ISSN: 0976-3104
SPECIAL ISSUE: (Emerging Technologies in Networking and Security (ETNS)

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

A SURVEY ON VARIOUS APPROACHES FOR VERIFYING CORRECTNESS AND COMPLETENESS OVER THE CLOUD DATA

S. Suvetha*,V. Arul, G. Sasikala

Dept of Computer Science and Engineering, Vivekanandha College of Engineering for women, Namakkal, T.N., INDIA

ABSTRACT

Aim: *Cloud computing is popularizing the computing archetype in which data is outsourced to a third-party service provider (server) for data mining Outsourcing. However, it raises a serious security issue: how can the client of weak computational power verify that the server returned correct mining result or not. By using homomorphic encryption algorithm can check the completeness and correctness of retrieved data from the server.*

Published on: 2nd -December-2016

KEY WORDS

Verifiable computing, Discovery, Data Mining as a Service.

***Corresponding author: Email:** suvethajan14@gmail.com; **Tel.:** +9500952218

INTRODUCTION

These days, the IT world is moving towards the pay-per-use paradigm named Cloud Computing. Companies of all sizes reduce their computing assets and shift to a use of computing resources in the clouds[1]. One consequence of this shift is that the IT world outside the clouds is moving to a use of weaker and smaller computer devices, like Virtualized Thin Desktops and Smart phones. Whenever stronger resources are needed, those devices can use the cloud. Data mining-as a service expertise to outsource their data mining needs to a third part service provider [4]. As an example, the operational transactional data from various stores of a supermarket chain can be shipped to a third party which provides mining services [3]. The supermarket management need not employ an in-house team of data mining experts [2]. Besides, they can cut down their local data management requirements because periodically data is shipped to the service provider who is in charge of maintaining it and conducting mining[6] on it in response to requests from business analysts of the supermarket chain. It is generally expected that the paradigm of “mining and management of data as service” will grow with the advent and popularity of cloud computing.

Cloud Computing

Cloud computing is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. Cloud services are given by an outsider entity, has many security and integrity problems [2]. Security in this context means that the client will receive an assurance that the computation performed by the server is correct, with the optional property that the client will be able to hide some of his data from the server [3]. Thus, the client must have some way of verifying the cloud’s computation. However, one basic problem is inherent in the model: How can a weak client verify the correctness of the cloud’s computation? Can the client be assured that the cloud server follows its declared strategy? These questions are not easily answered by the existing tools of security and cryptography [7]. There are many possible reasons for a cloud to cheat on answers. For example When perform a web search, expect that the list of links returned will be relevant and complete. As heavily rely on web searching, an often overlooked issue is that search engines are outsourced computations [13]. That is, users issue queries and have no intrinsic way of trusting the results they receive, thus introducing a modern spin on Cartesian doubt. This philosophy once asked if can trust our senses now it should ask if can trust our search results.

Some possible attack scenarios that arise in this context include the following:

1. A news web site posts a misleading article and later changes it to look as if the error never occurred.
2. A company posts a back-dated white paper claiming an invention after a related patent is issued to a competitor.
3. An obscure scientific web site posts incriminating data about a polluter, who then sues to get the data removed, in spite of its accuracy.
4. A search engine censors content for queries coming from users in a certain country, even though an associated web crawler provided web pages that would otherwise be indexed for the forbidden queries[5][8].

Verifiable Computing

The cloud would like to improve its revenue by computing things with minimal resources while charging for more [9]. This problem of verifiable computation was tackled in many previous works in the theoretical computer science community, most notably by using Probabilistically Checkable Proofs[12] . Other recent works use fully homomorphic encryption and get amortized performance advantages. Today, a common way to verify computations is replication. However, replication may not be verifiable. It also requires assumptions about failure independence. Another technique is auditing but if the performer understands the computation better than the requester, the performer can alter strategic bits, undetected by an audit[20]. A final technique is trusted computing , but it assumes that some component the hardware, the hypervisor, a higher layer is not physically altered. Propose homomorphic encryption verification approaches to check whether the server has returned correct and complete frequent itemsets[25]. Our homomorphic encryption approach can catch incorrect results with high probability, while our deterministic approach measures the result correctness with 100 % certainty. It also design efficient verification methods for both cases that the data and the mining setup are updated. It demonstrate the effectiveness and efficiency of our methods using an extensive set of empirical results on real datasets[12]. An interesting direction to explore is to extend the model to allow the client to specify her verification needs in terms of budget (possibly in monetary format) besides precision and recall threshold.

LITERATURE SURVEY

The following papers are surveyed. It is consists of various verification approaches. These approaches are how the client of weak computational power verify that the server returned correct mining results or not. These following approaches are used to verify the retrieved result from the server.

In[1] R. Canetti, B. Riva, and G. N. Rothblum, presented a paper on "**Verifiable computation with two or more clouds"** server using one cloud, the client uses two or more different clouds to perform the computation[1]. The client can verify the correct result of the computation, as long as at least one of the clouds is honest. It believes that such addition suits the world of cloud computing where cloud providers have incentives not to collude, and the client is free to use any set of clouds he wants. Our results are two fold[7]. First, they show two protocols in this model:1. A computationally sound verifiable computation for any efficiently computable function, with logarithmically numerous rounds, based on any collision-resistant hash family. 2. A 1-round (2-messages) unconditionally sound verifiable computation for any function computable in log-space uniform NC. Second, It show that our first protocol works for essentially any sequential program, and they present an implementation of the protocol, called quin, for Windows executables. Also describe its architecture and experiment with several parameters on live clouds[20].

In[2] F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi, and W. Hui Wang,presented a paper on **"Privacy-preserving data mining from outsourced databases"** , the problem of outsourcing the association rule mining task within a corporate privacy-preserving framework[19]. Association rule mining has the objective of discovering groups of products, or items, that are repeatedly purchased together by the supermarket's customers: the expected output of such task, given the sale transaction database as input, is the list of all possible groups of things, such as {milk, beer, diapers}, that occur together in a fraction of the market baskets that is statistically significant[2]. The complexity of this task is evident: there are tens of thousands of distinct products in the variety of a supermarket, and therefore the number of potential candidate groups of products quickly explodes with the size of the group[23]. Our encryption scheme is based on 1–1 substitutions together with addition of fake transactions such that the transformed database satisfies k-anonymity with respect to items and itemsets. This computational complexity motivates the introduction of an outsourcing model, where the data owner, like our supermarket, gives the data in outsourcing to a service provider to obtain an association rule mining service from it, within a privacy-preserving framework, i.e., without disclosing neither the sale data nor the information deriving from the mining analysis[25].

In[3] R. Liu, H. Wang, A. Monreale, D. Pedreschi, F. Giannotti, and WengeGuo, presented a paper on **"Audio: An integrity auditing framework of Outlier mining- as-a-service systems"** The AUDIO, an integrity auditing framework for the specific task of distance-based outlier mining outsourcing[3]. It provides efficient and practical verification approaches to check both completeness and correctness of the mining results. The key idea of our approach is to insert a small amount of artificial tuples into the outsourced data[18]. The artificial tuples will produce artificial outliers and non-outliers that do not exist in the original dataset. The server's answer is verified by analyzing the presence of artificial outliers/non-outliers, obtaining a probabilistic guarantee of correctness and completeness of the mining result. Our empirical results show the effectiveness and efficiency of our method [5].

In[4] S. Setty, A. J. Blumberg, and M. Walfish, presented paper on **"Toward practical and unconditional verification of remote computations"** it propose a new line of systems research: using the machinery of PCPs, can build a system that (i) has practical performance, (ii) is simple to implement, and (c)provides unqualified guarantees? Note that (a) and (b)contrast with PCPs as used in the theory literature and (c)contrasts with current systems approaches. To illustrate the promise of this line of research, do the following: (1) Identify work in the PCP literature that provides a base for systems research [4]. First looked to the PCP literature, then observed that efficient argument systems (PCP variants in which the server proves that it has a proof by answering questions interactively) are promising, and then noticed that a particular argument system could lead to a practical solution[19]. (2) Refine the approach of into a design that is practical over a limited domain . It applied refinements to shrink program encoding (via arithmetic circuits instead of Boolean circuits), enable batched proofs (which enhances performance for computations that can be decomposed into parallel pieces), and improve amortization (by moving more of the work to a setup phase). These innovations are essential to practical performance. (3) Implement this design to demonstrate its practicality. To our knowledge, PCP theory has never before found its way into any efficient implementation. Thus, believe that our implementation, though limited, is a contribution. Our implementation is also comparatively simple; it could conceivably be formally verified. (4) Articulate a research agenda for extending the reach of our approach . Our ultimate goal is a practical system for general-purpose verified computation [17]. The four contributions above provide a concrete foundation for our position, which is that PCP-based verifiable computation can be a systems problem, not just a theory problem. They need this foundation because PCPs are thought to be impractical; indeed, our prior designs were too expensive by over 11 orders of magnitude. Even our prototype achieves goals (a)–(c) above only over a limited domain [22].

In[5] S. Benabbas, R. Gennaro, and Y. Vahlis, presented a paper on **"Verifiable delegation of computation over large datasets"** this learn the problem of computing on large datasets that are stored on an untrusted server[5]. They follow the approach of amortized verifiable computation introduced by Gennaro, Gentry, and Parno. This present the first practical verifiable computation scheme for high degree polynomial functions[23]. Such functions can be used, for example, to make predictions based on polynomials fitted to a large number of sample points in an experiment. Our second result is a primitive which call a verifiable database (VDB). Here, a weak client outsources a large table to an untrusted server, and makes retrieval and update queries. For each query, the server provides a response and a proof that the response was computed correctly[26]. The goal is to minimize the resources required by the client. This is made particularly challenging if the number of update queries is unbounded. It presents a VDB scheme based on the hardness of the subgroup membership problem in composite order bilinear groups[19].

In[6] R. Canetti, B. Riva, and G. N. Rothblum, presented a paper on **"Practical delegation of computation using multiple servers"**this demonstrate a relatively efficient and general solution where the client delegates the computation to several servers, and is guaranteed to determine the correct answer as long as even a single server is honest [6]. It show: A protocol for any efficiently computable function, with logarithmically many rounds, based on any collision resistant hash family[10]. The protocol is set in terms of Turing Machines but can be adapted to other computation models. An adaptation of the protocol for the X86 computation model and a prototype implementation, called Quin, for Windows executables[21]. It describe the architecture of Quin and experiment with several parameters on live clouds. Also show that the protocol is practical, can work with nowadays clouds, and is efficient both for the servers and for the client.

In[7] D. Fiore and R. Gennaro, presented a paper on **“Publicly verifiable delegation of large polynomials and matrix computations, with applications”** The Outsourced computations (where a client requests a server to perform some computation on its behalf) are becoming increasingly important due to the rise of Cloud Computing and the proliferation of mobile devices[7]. Since cloud providers may not be trusted, a crucial problem is the verification of the integrity and correctness of such computation, possibly in a public way, i.e., the result of a computation can be verified by any third party, and requires no secret key { akin to a digital signature on a message. It present new protocols for publicly verifiable secure outsourcing of Evaluation of High Degree Polynomials and Matrix Multiplication[28]. It can be used for amortized model. Optimal Verification of Operations on Dynamic Sets it present a new authenticated data structure scheme that allows any entity to publicly verify the correctness of primitive sets operations such as intersection, union, subset and set difference. Based on a novel extension of the security properties of bilinear-map accumulators as well as on a primitive called accumulation tree, our authenticated data structure is the first to achieve optimal verification and proof complexity (i.e., only proportional to the size of the query parameters and the answer), as well as optimal update complexity (i.e., constant), and without bearing any extra asymptotic space overhead[13]. Queries (i.e., constructing the proof) are also efficient, adding a logarithmic overhead to the complexity needed to compute the actual answer.

In[8] M. T. Goodrich, C. Papamanthou, D. Nguyen, R. Tamassia, C. V. Lopes, O. Ohrimenko, and N. Triandopoulos, presented a paper on **“Efficient verification of web-content searching through authenticated web crawlers,”**It consider the problem of verifying the correctness and completeness of the result of a keyword search[8]. They introduce the concept of an authenticated web crawler and present its design and prototype implementation. An authenticated web crawler is a trusted program that computes a specially- crafted signature over the web contents it visits. This signature enables (i) the verification of common Internet queries on web pages, such as conjunctive keyword searches this guarantees that the output of a conjunctive keyword search is correct and complete[26]; (ii) the verification of the content returned by such Internet queries this guarantees that web data is authentic and has not been maliciously altered since the computation of the signature by the crawler. In our solution, the search engine returns a cryptographic proof of the query result. Both the proof size and the verification time are proportional only to the sizes of the query description and the query result, but do not depend on the number or sizes of the web pages over which the search is performed[15]. As experimentally demonstrate, the prototype implementation of our system provides a low communication overhead between the search engine and the user, and fast verification of the returned results by the user.

In[9] B. Parno, M. Raykova, and V. Vaikuntanathan, **“How to delegate and verify in public: Verifiable computation from Attribute-based encryption”**The outsourcing computation is useful only when the returned result can be trusted, which makes verifiable computation (VC) a must for such scenarios[9]. In this work extend the definition of verifiable computation in two important directions: public delegation and public verifiability, which have important applications in many practical delegation scenarios. Yet, existing VC constructions based on standard cryptographic assumptions fail to achieve these properties. As the primary contribution of our work, establish an important (and somewhat surprising) connection between verifiable computation and attribute-based encryption (ABE), a primitive that has been widely studied. Namely, it show how to construct a VC scheme with public delegation and public verifiability from any ABE scheme[16]. The VC scheme verifies any function in the class of functions covered by the permissible ABE policies (currently Boolean formulas). This scheme enjoys a very efficient verification algorithm that depends only on the output size. Efficient delegation, however, requires the ABE encryption algorithm to be cheaper than the original function computation[17]. Strengthening this connection, It show a construction of a multi-function verifiable computation scheme from an ABE scheme with outsourced decryption, a primitive defined recently by Green, Hohen berger and Waters[18]. A multi-function VC scheme allows the verifiable evaluation of multiple functions on the same preprocessed input. In the other direction, They also explore the construction of an ABE scheme from verifiable computation protocols.

In[10] Justin Thaler, Mike Roberts, Michael Mitzenmacher, and Hanspeter Pfister presented a paper on **“Verifiable Computation with Massively Parallel Interactive Proofs”** It assess the potential of parallel processing to help make practical verification a reality, identifying abundant data parallelism in a state-of-the-art general purpose protocol for verifiable computation[10]. It implement this protocol on the GPU, obtaining 40-120 server-side speedups relative to a state-of-the-art sequential implementation[22]. For benchmark problems, our implementation thereby reduces the slowdown of the server to within factors of 100-500 relative to the original computations requested by the client. Furthermore, it reduce the already small runtime of the client by 100. Our results demonstrate the immediate practicality of using GPUs for verifiable computation, and more generally, that protocols for verifiable computation have become sufficiently mature to deploy in real cloud computing systems[27].

COMPARATIVE ANALYSIS OF VERIFICATION APPROACHES

This section presents the comparison of different verification approaches.

Table: 1. Comparison of different Verification Approach

S.NO	PAPER TITLE	TECHNIQUES	ADVANTAGES	DISADVANTAGES
[1]	Verifiable Computation with Two or More Clouds	The interactive-proof model, Efficient-players refereed games (epRG).	It is easier to understand and to implement, and therefore might be adopted for real-world uses.	For each experiment ran the protocol several times with one cheating cloud that cheats on one out of three randomly chosen states.
[2]	Privacy-Preserving Data Mining from Outsourced Databases	Association rule mining, Privacy-preserving data mining (PPDM)	Effective, good privacy and accuracy	An individual item, a transaction, or the server can always be controlled to be a threshold chosen by the owner, by setting the anonymity threshold k.
[3]	AUDIO: An Integrity Auditing Framework of Outlier-Mining-as-a-Service Systems	AUDIO: An Integrity Auditing Framework of Outlier	It feasible to efficiently verify the outlier mining results of large databases.	It may not be able to catch the malicious server as it may launch verification-aware cheating
[4]	Toward practical and unconditional verification of remote computations	Probabilistically Checkable Proofs	realism, simplicity, and unconditional assurance	but it uses a limited domain only.
[5]	Verifiable Delegation of Computation over Large Datasets	amortized verifiable computation	It allows the client to insert and delete values, as well as update the value at any cell by sending a single group element to the server after retrieving the current value stored in the cell.	Implementation cost is high
[6].	Practical Delegation of Computation using Multiple Servers	Probabilistic Checkable	It is efficiently computable function, both for the servers	The main downside of this approach is the need for an honest majority of clouds.

		Proofs	and for the client.	
[7]	Publicly Verifiable Delegation of Large Polynomials and Matrix Computations, with Applications	amortized model	It is faster verification, a constant amount of computation, The result published is secure only under a weaker “selective” notion of security, where the adversary must commit in advance to the input point x on which it is going to cheat.	convolution is high
[8]	Efficient Verification of WebContent Searching Through Authenticated Web Crawlers	Web Crawlers, kew word search scheme	Get to gather the data you want	Traffic may be identified as abusive or suspicious and blocked .It may be constrained by limits in bandwidth, processing, or storage
[9]	How to Delegate and Verify in Public:Verifiable Computation from Attribute-based Encryption	ABE scheme attribute based encryption (one-key secure) , non-interactive verifiable computation	generality, efficiency, and adaptive security.	The attacker can easily recognize the key value.
[10]	Verifiable Computation with Massively Parallel Interactive Proofs	Interactive Proofs	It saves space and time for the verifier even when outsourcing a single computation, while saves time for the verifier only when batching together several dozen computations and amortizing the verifier’s cost over the batch.	parallel run due to the slow process.

POSSIBLE SOLUTION

In existing work probabilistic approach has been used for verifying the result. That approach used the key value as a whole value for the attacker can easily identify that value. They can use homomorphic algorithm for verification approach and it can split the key value. For example consider key value as 100 and split the key value like as 60 and 40. The attacker does not find out those key value. It also visualize and determine the how many truly relevant results are returned.

CONCLUSION

It presents an various verification approaches available in cloud computing. But the main challenge in cloud computing is security and integrity problems. In security in this context means that the client will receive an assurance that the computation performed by the server is correct. Thus, the client must have some way of verifying the cloud’s computation. They can use homomorphic encryption algorithm for verification approach.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENT

None

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Boxiang Dong, Ruilin Liu, and Hui (Wendy) Wang "Trust-but-Verify: Verifying Result Correctness of Outsourced Frequent Itemset Mining in Data-Mining-As-a-Service Paradigm" IEEE TRANSACTIONS ON SERVICES COMPUTING, 9(1) JANUARY/FEBRUARY 2016.

[2] R Canetti, B Riva, GN Rothblum. [2011] Verifiable computation with two or more clouds, *in Proc. Workshop Cryptography Security Clouds*.

[3] F Giannotti, LVS Lakshmanan, A Monreale, D Pedreschi, W Hui Wang. [2011] Privacy-preserving data mining from outsourced databases, in Proc. 3rd Int. Conf. Comput., Privacy Data Protection, , pp. 411–426.

[4] R Liu, H Wang, A Monreale, D Pedreschi, F Giannotti, Wenge Guo. [2012] Audio: An integrity auditing framework of Outliermining- as-a-service systems, *in Proc. Eur. Conf. Mach. Learning Knowl. Discovery Databases*, pp. 1–18.

[5] S Setty, AJ Blumberg, M Walfish. [2013] Toward practical and unconditional verification of remote computations. in Proc. 13th USENIX Conf. Hot Topics Operating Syst., p. 29.

[6] S Benabbas, R Gennaro, Y Vahlis. [2011] Verifiable delegation of computation over large datasets, in *Proc 31st Ann. Conf Adv Cryptol*, pp. 111–131.

[7] R Canetti, B Riva, GN Rothblum. [2011] Practical delegation of computation using multiple servers, in *Proc 18th ACM Conf Comput Commun Security*, pp. 445–454.

[8] D Fiore, R Gennaro. [2012] Publicly verifiable delegation of large polynomials and matrix computations, with applications, in Proc. ACM Conf. Comput. Commun. Security, pp. 501–512.

[9] MT Goodrich, C Papamanthou, D Nguyen, R Tamassia, CV Lopes, O Ohrimenko, N Triandopoulos. [2012] Efficient verification of web-content searching through authenticated web crawlers, *Proc. VLDB Endowment*, 5: 920–931.

[10] B Parno, M Raykova, V Vaikuntanathan. [2012] "How to delegate and verify in public: Verifiable computation from Attribute-based encryption," in Proc. 9th Theory Cryptography Conf , pp. 422–439.

[11] Justin Thaler_, Mike Roberts_, Michael Mitzenmacher, and Hanspeter Pfister "Verifiable Computation with Massively Parallel Interactive Proofs Harvard University, School of Engineering and Applied Sciences,2014

[12] M Yiu, I Assent, CS Jensen, P Kalnis.[2012] Outsourced Similarity Search on Metric Data Assets. In TKDE, 24

[13] AR Sadeghi, T Schneider, M Winandy.[2010] Token-based cloud computing: secure outsourcing of data and arbitrary computations with lower latency. In TRUST,

[14] M Bellare, B Waters, S Yilek.[2011]\Identity-based encryption secure against selective opening attack." To appear in TCC.

[15] A Lewko, Y Rouselakis, B Waters.[2011] \Achieving leakage resilience through dual system encryption." To appear in TCC.

[16] B Parno, M Raykova, V Vaikuntanathan.[2012] How to delegate and verify in public: Verifiable computation from attribute-based encryption. TCC.

[17] S Benabbas, R Gennaro, Y Vahlis. Verifiable delegation of computation over large datasets. In P. Rogaway, editor, Advances in Cryptology { CRYPTO 2011, volume 6841 of Lecture Notes in Computer Science, pages 111{131, Santa Barbara, CA, USA, Aug. 14{18, 2011. Springer, Berlin, Germany.

[18] P Mohassel. [2011] Efficient and secure delegation of linear algebra. Cryptology ePrint Archive, Report 2011/605.

[19] C Papamanthou, E Shi, R Tamassia.[2011] Signatures of correct computation. *Cryptology ePrint Archive,Report* 2011/587.

[20] N Bitansky, R Canetti, A Chiesa, E Tromer.[2011] From extractable collision resistance to succinct non-interactive arguments of knowledge, and back again. Cryptology ePrint Archive, Report 2011/443.

[21] D Boneh, A Sahai, B Waters. [2011]Functional encryption: Definitions and challenges. In Proceedings of the Theory of Cryptography Conference (TCC),

[22] R Canetti, B Riva, GN Rothblum.[2011] Two 1-round protocols for delegation of computation. *Cryptology ePrint Archive*, Report 2011/518

[23] C Gentry, D Wichs. [2011] Separating succinct non-interactive arguments from all falsifiable assumptions. In Proceedings of the ACM Symposium on Theory of Computing (STOC).

ISSN: 0976-3104

[24] S Goldwasser, H Lin, A Rubinstein.[2011] Delegation of computation without rejection problem from designated verifier CS-proofs. Cryptology ePrint Archive, Report 2011/456.

[25] S Setty, R McPherson, AJ Blumberg, and M. Walfish. Making argument systems for outsourced computation practical (sometimes). In Proc. NDSS, 2012.

[26] J Applequist. New assured cloud computing center to be established at Illinois. May 2011.

[27] G Cormode, J Thaler, K. Yi.[2011] Verifying computations with streaming interactive proofs. In Proc. VLDB Endowment.

[28] G Cormode, M Mitzenmacher, J Thaler.[2012] Practical verified computation with streaming interactive proofs. In Proc. ITCS.