundation.