Future Generation Computer Systems 72 (2017) 288–304

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Towards an asynchronous aggregation-capable watermark for end-to-end protection of big data streams

Arezou Soltani Panah [a,*], Ron van Schyndel [a], Timos Sellis [b]

[a] Department of Computer Science and Information Technology, RMIT University, Melbourne, VIC 3001, Australia
[b] Swinburne University of Technology, Melbourne, VIC 3122, Australia

HIGHLIGHTS

- We present a novel low complexity watermarking scheme for data streams integrity,
- Multiple signatures can be added to the data stream(s) in an asynchronous fashion,
- The verification process does not need time-synchronization for successful decoding,
- Our method can be coupled with a transport layer security to afford maximal protection.

ARTICLE INFO

Article history:
Received 15 November 2015
Received in revised form
1 August 2016
Accepted 1 September 2016
Available online 10 September 2016

Keywords:
Tamper proofing methods
Sensor networks
Data aggregation
Asynchronous data streams
Digital watermarking
Pseudorandom sequences

ABSTRACT

We consider how an untrusted data aggregator can be assessed over multiple data streams. The aggregator could be the sink node in a sensor network where all the sensory data are gathered, or a smart-meter responsible for computing power measurements of a group of households, or any other entity that is basically in charge of answering aggregation queries such as average or summation in a data streaming environment. In these applications, important decisions are made based on the aggregated results and therefore, it is vitally important to investigate the authenticity and integrity of aggregated values. One possible approach for solving this problem is marking the data before sending it out to the aggregators (i.e. marked at the *point of origin*) such that the existence of those marks could be verified subsequently after the aggregation process. Our goal is to produce hidden marks that remain detectable after the aggregation and thereby not only the trustworthiness of every individual data source, but also the trustworthiness of the aggregators could be verified. This problem is referred to secure data aggregation that has been investigated by means of digital watermarking and steganography techniques in recent years. Data synchronization is a serious problem which was not addressed in the current schemes, though. Therefore, in this paper, a new watermarking construction is proposed that provides 'synchronization marks' in the aggregated data stream and helps protect the data itself at the end-points. Our method works at the data layer so standard transport layer security methods can be used to protect the transport of data if it is required. Finally, a set of experiments are conducted using synthesized and real sensory data as a proof of concept.

Crown Copyright © 2016 Published by Elsevier B.V. All rights reserved.

1. Introduction

Enormous quantities of data flow through today's computer networks every day. Often, it is necessary to analyze this massive volume of data and compute aggregates and statistics in order to detect events and trends as those values offer significant insights for data analysis purposes. The need for processing such a big dataset has led to the study of the data aggregation model of computation, where multiple data streams are combined together based on an aggregation function such as summation or average [1]. One application of this model is sensor networks where nodes are organized into a routing tree rooted at the base station to perform the in-network data aggregation for the purpose of energy efficiency [2]. Another example is smart-meter reading aggregation (either spatial or temporal) for computing the power consumption of a group of households [3].

On the other hand, it may be necessary that the authenticity of the aggregated data is verified at the final destination because

* Corresponding author.
E-mail address: arezou.soltanipanah@rmit.edu.au (A. Soltani Panah).

http://dx.doi.org/10.1016/j.future.2016.09.001
0167-739X/Crown Copyright © 2016 Published by Elsevier B.V. All rights reserved.

vital decisions are going to be made based on the final results. This problem is usually referred to as 'secure data aggregation' in sensor networks [2]. The aggregate–commit–prove method proposed by Przydatek et al. [4] also shares the same concept in which the aggregator not only performs the aggregation operation, but also proves the correctness of the performed task. A promising solution to this problem is marking data with secret and robust digital watermarks.

Digital watermarking is a proven technique usually in multimedia domain for copyright protection [5]. Recently, there has been an explosion in non-media applications of digital watermarking among which are time-series, biological sequences, graph-structured, spatial, spatiotemporal and data streams [6]. Regardless of content type, watermarking aims at embedding (secret) information *within* the original content such that it can be detected or extracted later to make assertions about the content. From the security perspective, cryptographic techniques can do better compared to digital watermarking counterparts, but they impose large computational overhead because decryption is necessary before performing further computation on the encrypted data and therefore these methods become unsuitable for resource-constrained environments in terms of energy efficiency [2].

In our previous work [7], we proposed a composite watermark-based solution for the secure data aggregation problem by embedding of several (near) orthogonal patterns into multiple data streams. Compared to similar works such as [2,8], our approach has the batch detection feature which means the existence of individual watermarks can be investigated by means of just one detection operation for the aggregated data, instead of one per stream and therefore the detection complexity is decreased. This is a great benefit when the network size increases. The scheme that we used for marking several data streams is based on the Spread Spectrum (SS) watermarking technique and is usually applied in aggregation supportive environments because of its characteristics such as simple embedding operations and resilience against various transformations [5].

The essence of SS watermarking is the addition of a pseudo-random noise pattern to the acquired data and the authenticity of data can be verified by finding the correlation between the watermarked data and embedded pattern. The host data could be either a single data stream or multiple data streams. For the latter case, it is possible that every intermediate node along the data path, 'sign off' the data en-route before forwarding to the next node by embedding its own unique signature within the aggregated data. This way an audit trail of the data processing is provided that is referred to as data provenance. This concept of data self-audit is crucial for assurance of data trustworthiness [8]. Apart from its good features, our previous method necessitates the existence of exact timing information for decoding purposes because the watermarks are embedded in absolute positions/times of the data streams and therefore a de-synchronization attack [9] could potentially invalidate the detection process.

De-synchronization attack is one of the most effective attacks against SS watermarking in which an attacker does not directly remove or jam a watermarked signal, but instead shifts and wraps the watermarked content so that it is no longer recognizable to the decoder. Let us assume the watermark w is added to the host signal x in order to construct the tamper-proof signal y as $y_i = x_i + scale(w_j)$, where $scale()$ updates the watermark amplitude to make it imperceptible from the noise floor of host signal as described [7]. The i and j shows the sequence numbers in y and w respectively. If the watermark synchronization point matches the data synchronization point, i is equal to j. It is possible that an attacker cyclically rotates the watermarked signal y by θ, so those indexes do not match any more, i.e. $|i - j| = \theta$ and it is very likely that the watermark to be missed at the decoder.

Our previous work had the restriction that θ is a constant value and is known to the decoder. In this work, we do not make this assumption anymore. Instead θ is floating and is unknown to the decoder. Moreover, if the watermarked signal y constitutes multiple watermarks, it is possible that individual watermarks experience different delays; so the offset θ could be different for each of them and therefore the watermark detection becomes a complex task. The goal of this paper is to extend the previous scheme with respect to de-synchronization attacks such that the decoder still be able to read multiple watermarks when they are floating in respect to the data.

The problem of designing watermarks that are resilient to de-synchronization attacks has been looked at before. However, those schemes are not directly applicable for our scenario. The reason is that for a composite watermark, the information embedded by multiple data sources could become nonaligned, not only due to lack of synchronization between encoder and decoder, but also because of the 'joining' of multiple asynchronous streams. The asynchrony could happen due to reasons such as sensor clock de-synchronization, jitter or sensor malfunction. Whenever, there is a need for reasoning about a time-varying environment such as a sensor network, the sensor observations must be time-stamped inside of the network. The in-network times-stamping implies that sensors themselves must be time-synchronized; otherwise, there will be ambiguity in reasoning about the environment.

The majority of the proposed aggregation protocols make an implicit assumption that a node's clock is kept synchronized all the time. This assumption comes from the fact that applications are not concerned with the details of synchronization and they simply expect a 'correct' clock to be available at any instant [10]. In the following section, we argue that, this assumption is fundamentally flawed and needs to be considered for the secure data aggregation problem as well. To the best of our knowledge, no one has looked at this constraint for the watermarked-based data aggregations.

One might argue that if multiple data streams are not synchronized, the whole application would fail. So, why it is still important to retrieve the watermarks? Our argument is that the amount of time lags/offsets among multiple data streams cannot be a large number because of period synchronization for sure. However, the upper bound of the offset depends on the synchronization interval and the precision of the synchronization scheme. Due to energy constraints limits, it is difficult to perform fine-grained time synchronization in sensor networks [11] and therefore the small value of misalignment in the aggregated stream can be acceptable by the application. As such, we attempt to ensure watermark survival when the offsets are relatively small. Section 2 elaborates on the details of the mentioned de-synchrony problem.

The approach that we take to this problem is by embedding two shifted watermark patterns instead of one for every data source. This way, the difference between those two shifted patterns remains the same even when the detection window leads or lags in respect to the encoding window timing. Unlike our previous work that we used one-dimensional (1D) patterns, in this work we use two-dimensional (2D) watermarks because of the following reasons:

1. It is easier to find multiple non-interfering binary watermarks in 2D space than 1D,
2. We used Distinct Sum Array construction [12] to extend the dimensionality of 1D sequences. This construction offers additional diversity and security over binary arrays, rendering them robust to cryptographic attack [13].
3. And finally, the 2D construction will increase the amount of payload information that one can embed, i.e. watermark capacity [14].

Download English Version:

https://daneshyari.com/en/article/4950248

Download Persian Version:

https://daneshyari.com/article/4950248

Daneshyari.com