International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

Lime-A Generic Data Lineage Framework In Malicious Environment

P. Swathi & V.Janaki
M.Tech,Department of CSE, Vaagdevi College of Engineering, Bollikunta, Warangal, Telangana, Mail id: swathi.peddapati@gmail.com

Professor,Department of CSE, Vaagdevi College of Engineering, Bollikunta ,Warangal, Telangana, Mail ID: janaki.vaag@gmail.com,

ABSTRACT:*Deliberate or unexpected spillage of secret information is without a doubt a standout amongst the most extreme security dangers that associations look in the advanced period. The danger now reaches out to our own lives: a plenty of individual data is accessible to interpersonal organizations and cell phone suppliers and is in a roundabout way exchanged to dishonest outsider and fourth gathering applications. In this work, we introduce a bland information genealogy system LIME for information stream over numerous elements that take two trademark, key parts (i.e., proprietor and purchaser). We characterize the correct security ensures required by such an information heredity component toward distinguishing proof of a blameworthy substance, and recognize the disentangling non-disavowal and genuineness suppositions. We at that point create and break down a novel responsible information exchange convention between two elements inside a malevolent domain by expanding upon careless exchange, powerful watermarking, and mark primitives. At long last, we play out an exploratory assessment to exhibit the reasonableness of our convention and apply our system to the vital information spillage situations of information outsourcing and interpersonal organizations. By and large, we consider LIME , our genealogy structure for information exchange, to be a key advance towards accomplishing responsibility by plan.*

1 INTRODUCTION

IN the digital era, information leakage through unintentional exposures, or intentional sabotage by disgruntled employees and malicious external entities, present one of the most serious threats to organizations. According to an interesting chronology of data breaches maintained by the Privacy Rights Clearinghouse (PRC), in the United States alone, 868;045;823 records have been breached from 4;355 data breaches made public since 2005 [1]. It is not hard to believe that this is just the tip of the iceberg, as most

cases of information leakage go unreported due to fear of loss of customer confidence or regulatory penalties: it costs companies on average $214 per compromised record [2]. Large amounts of digital data can be copied at almost no cost and can be spread through the internet in very short time. Additionally, the risk of getting caught for data leakage is very low, as there are currently almost no accountability mechanisms. For these reasons, the problem of data leakage has reached a new dimension nowadays. Not only companies are affected by data leakage, it is also a concern to individuals. The rise of social networks and smartphones has made the situation worse. In these environments, individuals disclose their personal information to various service providers, commonly known as third party applications, in return for some possibly free services. In the absence of proper regulations and accountability mechanisms, many of these applications share individuals' identifying information with dozens of advertising and Internet tracking companies. Even with access control mechanisms, where access to sensitive data is limited, a malicious authorized user can publish sensitive data as soon as he receives it. Primitives like encryption offer protection only as long as the information of interest is encrypted, but once the recipient decrypts a message, nothing can prevent him from publishing the decrypted content. Thus it seems impossible to prevent data leakage proactively. Privacy, consumer rights, and advocacy organizations such as PRC [3] and EPIC [4] try to address the problem of information leakages through policies and awareness. However, as seen in the following scenarios the effectiveness of policies is questionable as long as it is not possible to provably associate the guilty parties to the leakages.

2 RELATED WORK

A preliminary shorter version of this paper appeared at the STM workshop . This version constitutes a significant extension by including the following contributions: We give a more detailed description of our model, a formal specification of the used primitives, an analysis of the introduced protocol, a discussion of implementation results, an application of our framework to example scenarios, a discussion of additional features and an extended discussion of related work. Clustering analysis is very useful to estimate the inter-entity similarity. One good example

of clustering based reranking algorithms is the Information Bottle based scheme developed by Hsu et al.[9]. In this method, the images in the initial results are primarily grouped automatically into several clusters. Then the re-

ranked result list is created first by ordering the clusters according to the cluster conditional probability and next by ordering the samples within a cluster based on their cluster membership value. In a fast and accurate scheme is proposed for grouping Web image search results into semantic clusters. It is obvious that the clustering based reranking methods can work well when the initial search results contain many near duplicate media documents. However, for queries that return highly diverse results or without clear visual patterns, the performance is not guaranteed.

3 THE LIME FRAMEWORK

As we want to address a general case of data leakage in data transfer settings, we propose the simplifying model LIME (Lineage in the malicious environment). With LIME we assign a clearly defined role to each involved party and define the inter-relationships between these roles. This allows us to define the exact properties that our transfer protocol has to fulfill in order to allow a provable identification of the guilty party in case of data leakage.

3.1 Model

As LIME is a general model and should be applicable to all cases, we abstract the data type and call every data item document. There are three different roles that can be assigned to the involved parties in LIME: data owner, data consumer and auditor. The data owner is responsible for the management of documents and the consumer receives documents and can carry out some task using them. The auditor is not involved in the transfer of documents, he is only invoked when a leakage occurs and then performs all steps that are necessary to identify the leaker. All of the mentioned roles can have multiple instantiations when our model is applied to a concrete setting. We refer to a concrete instantiation of our model as scenario. In typical scenarios the owner transfers documents to consumers. However, it is also possible that consumers pass on documents to other consumers or that owners exchange documents with each other. In the outsourcing scenario [6] the employees and their employer are owners, while the outsourcing companies are untrusted consumers. In the following we show relations between the different entities and introduce optional trust assumptions. We only use these trust assumptions because we find that they are realistic in a real world scenario and because it allows us to have a more efficient data transfer in our framework. At the end of this section we explain how our framework can be applied without any trust assumptions. When documents are transferred from one owner to another one, we can assume that the transfer is governed by a non-repudiation assumption. This

means that the sending owner trusts the receiving owner to take responsibility if he should leak the document. As we consider consumers as untrusted participants in our model, a transfer involving a consumer cannot be based on a non-repudiation assumption. Therefore, whenever a document is transferred to a consumer, the sender embeds information that uniquely identifies the recipient. We call this fingerprinting. If the consumer leaks this document, it is possible to identify him with the help of the embedded information. As presented, LIME relies on a technique for embedding identifiers into documents, as this provides an instrument to identify consumers that are responsible for data leakage. We require that the embedding does not not affect the utility of the document. Furthermore, it should not be possible for a malicious consumer to remove the embedded information without rendering the document useless. A technique that can offer these properties is robust watermarking. We give a definition of watermarking and a detailed description of the desired..

4 ACCOUNTABLE DATA TRANSFER

In this section we specify how one party transfers a document to another one, what information is embedded and which steps the auditor performs to find the guilty party in case of data leakage. We assume a public key infrastructure to be present, i.e., both parties know each others signature verification key.

4.1 Trusted Sender

In the case of a trusted sender it is sufficient for the sender to embed identifying information, so that the guilty party can be found. As the sender is trusted, there is no need for further security mechanisms. we present a transfer protocol that fulfills the properties of correctness and no denial as. As the sender is trusted to be honest, we do not need the no framing property. The sender, who is in possession of some document D, creates a watermarking key k, embeds a triple consisting of the two parties' identifiers and a timestampt into D to create Dw ¼WðD; s; kÞ. He then sends Dw to the recipient, who will be held accountable for this version of the document. As the sender also knows Dw, this very simple protocol is only applicable if the sender is completely trusted; otherwise the sender could publish Dw and blame the recipient.

4.2 Untrusted Sender

In the case of an untrusted sender we have to take additional actions to prevent the sender from cheating, i.e., we have to fulfill the no framing property. To achieve this property, the sender divides the original document into n parts and for each part he creates two differently watermarked versions. He then transfers one of

each of these two versions to the recipient via OT2

1 . The recipient is held accountable only for the document with the parts that he received, but the sender does not know which versions that are. The probability for the sender to cheat is therefore 12n. We show the protocol and provide an analysis of the protocol properties . First, the sender generates two watermarking keys k1 and k2. It is in his own interest that these keys are fresh and distinct. The identifying information that the sender embeds into the documentD is a signed statement s ¼ ½CS; CR; t_skCR containing the sender's and recipient's identifiers and a timestamp t, so that every valid watermark is authorized by the recipient. The sender computes the watermarked document splits the document D0 into n parts and creates two different versions

4.3 Data Lineage Generation

The auditor is the entity that is used to find the guilty party in case of a leakage. He is invoked by the owner of the document and is provided with the leaked document. In order to Protocol for trusted senders: The sender watermarks the original document with a signed statment containing the participants' identifiers and a timestamp, furthermore, sends the watermarked archive to the beneficiary.

locate the liable party, the examiner continues in the accompanying way:

1) The reviewer at first takes the proprietor as the present suspect.

2) The reviewer attaches the present suspect to the heredity.

3) The inspector sends the spilled report to the momentum suspect and requests that he give the location keys k1 and k2 for the watermarks in this archive and in addition the watermark s. On the off chance that a non-daze watermarking plan is utilized, the examiner furthermore asks for the unmarked variant of the report.

4) If, with key k1, s can't be recognized, the examiner proceeds with 9.

5) If the present suspect is believed, the evaluator watches that s is of the frame where CS is the identifier of the present suspect, takes CR as present suspect and proceeds with 2.

6) The reviewer checks that s is of the frame ½CS; CR; t_skCR where CS is the identifier of the present suspect. He likewise checks the legitimacy of the mark.

7) The examiner parts the archive into n parts and for each part he tries to identify 0 and 1 with key k2. On the off chance that none of these or both of these are perceptible, he proceeds with 9. Else he sets b0i as the identified piece for the ith part. He sets b0 ¼ b01 . . . b0n.

8) The inspector solicits CR to demonstrate his decision from b ¼ b1 _ bn for the given timestamp t by displaying the . On the off chance that CR can't give a right verification (i.e., mi;bi is of the wrong shape or the mark is invalid) or if b ¼ b0, at that point the inspector takes CR as present suspect and proceeds with 2.

9) The examiner yields the ancestry. The last passage is in charge of the spillage.

CONCLUSION AND FUTURE DIRECTIONS

We exhibit LIME, a model for responsible information exchange over various substances. We characterize taking part parties, their between connections and give a solid instantiation for an information exchange convention utilizing a novel mix of neglectful exchange, powerful watermarking and computerized marks. We demonstrate its rightness and demonstrate that it is feasible by giving microbenchmarking comes about. By displaying a general relevant system, we present responsibility as ahead of schedule as in the plan period of an information exchange foundation. In spite of the fact that LIME does not effectively anticipate information spillage, it presents receptive responsibility. In this way, it will prevent malevolent gatherings from releasing private archives and will empower legitimate (yet imprudent) gatherings to give the expected assurance to delicate information. LIME is adaptable as we separate between put stock in senders (typically proprietors) and untrusted senders (normally buyers). On account of the put stock in sender, an extremely straightforward convention with minimal overhead is conceivable. The untrusted sender requires a more muddled convention, yet the outcomes are not founded on trust suppositions and along these lines they ought to have the capacity to persuade an impartial substance (e.g., a judge). Our work likewise spurs additionally examine on information spillage identification systems for different report sorts and situations. For instance, it will be an intriguing future research heading to plan an obvious genealogy convention for inferred information..

REFERENCES

[1] Chronology of data breaches [Online]. Available: http://www. privacyrights.org/data-breach, 2014.

[2] Data breach cost [Online]. Available: http://www.symantec. com/about/news/release/article.jsp?prid=20110308_01, 2011.

[3] Privacy rights clearinghouse [Online]. Available: http://www. privacyrights.org, 2014.

[4] (1994). Electronic privacy information center (EPIC) [Online]. Available: http://epic.org, 1994.

[5] Facebook in privacy breach [Online]. Available: http://online.wsj. com/article/SB10001424052702304772804575558484075236968. html, 2010.

[6] Offshore outsourcing [Online]. Available: http://www.computerworld. com/s/article/109938/Offshore_outsourcing_cited_in_Florida_data_leak, 2006.

[7] A. Mascher-Kampfer, H. St€ogner, and A. Uhl, "Multiple re-watermarking scenarios," in Proc. 13th Int. Conf. Syst., Signals, Image Process., 2006, pp. 53–56.

[8] P. Papadimitriou and H. Garcia-Molina, "Data leakage detection," IEEE Trans. Knowl. Data Eng., vol. 23, no. 1, pp. 51–63, Jan. 2011.

[9] Pairing-based cryptography library (PBC) [Online]. Available: http://crypto.stanford.edu/pbc, 2014.

[10] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," IEEE Trans. Image Process., vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[11] B. Pfitzmann and M. Waidner, "Asymmetric fingerprinting for larger collusions," in Proc. 4th ACM Conf. Comput. Commun. Security, 1997, pp. 151–160.

[12] S. Goldwasser, S. Micali, and R. L. Rivest, "A digital signature scheme secure against adaptive chosen-message attacks," SIAM J. Comput., vol. 17, no. 2, pp. 281–308, 1988.

[13] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi, "A computational model for watermark robustness," in Proc. 8th Int. Conf. Inf. Hiding, 2007, pp. 145–160.

[14] J. Kilian, F. T. Leighton, L. R. Matheson, T. G. Shamoon, R. E. Tarjan, and F. Zane, "Resistance of digital watermarks to collusive attacks," in Proc. IEEE Int. Symp. Inf. Theory, 1998, pp. 271–271.

[15] M. Naor and B. Pinkas, "Efficient oblivious transfer protocols," in Proc. 12th Annu. ACM-SIAM Symp. Discrete Algorithms, 2001, pp. 448–457.

[16] GNU multiple precision arithmetic library (GMP) [Online]. Available: http://gmplib.org/, 2014.

[17] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the Weil pairing," in Proc. 7th Int. Conf. Theory Appl. Cryptol. Inf. Security: Adv. Cryptol., 2001, pp. 514–532.

[18] W. Dai. Crypto++ Library [Online]. Available: http://cryptopp. com, 2013.

[19] P. Meerwald. Watermarking toolbox [Online]. Available: http:// www.cosy.sbg.ac.at/ pmeerw/Watermarking/source, 2010.

[20] Y. Ishai, J. Kilian, K. Nissim, and E. Petrank, "Extending oblivious transfers efficiently," in Proc. 23rd Annu. Int. Cryptol. Conf. Adv. Cryptol., 2003, pp. 145–161.