NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 2, ISSUE 12, DEC.-2015

ENHANCED TIME SERIES PATTERN BASED EFFECTIVE NOISE GENERATION FOR PRIVACY PROTECTION ON CLOUD

Ashwini H Gajare
Department of Computer Engg.,
Savitribai Phule Pune University,Pune, India

ABSTRACT

Cloud computing can manage various IT resources and provide virtual scalable IT services under its openness and virtualization features. Hence, cloud customers can save huge capital investments in their own infrastructure by deploying or utilizing these IT services through cloud. Due to the openness and virtualization, various malicious service providers may exist in these cloud environments, and some of them may record service data from a customer and then collectively deduce the customer's private information without permission. Therefore, from the perspective of cloud customers, it is essential to take certain technical actions to protect their privacy at client side. Noise obfuscation is an effective approach in this regard by utilizing noise data. For instance, noise service requests can be generated and injected into real customer service requests so that malicious service providers would not be able to distinguish which requests are real ones if these requests occurrence probabilities are about the same, and consequently related customer privacy can be protected. Currently, existing representative noise generation strategies have not considered possible fluctuations of occurrence probabilities. In this case, the probability fluctuation could not be concealed by existing noise generation strategies, and it is a serious risk for the customer's privacy. To address this probability fluctuation privacy risk, we systematically develop a novel time-series pattern based noise generation strategy for privacy protection on cloud. First, we analyze this privacy risk and present a novel cluster based algorithm to generate time intervals dynamically.

INDEX TERMS — Cloud computing, privacy protection, noise obfuscation, noise generation, time-series pattern, cluster.

INTRODUCTION

Cloud privacy protection is a joint research frontier for both cloud computing and privacy protection. For instance, due to the openness and virtualization features, various malicious service providers may exist in cloud environments out of cloud customer's control. Meanwhile, these customers are quite hard to distinguish these malicious ones for the same reason. As a result, some of these malicious service providers can collect these customers' service data, such as service requests or communication logs, and then deduce their privacy without authorization or permission. Therefore, certain technical actions should be taken to protect their privacy automatically at client side. That is the cloud privacy protection at client side which is one essential aspect of the entire cloud privacy protection

For service providers, it is a common phenomenon to collect their customers' information, like service requests. From large to small firms, they commonly use them to analyze customers' behaviors, habits, and other sensitive information [5]. Most ethical ones have adequate self-control to use the information conforming to privacy-related regulations and Policies, but some others may abuse this information in unethical ways. Besides, these features also make customers difficult to distinguish which service providers are trustworthy (Ethical or unethical). Existing representative privacy protection approaches at server side have not taken this situation into a thorough consideration. For such type of cloud privacy risks, it is natural that customer privacy should be protected by taking certain technical actions automatically at client side, without involvement of service providers.

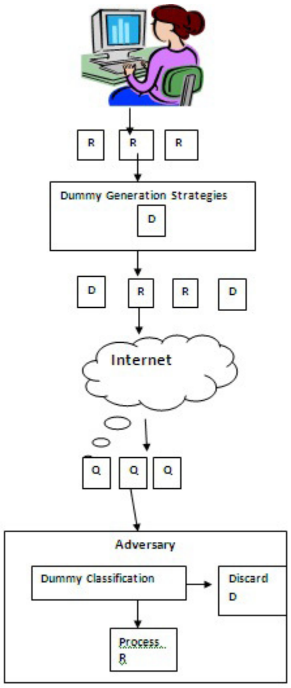
Let us take a weather service on cloud as a motivating example. One customer, who often travels to one city in Australia, like 'Sydney', checks the weather report regularly from a weather service on cloud before departure. The frequent appearance of service requests about the weather report for 'Sydney' can reveal the privacy that the customer usually goes to 'Sydney'. But if a system automatically injects other requests like 'Perth' or 'Darwin' into the 'Sydney' queues, the service provider cannot distinguish which ones are real and which ones are 'noise'. These requests should be responded and hard to reveal the location privacy of the customer. In such cases, the 'Sydney' privacy can be protected by noise obfuscation in general. One approach for improving this process is to decrease noise requests as in , under the pay-as-you-go style of cloud computing. But, given the privacy risk identified in this paper earlier, the customer could go to 'Sydney' in this month and 'Perth' in the next month. So, probabilities of real requests may have some fluctuations: 'Sydney' request is high in this month and low in the next month; 'Perth' request is low in this month and high in the next month. In the view of an entire service period, occurrence probabilities of 'Sydney' and Perth' may be about the same already for noise obfuscation and privacy protection. But in the view of time intervals, customer privacy can still be deduced, because these unconcealed fluctuations can reveal that the person goes to 'Sydney' in this month and 'Perth' in the next month. To address this, the goal of privacy protection in this paper is to keep occurrence probabilities of final requests to be about the same at any time intervals, instead of only in the entire time period. To achieve this goal, we will forecast these fluctuations by time-series patterns and generate noise service requests. In this example, privacy is the location information in service requests, not actual requests. We consider customer privacy without specific data structures as a general case for noise obfuscation in this paper, and we can extend this paper's work into other privacy types. In this example, a time interval is one month, and a time segment could be four months with a whole probability fluctuation. Besides, a time element could be one day which denotes the minimum time unit used in this example. According to the previous discussions, how to generate time intervals is a crucial issue in this paper, due to that they can decide the expressions of probability fluctuations. In the motivating example, the issue is why the time intervals are 'this month' and 'the next month'. For privacy attackers, time intervals are a mechanism to view probability fluctuations introduced before by controlling the expression of them. For instance, in the motivating example, if the length of time intervals is two months, these probability fluctuations about: 'Sydney' is high in this month and low in the next month, 'Perth' is low in this month and high in the next month, may not be expressed: they are about the same in the two-month period. Hence, for privacy protection at client side, noise obfuscation has to consider these time intervals which are utilized by privacy attackers at server side. Briefly, time intervals that are too long may cause protection failing, whilst too short may cause unnecessary cost. In Section 3, we will discuss this in detail.

We will extend these time intervals to be flexible and generate them to withstand various privacy attackers for noise obfuscation and privacy protection. In summary, the contributions of this paper are:

1. We first investigate the fluctuations of occurrence probabilities which can jeopardize existing noise obfuscation and threat customer privacy, as the probability fluctuation privacy risk introduced before.
2. With the novel privacy risk model addressed analyze the withstanding between privacy attackers and privacy protectors (PP) in terms of time intervals and probability fluctuations for noise obfuscation.
3. The novel dynamic cluster based time interval generation algorithm (CTIG) for noise generation is presented to provide dynamic time intervals for probability forecasting and noise obfuscation.
4. Based on time intervals, the novel time-series pattern based forecasting algorithm (TPF) is proposed to abstract past probability fluctuations, and forecast future occurrence probabilities (probability fluctuations).
5. Our novel time-series pattern based noise generation strategy is presented to improve the effectiveness of privacy protection on noise obfuscation to withstand the probability fluctuation privacy risk based on the above model and algorithms in cloud environments.

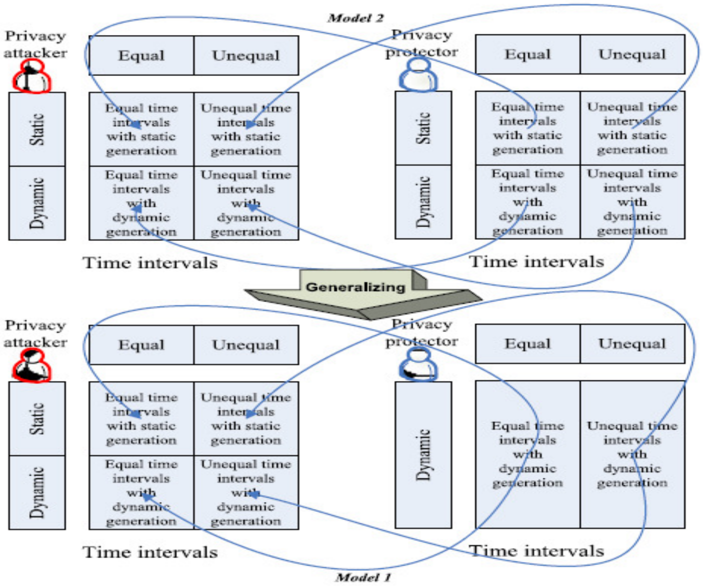
ABSTRACT MODEL OF PROPOSED SYSTEM

In architecture diagram from top to bottom the _gure displays a user issuing real queries R. The OB-PWS tool installed in the user's computer receives as input the users real queries R and automatically generates dummy queries D according to its dummy generation strategy DGS and associated semantic classification algorithm SCADGS. Both real and dummy queries are sent to the adversarial web search service provider, who (ideally) cannot distinguish them and thus are represented as Q. The observed profile is a representation of all Q queries according to some SCA of the adversary's choice. Further, the adversary can implement dummy classification DCA and profile filtering PFA algorithms that exploit vulnerabilities in the DGS. The former is used to classify queries Q as real QR or dummy QD, while the latter reverses the obfuscation introduced by the DGS in Y in order to obtain the filtered profile Z. The DCA and PFA are applied iteratively (using an SCA to translate queries to semantic categories) to both reduce the amount of noise and enhance the distinguish ability of real and dummy queries.

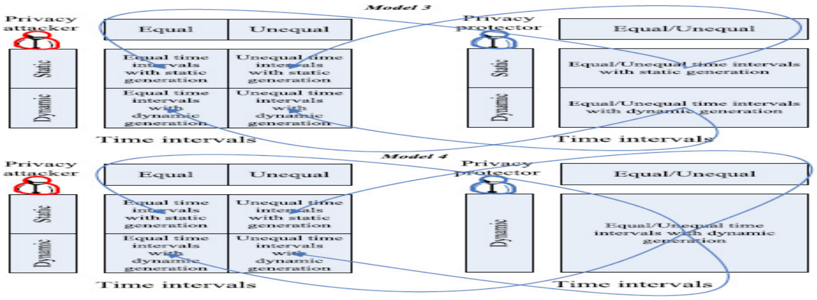


PRIVACY RISK MODEL

In this section, we discuss privacy risk models to analyze the withstanding between privacy attackers (PAs) and privacy protectors about time intervals and probability fluctuations in terms of noise obfuscation. In other words, as a thorough investigation of the probability fluctuation privacy risk compared to , we present the privacy risk models. It is the basis of our novel TPNGS .


Model 2
Privacy attacker
Equal
Unequal
Static
Dynamic
Equal time intervals with static generation
Unequal time intervals with static generation
Equal time intervals with dynamic generation
Unequal time intervals with dynamic generation
Time intervals
Privacy protector
Equal
Unequal
Static
Dynamic
Equal time intervals with static generation
Unequal time intervals with static generation
Equal time intervals with dynamic generation
Unequal time intervals with dynamic generation
Time intervals
Generalizing
Privacy attacker
Equal
Unequal
Static
Dynamic
Equal time intervals with static generation
Unequal time intervals with static generation
Equal time intervals with dynamic generation
Unequal time intervals with dynamic generation
Time intervals
Privacy protector
Equal
Unequal
Dynamic
Equal time intervals with dynamic generation
Unequal time intervals with dynamic generation
Time intervals
Model 1

In real cloud environments, PPs are difficult to recognize PAs' actions. Hence, one noise obfuscation strategy has to face various possible PAs with different methods of time interval generation discussed before, concurrently. Specifically, based on Model 1, we will evaluate our novel strategy and design corresponding processes to illustrate various
cases: four types of PAs with four methods for time interval generation: 1) ES—equal time intervals with static generation, 2) ED—equal time intervals with dynamic generation,
3) US—unequal time intervals with static generation and 4) UD—unequal time intervals with dynamic generation.


Model 3
Privacy attacker
Equal
Unequal
Static
Dynamic
Equal time intervals with static generation
Unequal time intervals with static generation
Equal time intervals with dynamic generation
Unequal time intervals with dynamic generation
Time intervals
Privacy protector
Equal/Unequal
Static
Dynamic
Equal/Unequal time intervals with static generation
Equal/Unequal time intervals with dynamic generation
Time intervals
Model 4
Privacy attacker
Equal
Unequal
Static
Dynamic
Equal time intervals with static generation
Unequal time intervals with static generation
Equal time intervals with dynamic generation
Unequal time intervals with dynamic generation
Time intervals
Privacy protector
Equal/Unequal
Dynamic
Equal/Unequal time intervals with dynamic generation
Time intervals

For PAs, there are two main types of time intervals: equal and unequal, and two main types of time interval generation: static and dynamic. Equal time intervals mean that all time intervals have the same length, such as one month for all. And unequal time intervals mean that time intervals have different lengths, such as three months, half a month and so on. Static time intervals mean that time intervals are generated by pre-setting without considering runtime service data. And dynamic time intervals mean that time intervals are generated dynamically, depended on runtime service data. Therefore, for PAs, there are four types of methods for time interval generation: 1) ES—equal time intervals with static generation, 2) ED—equal time intervals with dynamic generation, 3) US—unequal time intervals with static generation and 4) UD—unequal time intervals with dynamic generation. In other words, there are four kinds of PAs. Therefore, to withstand these PAs, PPs have to consider them all for noise obfuscation. In Fig we introduce privacy risk models to discuss the withstanding between PPs and PAs in terms of time intervals for noise obfuscation, under the probability fluctuation privacy risk. In other words, in time interval generation for noise obfuscation and privacy protection, PPs can utilize some methods to withstand PAs' methods.

NOVEL PRIVACY RISK MODEL AND ITS SPECIFIC CASE

In the cloud, PPs are impossible to settle down PAs' static time intervals, or specific methods of dynamic generation. Hence, PPs have to analyze time intervals by themselves. In other words, they need to investigate past occurrence probabilities, and generate time intervals from them. It is reasonable that these time intervals should express probability fluctuations sufficiently. As described by Model 1 in Fig. , this idea can utilize the dynamic generation to withstand PAs' static and dynamic generations. Specifically, we will introduce a dynamic time interval generation algorithm (CTIG), which is a fundamental part of our novel TPNGS. Based on the results of CTIG, PPs can utilize time intervals to describe probability fluctuations, and support time-series pattern forecasting to guide noise generation. Briefly, Model 1 reflects our novel privacy risk model and the primary task of this paper. Hence, we obtain Model 1 in which SPA is the set of privacy attackers:
{ES, ED, US, UD} and SPP is the similar set of privacy protectors: {ED, UD}

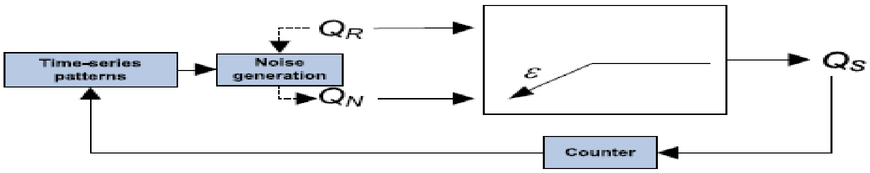
{ES,ED,US,UD}=> {US,UD}… Model 1 SPA maps SPP (1).

3Besides, compared to Model 1, Model 2 is a specific and idealized case which requires that: PPs know PAs' static time intervals, or specific methods of dynamic generation.
As described by Model 2 in equation (2), based on PAs' time intervals, PPs can 'accurately' utilize them to describe probability fluctuations directly, and support time-series pattern forecasting to guide noise generation. And the withstanding is simple, compared to Model 1.

{ES, ED, US, UD} =>{ES,ED,US, UD}.. Model 2 SPA maps SPP (2).

NOISE INJECTION MODEL

Our time-series pattern based noise injection model is modified to fulfill our time-series pattern idea as shown in Fig.
QR: queue of customer's real service requests to be protected.
QN: queue of noise service requests to be injected in QR.
QS: queue of final service requests composing of QR and QN.



Q: a set of service requests, and Q = {q1,q2,…,qi,…,qn} ; qng. Every service request in QR;QS and QN is from this set. Hence, in the view of service providers, one request in the queue of final service request QS could be from real requests QR or noise requests QN.

CONCLUSION & FUTURE WORK

In open and virtualized cloud environments, some malicious service providers may focus on Customer service data and collectively deduces customer privacy without permission. Noise obfuscation is an effective approach in this regard. For example, it generates and injects noise service requests into real ones to ensure that their occurrence probabilities are about the same so that service providers cannot distinguish which requests are real ones. However, existing representative noise generation strategies have not considered occurrence probability fluctuations. In fact, such occurrence probabilities could fluctuate at some time segments of the entire time period, which cannot be concealed by existing noise obfuscations. Hence, malicious service providers are still able to deduce customer privacy from these probability fluctuations. To address this probability fluctuation privacy risk, we developed a novel time series pattern based noise generation strategy for privacy protection on cloud.

In future, based on TPNGS, we plan to investigate how to protect customer privacy in the Scenario where multiple malicious service providers may collaborate with each other to threat noise obfuscation.

ACKNOLEDGEMENTS

We take this opportunity to thank our project guide Prof.A.M.Pawar and Head of the Department Prof.S.M.Sangve from ZCOER SPPU Pune for their valuable guidance and for providing all the necessary facilities, & their stimulating suggestions in this paper.

REFERENCES

[1] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, *Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.*

[2] M. Armbrust, A. Fox, R. Gri_th, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, *Above the Clouds: A Berkeley View of Cloud Computing, Comm. ACM, vol. 53, no. 6, pp. 50-58, 2010.*

[3] W. Jansen and G. Timothy, *Guidelines on Security and Privacy in Public Cloud Computing. National Inst. Standard and Technology, Special Publication 800-144, Dec. 2011.*

[4] G. Zhang, Y. Yang, X. Liu, and J. Chen, *A Time-Series Pattern Based Noise Generation Strategy for Privacy Protection in Cloud Computing, Proc. 12th IEEE/ACM Intl Symp. Cluster, Cloud and Grid Computing (CCGrid 12), pp. 458-465, May 2012.*

[5] B.C.M. Fung, K. Wang, R. Chen, and P.S. Yu, *Privacy-Preserving Data Publishing: A Survey of Recent Developments, ACM Computing Surveys, vol. 42, no. 4, pp. 1-53, 2010.*

[6] C.A. Ardagna, M. Cremonini, S. de Capitani di Vimercati, and P. Samarati, *An Obfuscation- Based Approach for Protecting Location Privacy, IEEE Trans. Dependable and Secure Computing, vol. 8, no. 1, pp. 13-27, Jan./Feb. 2011.*

[7] E. Balsa, C. Troncoso, and C. Diaz, *OB-PWS: Obfuscation-Based Private Web Search, Proc. IEEE Symp. Security and Privacy, pp. 491-505, May 2012.*

[8] X. Liu, Z. Ni, D. Yuan, Y. Jiang, Z. Wu, J. Chen, and Y. Yang, *A Novel Statistical Time-Series Pattern Based Interval Forecasting Strategy for Activity Durations in Worked Systems, J. Systems and Software, vol. 84, no. 3, pp. 354-376, 2011.*

[9] S. Ye, F. Wu, R. Pandey, and H. Chen, *Noise Injection for Search Privacy Protection, Proc. Intl Conf. Computational Science and Eng. (CSE 09), pp. 1-8, Aug. 2009.*

[10] C. Wang, S.S.M. Chow, Q. Wang, K. Ren, and W. Lou, *Privacy- Preserving Public Auditing for Secure Cloud Storage, IEEE Trans. Computers, vol. 62, no. 2, pp. 362-375, Feb. 2013.*