This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2010 proceedings

SAS: A Secure Data Aggregation Scheme in Vehicular Sensing Networks

Qi Han†, Suguo Du†, Dandan Ren† and Haojin Zhu‡
† Management Science Department,
{hanqi, sgdu, rendandan}@sjtu.edu.cn
‡ Computer Science and Engineering Department,
Shanghai Jiao Tong University, Shanghai, China
zhu-hj@cs.sjtu.edu.cn

***Abstract*—Vehicular ad hoc networks support a wide range of promising applications including vehicular sensing networks, which enable vehicles to cooperatively collect and transmit the aggregated traffic data for the purpose of traffic monitoring. The reported literatures mainly focus on how to achieve the data aggregation in dynamic vehicular environment while the security issue especially on the authenticity and integrity of aggregation results receive less attention. In this study, we introduce a secure probabilistic data aggregation scheme based on Flajolet-Martin sketch and *sketch proof* technique. We also discuss the tradeoff between the bandwidth efficiency and the estimation accuracy. Extensive simulations and analysis demonstrate the efficiency and effectiveness of the proposed scheme.**

I. Introduction

With the advancement of wireless technology, vehicular communication networks, also known as Vehicular Ad Hoc Networks (VANETs), are emerging as a promising approach to increase road safety, efficiency and convenience. Although the primary purpose of vehicular networks is to enable communication-based automotive safety applications, e.g., cooperative collision warning (CCW), VANETs also allow a wide range of promising applications such as traffic monitoring and data collecting, which are regarded as an important component of future intelligent transportation systems (ITS).

As shown in [1], Departments of Transportation in the US must collect various types of data (e.g., average speed or traffic density) for traffic monitoring purposes. Traditionally, these important data are collected by technologies such as inductive loop detectors (ILD), video detection systems, acoustic tracking systems, or microwave radar sensors, which may be susceptible to failure or suffer from a high maintenance cost. On the other hand, cooperative data collection and dissemination in VANETs allow the traffic monitoring performed in a more cost-effective way [2]. Specifically, each vehicle collects its own or neighboring information (e.g., its current speed or neighboring traffic) and then transmits them to the remote roadside units (RSUs) via vehicle-to-vehicle(V2V) and vehicle-to-infrastructure(V2I) communications while the RSUs could be deployed at various points of interest along the roadway and can be used to collect data from locations up to tens of kilometers away. In this study, we coin the vehicular networks which are designed for traffic sensing and monitoring as the *vehicular sensing networks*.

One of the major challenges of vehicular sensing networks is high overhead of transmitted sensing data. Each sensing result is essentially some spatial-temporal measured values (speed, traffic density), which are determined jointly by the position of vehicles (i. e., a road segment or a small area) and the point in time when the observation has been made. Such sensing data is periodically broadcasted. Upon reception of such a broadcast, the intermediate receivers/forwarders incorporate the received data into their local reports, and then broadcast them again. Unfortunately, such a periodically boardcast introduces a high traffic load or even *traffic storm*. This problem is more serious in the scenario of high vehicle density, which could be found on multi-laned highways in congestion situations. On the other hand, in most cases, drivers or monitors do not need exact individual reports, but only an overview of the general average speed on the road ahead [3]. This motivates the data aggregation issues in vehicular networks, including Flajolet-Martin sketch based probabilistic aggregation [4], fuzzy aggregation [3] and others [5], [6]. However, most of them are mainly focusing on how to achieve the data aggregation in dynamic vehicular environment, while the security issues especially on how to make sure the authenticity and integrity of aggregation results receive less attention. Since aggregation operation could be performed by any intermediate forwarding vehicle, any malicious attacker could easily launch the attacks towards the data aggregation process by modifying the aggregated result or simply inserting invalid sensing data.

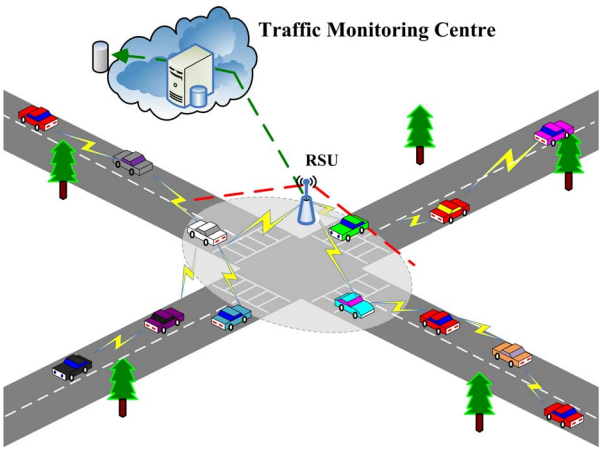

Fig. 1. Overview of Vehicular Sensing Network

978-1-4244-6404-3/10/$26.00 ©2010 IEEE

Secure data aggregation represents a great challenge in vehicular sensing networks due to their unique network characteristics including: highly dynamic network topology, intermittent connectivity as well as potentially a huge number of VANET nodes. These unique characteristics make the secure data aggregation in traditional wireless sensor networks such as [7], which always assume either a static network topology or aggregation structure, unsuitable for vehicular sensing networks.

Therefore, to achieve secure and efficient data sensing and collection, in this paper, we present the SAS, a secure data aggregation scheme for vehicular sensing networks, which is based on Flajolet-Martin sketch and a series of *sketch proof* techniques. We also discuss the tradeoff between the bandwidth efficiency and the estimation precision. Extensive simulations and analysis demonstrate the efficiency and effectiveness of the proposed scheme.

The remainder of the paper is organized as follows.In Section II, we present the system model, attack model, security assumptions and the design goals. Section III gives a review on some preliminary backgrounds. In Section IV, the proposed SAS is presented in detail. Performance analysis is given in Section V, followed by the conclusion in Section VI.

II. System Model and Design Goal

A. Network Model

In this paper, we consider a general vehicular sensing network model, which is mainly comprised of three components: traffic monitoring centre (TMC), RSUs and vehicles. As shown in Fig. 1, RSUs could be selectively deployed at some positions (e.g., intersections) to collect the traffic information (e.g., average speed) within a certain area. Due to high maintenance cost, RSUs could be only deployed intermittently to reduce the deployment cost. We assume that each vehicle, which is equipped with an on-board unit (OBU), has the capability of data collecting and reporting. The transmitted sensing data are propagated via V2V and V2I communications to the RSUs, which then forward them to the TMC. SAS is based on the distributed aggregation model similar to [4], which does not require any group/cluster formulation.

B. Security Assumptions

We assume that each OBU either shares a distinct secret symmetric key with TMC or obtains a public/private key pair, which is issued by TMC. Whether using shared secret key or public key depends on different system requirements.

C. Attack Model

In this study, we assume that the TMC and RSUs are trusted while vehicles (including the sensing vehicles and aggregator vehicles) are potentially malicious and can thus launch various attacks including fabricating, duplicating and computing the aggregation incorrectly, etc. We do not consider denial-of-service attacks where aggregator vehicles fail to or refuse to provide any acceptable result. A malicious sensor can always report an arbitrary sensing report, which fundamentally cannot be prevented. So we do not aim at preventing such a attack.

D. Design Goals

- **Security Goal:** The security goal of SAS is to enable the TMC to verify whether an aggregate sensing report is correct or not. Specifically, TMC should accept a reported aggregate report if and only if it equals to the output of a correct execution of the aggregation function over all of the sensing report provided by the qualified vehicles in the most recent epoch,
- **Efficiency and Effectiveness Goal:** The efficiency goal of SAS is to minimize the transmission overhead and, at the same time, to ensure a certain sensing accuracy. However, computational cost is not a major concern of this paper since VANET is generally assumed to have unlimited computational capability [8].

III. Preliminaries: One-way chains and MAX protocols

One-way chain is a widely used cryptographic primitive, which is based on some one-way function F and a secret seed s. The one-way function F is easy to compute but computational infeasible to invert. The chain has the sequence of values $F(s)$, $F(F(s))$, $F(F(F(s)))$, ... Throughout this paper, we use $F^x()$ to denote recursively applying the function F for x times. Thus the x'th value in the sequence is $F^x(s)$. For example, given two positive integers m and n, where $m < n$, it is easy to compute the $F^n(s)$ by functioning forward the value of $F^m(s)$ for $(n-m)$ times with the function F. On the other hand, it is infeasible to compute the value of $F^m(s)$ by functioning backward the value of $F^n(s)$. One-way chain has been widely used in many security topics such as micropayment. Recently, in [7], the authors take advantage of one-way chains to construct a MAX protocol, which could ensure the aggregated maximum message cannot be inflated or deflated. However, MAX protocol is not designed for probabilistic aggregation. Further, the network topology considered in [7] is for a sensor networks with statistic network topology. In SAS, what we consider is a dynamic network topology and probabilistic aggregation model.

IV. A Secure Data Aggregation Scheme for Vehicular Sensing Networks

In this section, we firstly introduce the concept of FM-Sketch, which is the foundation of probabilistic data aggregation in vehicular networks. We then propose a secure data aggregation scheme based on our proposed *sketch proof* technique.

A. FM sketches based Data Aggregation in VANETs

A Flajolet-Martin sketch (or "FM sketch") is a data structure for probabilistic counting of distinct elements that has been introduced in [9]. FM sketch represents an approximation of a positive integer by a bit field $s = s_1, \ldots, s_w$ of length w, where $w \geq 1$. The bit field is initialized to zero at all positions. To add an element x to the sketch, it is hashed by a hash function h with geometrically distributed positive integer output, where $P(h(x) = i) = 2^{-i}$. The entry $s_{h(x)}$ is then set

to one. After processing all objects, FM finds the first bit of the sketch that is still 0. Let the position of this bit be k; then the number of distinct objects is estimated as $n = 1.29 \times 2^k$.

The variance of n is quite significant [4], and thus the approximation is not very accurate. To overcome this, instead of using only one sketch, a set of sketches can be used to represent a single value to achieve trade off between the accuracy and memory. The respective technique is called Probabilistic Counting with Stochastic Averaging (PCSA) in [9]. With PCSA, each added element is first mapped to one of the sketches by using an equally distributed hash function, and is then added there. If m sketches are used, denoted by $S_1, ..., S_m$, let $a_1, a_2, , a_m$ be the positions of the first 0 in the m sketches respectively, the estimate for the total number of distinct items added is then given by $n = 1.29 \times 2^{k_a}$, where $k_a = (1/m)\sum_{i=1}^{m}(a_i)$.

Sketches can be merged to obtain the total number of distinct elements added to any of them by a simple bit-wise OR. Important here is that, by their construction, repeatedly combining the same sketches or adding already present elements again does not change the results, no matter how often or in which order these operations occur. This makes FM sketches ideally suited for VANET aggregation.

For the purpose of discussion, let us consider a specific application. Assume that we are interested in monitoring the average speed within a certain area. As the first step, we use a sketch for each road segment and approximate the sum of speeds of vehicles within this road segment. For the second step, we will calculate the average speed by dividing the speed sum by the number of vehicles involved. In the following sections, we will discuss how to generate the sketch proof and secure sketch aggregation.

B. Sketch Proof Generation

According to the FM sketch definition, given the ID i and speed v_i, a vehicle may add the tuples $(i,1), \ldots, (i, v_i)$ to the sketch by hashing them and setting the respective bit position $(h(i,1), \ldots, h(i, v_i)$ to 1. The malicious attackers may launch two kinds of attacks towards the FM sketch: inflation attack and deflation attack.

We start from three basic pieces of information that each sensor generates in our protocol. Let $\Lambda^i = \{\ell_1, \ldots, \ell_{v_i}\}$ denote v_i 1-bit positions generated by i. Given that ψ_i is the position of first 0-bit, Λ^i could be represented as the union of two subsets $\Lambda^i_{\psi_i} = \{1, \ldots, \psi_i - 1\}$ and $\overline{\Lambda^i_{\psi_i}} = \{\ell_{\psi_i}, \ldots, \ell_{v_i}\}$, where ℓ_{ψ_i} represents the first 1-bit larger than ψ_i. Thus, each vehicle i generates

1) $s_i^+ = \{i, \psi_i, Loc\#, epoch\#, MAC_{K_i}(\omega||Loc\#||epoch\#)|\omega \in \Lambda^i_{\psi_i}\}$, which is called vehicle i's *inflation-free proof*. Here, $Loc\#$ and $epoch\#$ refer to the road segment number and time slot number, respectively.
2) $s_i^- = MAC_{K_i}(Loc\#||epoch\#)$, which is called vehicle i's *deflation-free proof*. This is basically the authentication code generated by the vehicle on the common information $Loc\#, epoch\#$.
3) $s_i^\times = \{\overline{\Lambda^i_{\psi_i}}, MAC_{K_i}(\omega||Loc\#||epoch\#)|\omega \in \overline{\Lambda^i_{\psi_i}}\}$, which is called vehicle i's *supplement security proof*.

In the follows, we will introduce these three security proofs one by one.

1) Inflation-free Proof: Inflation-free proof is basically the authentication code generated by the vehicles on the 1-bit positions, which are smaller than the position of first 0. To prevent the inflation attacks, it is sufficient to require that each 1-bit, whose position is less than ψ_i, should be authenticated by a single signed value from one of the sensing vehicles that turn it on. We define two extra operations for inflation-free proofs:

- **Merging Operation** $\oplus$: Consider two sketches Λ^i and Λ^j (for simplicity of presentation, we assume $\psi_i > \psi_j$). Let ψ_m be the globally maximum value of first 0-bit after sketch merging , which corresponds to the new $\Lambda_{\psi_m} = \{1, \ldots, \psi_m - 1\}$ and $\overline{\Lambda_{\psi_m}} = \Lambda^i \cup \Lambda^j \backslash \Lambda_{\psi_m}$. We define

$$\oplus_{\omega=i,j} s^+_{\psi_w} = s^+_{\psi_i} \cup s_i^\times(\Lambda_{\psi_m}) \cup s_j^\times(\Lambda_{\psi_m}),$$

  where $s_i^\times(\Lambda_{\psi_m})$ is the operation that picks up all the supplement security proof whose positions are less than ψ_m. In other words, to generate inflation-free proof for the merged sketches, the aggregator could first pick up the inflation-free proof $s^+_{\psi_i}$ of the sketch with a higher 0-bit position ψ_i. For the remaining 1-bit positions $\psi_i, \ldots, \psi_m - 1$, the aggregator could pick up the inflation-free proofs either from $s_i^\times$ or $s_j^\times$. Note that, if a 1-bit is authenticated by multiple MACs generated by multiple vehicles, aggregators could choose inflation-free proof of vehicles with a lower ID.
- **Aggregation operation** $\otimes$: The MACs of s_i^+ could be further aggregated. For example, if MAC is generated by symmetric key based hash function (e.g., MD5 or SHA-1), then $\otimes$ can be simple XOR; if MAC is signatures, $\otimes$ could be achieved by using aggregate signature technique such as [10].

With merging operation and aggregation operation, size of inflation-free proof could be minimized to $|ID| * N_{1-\text{bit}} + |MAC|$, where $|ID|$ and $|MAC|$ refer to the size of vehicle ID and MAC, respectively, and denote $N_{1-\text{bit}}$ as the number of 1-bits.

2) Deflation-free Proof: Deflation attack is defined as that the malicious aggregators may try to turn 1-bits into 0-bits, removing the corresponding MACs from the security proofs. To prevent deflation attack, SAS adopts the hash-chain based MAX protocol, which is introduced in [7]. The basic idea is to construct one-way chains whose seeds are all the s_i^-. Specifically, given the one way function $F()$, vehicle node i reports to the aggregator $F^{\psi_0}(s_i^-)$. In a case of multiple sketch aggregation, let ψ_m be the maximum positions of seen by the aggregator, The aggregator can obtain $F^{\psi_m}(s_i^-)$ by performing hash operations on $F^{\psi_0}(s_i^-)$ by $\psi_m - \psi_0$ times. After obtaining all the $F^{\psi_m}(s_i^-)$, a new operation is introduced in [7] to reduce the transmission cost, which is shown as follows:

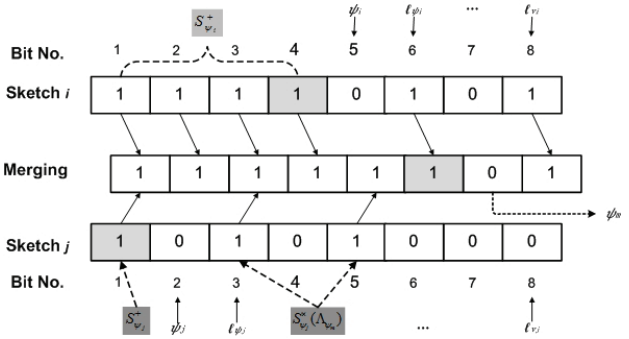

Fig. 2. Sketch Generation and Sketch Proof

- **Hash chain folding operation** $\odot$: The aggregator could use the folding function $\odot$ to fold all the hash chains into a single one $\odot F^{\psi_m}(s_i^-)$. Obviously, due to adoption of one-way function, it is impossible for the attackers to generate a new security proof for $\psi_i < \psi_m$, which prevents the deflation attack.

Note that, one-way chains should be rolled forward even after they have been folded together with an operation like $\odot$. Therefore, it required the one way function should achieve homomorphic property in that $F(x_1 \odot x_2) = F(x_1) \odot F(x_2)$. There are a wide range of cryptographic tool such as RSA encryption could support such kind of homomorphic property. In this case, $\odot$ could be defined as modular multiplication.

The size of deflation-free proof is a constant number $|F()|$, which represents the size of one way function output. If choosing RSA as the cryptographic tool, $|F()| = 1024$.

3) Supplement Security Proof : Supplement security proof enables the aggregator to derive the new inflation-free proof when ψ_0 changes because of the merge of sketches. Therefore, SAS records all 1-bits whose positions are larger than ψ_m, and their corresponding MACs as the supplement security proof. Since they are not continuous, supplement security proof cannot be aggregated. Further, we denote $s_i^{\times}(\overline{\Lambda_{\psi_m}})$ as the set of all the supplement security proofs whose positions are not less than ψ_m.

C. Sketch Proof Aggregation

Multiple sketches could be aggregated during their propagation process and sketch proofs could be aggregated along with sketches merging. Without loss of generality, we discuss aggregation algorithm SA only for two sketch proofs and more than two sketch aggregation can be aggregated by applying $SA()$ multiple times.

Consider two sketches Λ^i and Λ^j and their corresponding sketch proofs $s_i^+, s_i^-, s_i^{\times}$ and $s_j^+, s_j^-, s_j^{\times}$. Let ψ_m be the globally maximum value of first 0-bit after sketch merging. SA() aggregates sketch proof by performing the following steps:

- Inflation-free Proof Aggregation: $\otimes(\oplus_{\omega=i,j} s_{\psi_w}^+)$;
- Deflation-free Proof Aggregation: $\odot_{\omega=i,j} F^{\psi_m}(s_{\omega}^-)$;
- Supplement Security Proof Updating:

$$s_i^{\times}(\overline{\Lambda_{\psi_m}}) \cup s_j^{\times}(\overline{\Lambda_{\psi_m}});$$

TABLE I
THE SIZE OF EACH COMPONENT OF SAS (BYTES)

	No SAS	SAS
T&L	$8 \times n$	8
ID	$8 \times n$	$8 \times n$
Data v	$8 \times n$	0
$Sketch_i$	0	$8 \times log_2(v_{max} \times n)$
$SketchProofs$	$8 \times n$	$8 \times log_2(v_{max} \times n) + 136$
Total Size	$32 \times n$	$8 \times n + 16 \times log_2(v_{max} \times n) + 144$

Note that such a sketch proof aggregation process could be performed fully distritbuted, which means it naturally supports hierarchical aggregation while does not require any aggregation architecture.

D. Sketch Proof Verification and Average Calculation

After the aggregation results and the security proof arrive at the TMC, TMC should verify the correctness of the inflation-free proof and deflation-free proof. To check the validity of inflation-free proof, TMC should performing the following operations in different MAC modes:

- Symmetric key mode: In this mode, TMC should recalculate the MAC of each 1-bit and then aggregate them into a single one. After that, TMC should check if the obtained result equals the received one.
- Signature mode: In this mode, TMC could batch verify the aggregated signatures by performing batch verification technique [10].

To verify the correctness of deflation-free proof, computes all individual s_{ω}^- and folds them together to create the $\odot_{\omega=1,2,...} F^{\psi_m}(s_{\omega}^-)$. The answer is accepted if and only if the calculated result equals the received one. Finally, by obtaining the ψ_m, the average speed could be computed as follows:

$$Speed_{average} = 1.29 \times 2^{\psi_m}/N_{ID},$$

where N_{ID} refers to the number of vehicles involved. Similar to the original FM-sketch, the accuracy of this average speed estimation could be further improved by introducing multiple sketches.

V. Performance Evaluations

In this section, we evaluate the performance of the proposed SAS in terms of the resultant communication cost and approximate accuracy. To demonstrate the superiority of SAS, we also compare SAS with non-aggregation transmission approach. In this part, we consider SHA-1 as the building blocks of MAC. Note that, asymmetric-key based MAC mode will have a similar communication cost if we choose short aggregate signature as the building blocks.

A. Transmission Overhead

One of the major advantages of SAS is the reduction of its transmission cost. The communication cost is determined by the size of aggregated security proof including inflation-free proof, deflation-free proof and supplement security proof.

As a typical example, we choose the 64 bits SHA-1 as the basic MAC technique and $RSA-1028$ as the basic one-way function tool. Table I summarizes the size of different components as well as the overall transmission overhead for non-aggregation transmission and SAS transmission. Here, we consider the worst case of our aggregation in that the size of supplement security proofs are bounded by $log_2(v_{max} \times n)$ [4], where v_{max} is the maximum speed for this road segment while n is maximum number of vehicles in this area. However, it is important to point out that, in practice, the size for supplement security proof should be much less than this bound since it will decrease along with the aggregation.

By choosing different number of sketches, we obtain the different communication cost of SAS under different vehicle numbers as well as different sketch numbers, which has been shown in Fig. 3. It is observed that, the probabilistic aggregation does not show its advantage when the number of vehicles is small. However, when the number of vehicles grows, he proposed SAS aggregation scheme could dramatically reduce the communication cost when the sketch number is small. It is also observed that the number of sketches plays an important role for the overall system performance in that a small sketch number such as 4 makes the proposed SAS to have a better performance while, when the sketch number is large such as 16, the advantage is not so obvious. Therefore, if an acceptable accuracy is guaranteed, the number of sketch should be as small as possible to achieve a better performance. In the next section, we will discuss the tradeoff of accuracy and the number of sketches.

B. Tradeoff of the Accuracy and Number of Sketches

According to [4], PCSA yields a standard error of approximately $0.78/\sqrt{m}$. By choosing different sketch number, we can obtain the corresponding standard error, which has been plot in Fig. 4. It is observed that the standard error decreases dramatically along with the increase of number of sketches in the beginning while keeps relatively stable after a specific threshold (e.g., 4 in Fig. 4). However, as we pointed out in previous section, in the vehicular sensing networks, a small number of sketches (e.g., 4) guarantee an acceptable standard error (e.g., 0.39). This further demonstrates the effectiveness of the proposed SAS.

VI. Conclusion and Future Work

Vehicular sensing networks have been envisioned to play an important role for future traffic monitoring applications. In this study, we propose a secure and efficient aggregation method based on FM-sketch and security proofs techniques. The extensive performance evaluations have demonstrated the efficiency and effectiveness of the proposed scheme. Our future work includes implementing SAS in a specific application scenario and evaluate its performance with more realistic simulations or even experiments.

Acknowledgment

This research is supported by the National Natural Science Foundation of China, Grant No. 70971086.

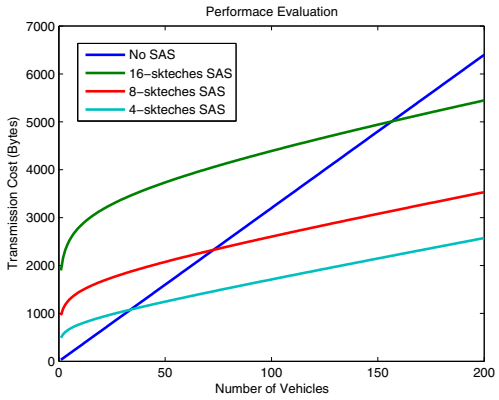

Fig. 3. Transmission Overhead of Various Secure FM Sketch

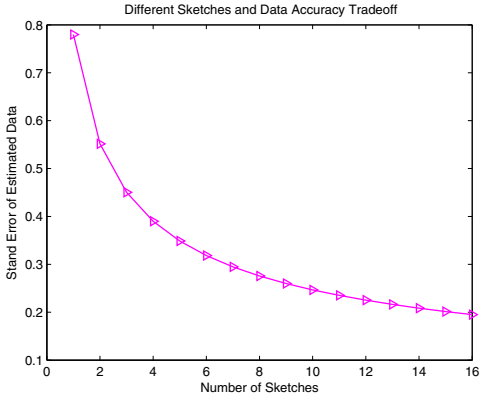

Fig. 4. Standard Error of SAS Secure Sketch

References

[1] M. Fontaine,"Traffic Monitoring," in *Vehicular Networks from Theory to Practice*, CRC Press, 2009.

[2] M. H. Arbabi, M. Weigle, "Using Vehicular Networks to Collect Common Traffic Data," in Proc. of *VANET'09*, 2009.

[3] S. Dietzel, B. Bako, E. Schoch, F. Kargl, "A Fuzzy Logic based Approach for Structure-free Aggregation in Vehicular Ad-Hoc Networks," in Proc. of *VANET'09*, 2009.

[4] C. Lochert, B. Scheuermann and M. Mauve, "Probabilistic aggregation for data dissemination in VANETs," in Proc. of *VANET'07*, 2007.

[5] Tamer Nadeem, Sasan Dashtinezhad, Chunyuan Liao, and Liviu Iftode, "TrafficView: Traffic Data Dissemination using Car-to-Car Communication," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 8, no. 3, pp. 6-19, 2004.

[6] Murat Caliskan, Daniel Graupner, and Martin Mauve, "Decentralized Discovery of Free Parking Places," in Proc. of *VANET'06*, New York, NY, USA, 2006, pp.30-39, ACM.

[7] S. Nath, H. Yu and H. Chan, "Secure Outsourced Aggregation via One-way Chains," in Proc. of *SIGMOD'09*, 2009.

[8] H. Zhu, X. Lin, R. Lu, P.H. Ho and X. Shen, "AEMA: An Aggregated Emergency Message Authentication Scheme for Enhancing the Security of Vehicular Ad Hoc Networks,"in Proc. of *IEEE ICC'08*, Beijing, China, May 19-23, 2008.

[9] P. Flajolet and G. N. Martin, " Probabilistic counting algorithms for data base applications," *Journal of Computer and System Sciences*, pp. 31, no.2, pp.182-209, Oct. 1985.

[10] D. Boneh, B. Lynn and H. Shacham, "Short Signatures from the Weil Pairing," in *Journal of Cryptology*, vol. 17, no. 4, pp. 297-319, 2004.