IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 10-13
www.iosrjournals.org

Privacy-Preserving Data Leak Detection

Nitin Naik[1], Aniket Nikam[2], Narendra Patil[3] &K.V.Ugale[4]
[1,2,3,4]*(Dept of Computer Engineering, KVNNIEER ,SPPU,(MS), India)*

Abstract : *Among différent data-leakage examples, man-like mistakes are one of the main causes of data loss. There have existence answers sensing indentent sensitive information for computer leak cause by man like mistake and to make ready for organisations. A common move near by is to screen what is in place for storing and send (power and so on) for made open to sensitive information. Such a move near usually has need of the detecting operation to be guided in secrecy. However, this secretiveness thing need is hard to give what is desired, need to in familiarization, as discovery serve may be put at risk. In this paper, we present a right not to be public keeping harmless data-leak discovery (OLD) answer to get answer to the question under expansion where a special group of sensitive knowledge for computer goes through process of digestion is use in discovery. The better chances of our elaborated way is that it enables the facts publisher to safely give powers the discovery operation to an almost upright, true given without letting be seen the sensible knowledge for computers to the given. We make, be moving in how Internet public alliance provides can over their customers OLD as an add-on public organization with strong right not to be public gives support to a proposition.*

Keywords: *Data Leak, Network Security, privacy collection intersection.*

I. Introduction

According to a report from Risk base Security (RBS), the number of leak sensitive data records has greater than before radically during the previous few years, i.e. from 412 million in 2012 to 822 million in 2013. Intentionally planned attacks, unintentional leaks (e.g., forward secret emails to unspecified email accounts), and human mistake (e.g., assigning the incorrect privilege) lead to most of the data-leak incident. Detecting and prevent data leaks requires a set of complementary solution, which may include data-leak detection, data imprisonment, stealthy malware detection, and rule enforcement. Network data-leak detection (DLD) classically performs deep packet examination (DPI) and searches for any occurrence of sensitive data patterns. DPI is techniques to examine payloads of IP/TCP packets for inspect application layer data, e.g., HTTP header/content. Alert are trigger when the amount of sensitive data establish in traffic passes a threshold. The detection system canbe deploy on a router or included into existing network interruption detection system (NIDS). Simple realizations of data-leak detection require the plain text sensitive data. though, this requirement is unwanted, as it may threaten the privacy of the sensitive information. If a detection system is compromise, then it may expose the plain text sensitive data (in memory). In adding, the data owner may need to subcontract the data-leak detection to providers, but may be unwilling to reveal the plain text sensitive data to them. so, one needs new data-leak detection solutions that allow the provider to scan content for leaks without learning the sensitive information.

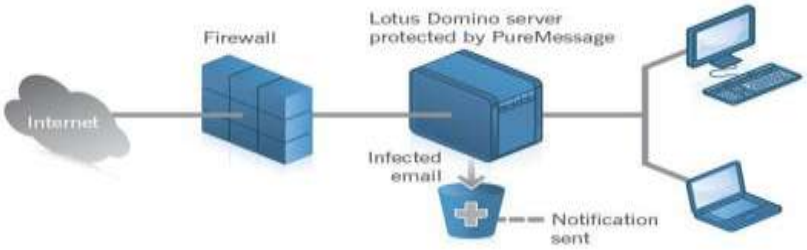
In that Paper, we propose a data-leak detection solution which can be outsourced and be deploy in a half honest detection surroundings. We design, apply, and evaluate our fuzzy fingerprint method that enhances data privacy during data- leak detection operations. Our approach is based on a fast and practical one-way calculation on the sensitive data (SSN records, classified papers, sensitive emails, etc.). It enables the data owner to firmly delegate the content-inspection task toDLD provider without exposing the sensitive data. Using our detection technique, the DLD provider, who is modeled as an honest-but-curious (aka semi-honest) ad-adversary, can only gain limited knowledge about the sensitive data from either the free digests, or the content being inspected. Using our technique, an Internet service provider (ISP) can perform detection on its customers transfer securely and provide data-leak detection as an add-on service for its customers. In another situation, individuals can mark their own sensitive data and ask the manager of their local network to detect data leaks for them. In our detection process, the data owner compute a special set of digests or fingerprints from the sensitive data and then disclose only a small amount of them to the DLD provider. The DLD provider compute fingerprints from network traffic and identify potential leaks in them. To prevent the DLD provider from gathering exact knowledge about the sensitive data, the collection of potential leaks is composed of leaks and noises. It is the data owner, who post-processes the potential leak sent back by the DLD provider and determines whether there is any real data leak.

IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 10-13
www.iosrjournals.org

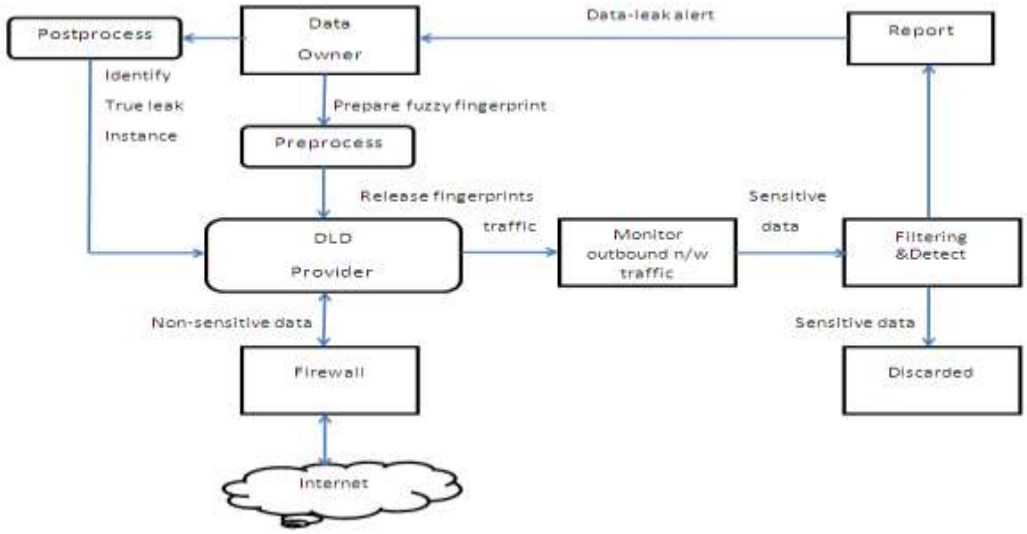
II. System Design

2.1 Existing System

Traditionally, leakage detection is handled by in watermarking, e.g., a unique code is embedded in each distributed copy. If that copy is a later discovered in the hands of an unauthorized party, the leaker can be identified. Watermarks can be very important in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be delete if the data recipient is malicious. E.g. A hospital may give patient records to researchers who will advise new treatments. Similarly, the company may have tie ups with other companies that require customer information. Different enterprise may outsource its information process, so and information must be given to other companies. We call the owner of the information the distributor and the supposedly trusted 3rd parties the agents.



2.2 Proposed System



Our aim is to sense when the distributor's responsive information has been leakage by agent, and if like to identify the agents that leak the data. Perturbation is a extremely helpful and safe method where the data is customized and made less responsive before being hand to agent. We develop up unobtrusive technique for detect leak of a set of objects or records. In this part we develop a model for assess the "Errors" of agent. We also present algorithms for distribute substance to agents, a way that improve our modification of identifies a leaker. At last, we also think the option of addition "Error" the substance to distributes set. Such substances do not correspond to real entity but appear sensible to the agents. In a sense, the fake substance act as a kind of watermark for the whole set, with leakage any single member. If it turn out an agent was given one or many fake substance that were leak, then the dispenser can be more guarantee that agent was responsible s.

2.3 Algorithm: SHINGLES ALGORITHM

A) Shingle:-
1) Start
2) A shingle window is used to generate q-grams on an input binary string first.
3) The fingerprints of q-grams are then computed.
4) Local feature preservation is accomplished through the use of shingles.
5) stop .

Rabin Fingerprint Algorithm:
1) Start
2) In fingerprinting, each shingle is treated as polynomial q(x).
3) Each coefficient of q(x) i.e. Ci(0 ¡ I ¡ k) is one bit in the shingle.
4) q(x) is mod by a selected irredvisble polynomial p(x).F=C1 xk-1+C2xk-2+..+Ck-1x+Ck mod p(x) It maps a k-bit shingle into apf bit fingerprint f where the degree of p(x) is pf+1.
5) If there is a data leak, there is match between two fingerprints from sensitive data and network traffic.
6) Stop.

2.4 Application:

1.Online banking system:- when we enter the data on on line system this has to fake page ,then aware from intruders
2.Data leakage system:- if we send data in network then this can not be leak to anywhere.
3.Store securely medical documents:- security provide to personal medical report
4.Online shopping:- aware from hacker's which is make the fake site's
5.In mall shopping:- secure ATM card and provide security to banking account

III. Result Analysis

3.1 Homepage

Distributor Login:-

3.2 Agent Login

IV. Conclusion

We planned fuzzy fingerprints, the privacies preserving data leak detection model and there it's accepting. Using particular digest, the contact of the responsive data is kept to a lowest during the detection. We have conducted extensive test to authenticate the correctly, confidentiality, and effectiveness of our solution. For prospect job, we plan to focus on architecture a hosts assisted technique for the all information leak detection for main scale organization.

Acknowledgements

It is my immense pleasure to work this Project Report on"Privacy-Preserving DataLeak Detection". It is only the blessing of my divine master which has prompted and mentally equipped me to undergo the study of this seminar. I would like to thank Dr.A.K.Dwivedi, Principal, K.V.N.Naik Institute of Engineering Education, Research. for giving me such an opportunity to develop practical knowledge about subject. I am also thankful to Prof. K. V. Ugale, Head of Computer Engineering Department for his valuable encouragement at every phase of my Project Report work and completion. I over my sincere thanks to my guide Prof. K. V. Ugale, who very affectionately encourages me to work on the subject and gave his valuable guidance time to time. While preparing this seminar I am very much thankful to him. I am also grateful to entire staff of Computer Engineering Department for their kind co-operation which helped me in successful completion of seminar.

References

[1] J. Jung, A. Sheth, B. Greenstein, D. Wetherall, G. Maganis, and T.Kohno, "Privacy oracle: A system for finding application leaks with black box differential testing", in Proc. 15th ACM Conf. Comput. Commun.
[2] Secur., 2008, pp. 279288
[3] A. Z. Broder, Some applications of Rabins fingerprinting method, in Sequences II. New York, NY, USA: Springer-Verlag, 1993, pp. 143152.
[4] X. Shu and D. Yao, "Data leak detection as a service", in Proc. 8th Int. Conf. Secur. Privacy Commun. Netw., 2012, pp. 222240.
[5] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher," Minwise independent permutations", J. Comput. Syst. Sci., vol. 60, no. 3, pp. 630659, 2000.
[6] K. Borders, E. V. Weele, B. Lau, and A. Prakash," Protecting confidential data on personal computers with storage capsules", in Proc. 18Th USENIX Secur. Symp., 2009, pp. 367382.
[7] H. Yin, D. Song, M. Egele, C. Kruegel, and E. Kirda, "Panorama: Capturing system-wide information flow for malware detection and analysis", in Proc. 14th ACM Conf. Comput. Commun. Secur., 2007, pp. 116127.
[8] Y. Jang, S. P. Chung, B. D. Payne, and W. Lee, Gyrus: "A framework for user-intent monitoring of text-based networked applications", in Proc. 23rd USENIX Secur. Symp., 2014, pp. 7993.
[9] G. Karjoth and M. Schunter, "A privacy policy model for enterprises", in Proc. 15th IEEE Comput. Secur. Found. Workshop, Jun. 2002, pp. 271281