Protecting Source-Location Privacy against Hotspot-Locating Attack in Wireless Sensor Networks using Partial Cloud-Based Scheme

S.Ramakrishnan, R.Velmani, N.Sakthivel

Abstract— The wireless sensor networks adversaries can make use of the traffic information's to locate the monitored objectsin platform as a service(Paas) , e.g., to hunt endangered animals or kill soldiers. In this paper, we first define a hotspot phenomenon that causes an obvious inconsistency in the network traffic pattern due to the large volume of packets originating from a small area. Second, we develop a realistic adversary model, assuming that the adversary can monitor the network traffic in multiple areas, rather than the entire network or only one area. Using this model, we introduce a novel attack called Hotspot-Locating where the adversary uses traffic analysistechniques to locate hotspots. Finally, we propose a cloud-based scheme for efficiently protecting source nodes' location privacy against Hotspot-Locating attack by creating a cloud with an irregular shape of fake traffic, to counteract the inconsistency in the traffic pattern and camouflage the source node in the nodes forming the cloud. To reduce the energy cost, clouds are active only during data transmission and the intersection of clouds creates a larger merged cloud, to reduce the number of fake packets and also boost privacy preservation. Simulation and analytical results demonstrate that our scheme can provide stronger privacy protection than routing-based schemes and requires much less energy than global-adversary-based schemes.

Index Terms— Wireless sensor network privacy, source-location privacy-preserving schemes, context privacy, and anonymity ,merged cloud.

1 INTRODUCTION

A wireless sensor network (WSN) consists of a large number of sensing devices, called sensor nodes, which are interconnected through wireless links to perform distributed sensing tasks. WSN have found many useful applications for automatic data collecting , such as habitat monitoring, military surveillance, and target tracking, for monitoring the activities of enemy soldiers or valuable assets, e.g., endangered animals. When a sensor node detects a soldier or an endangered animal, it reports the event to the data collector called the Sink. This data transmission may occur via multi hop transmission, where the sensor nodes act as routers. In this paper, we consider habitat monitoring applications where the WSN is deployed for monitoring pandas. For example, a WSN has been deployed by the Save-The-Panda Organization to monitor pandas in a wild habitat . While pandas move in the network, their presence and activities are periodically sensed by the sensor nodes and reported to the Sink.However, WSNs are usually deployed in open and large areas that are unattended and lack of protected physical boundary, which makes the networks vulnerable to many threats. Since the sensed data are typically transmitted through wireless channels, adversaries can eavesdrop on the open and shared wireless medium and make use of traffic information to locate source nodes to hunt pandas.

Therefore, preserving source nodes' location privacy is essential due to the easiness of locating pandas and their furs'. large market value, e.g., a piece of a panda's fur was sold in China for $66,500 in 2003. The privacy threats can usually be classified into: content privacy and contextual privacy . For the content privacy threat, the adversary attempts to observe the content of the packets sent in the network to learn the sensed data and the identities and locations of the source nodes. This privacy threat can be countered by encrypting the packets' contents and using pseudonyms instead of the real identities. For the contextual privacy threat, the adversary eavesdrops on the network transmissions and uses traffic analysis techniques to deduce sensitive information, including whether, when, and where the data are collected. Actually, the act of packet transmission itself reveals information even if the packets are strongly encrypted and the adversary could not interpret them.
The existing source location privacy-preserving schemes can be classified into global-adversary-based and routing-based schemes.

2 RELATED WORKS

Recently, location privacy in wireless and wired networks has gained much attention. Different schemes have been developed to protect users' privacy in location tracking systems. Which determine the users' positions for location-based services. Location privacy in these schemes is content oriented, where location information is collected and protected as the users' private data. An Onion routing provides the anonymous communications for the Internet by hiding the identities of the end users of a communication session. The proposed schemes in conceal the nodes' network/MAC addresses in order to achieve anonymous communications for mobile ad hoc networks. However, these schemes employ different network and threat models from the ones suitable for the source location
privacy problem in sensor networks

Routing-based schemes preserve source nodes' location privacy by sending packets through different routes to make back tracing the movement of the packets from the Sink to the source nodes infeasible. In a random-walk-based privacy-preserving scheme, called Phantom, is proposed.Each packet takes a random walk to a random location before . it is sent to the Sink. However, the scheme fails if the adversary's overhearing range is more than the sensor nodes' transmission range.

Global-adversary-based schemes assume that adversaries can monitor the traffic of the entire network.Each node has to periodically send packets, and send dummy packets if it does not have sensed data so that it is infeasible for the adversaries to distinguish between the real and dummy packets.

3 NETWORK AND ADVERSARY MODELS

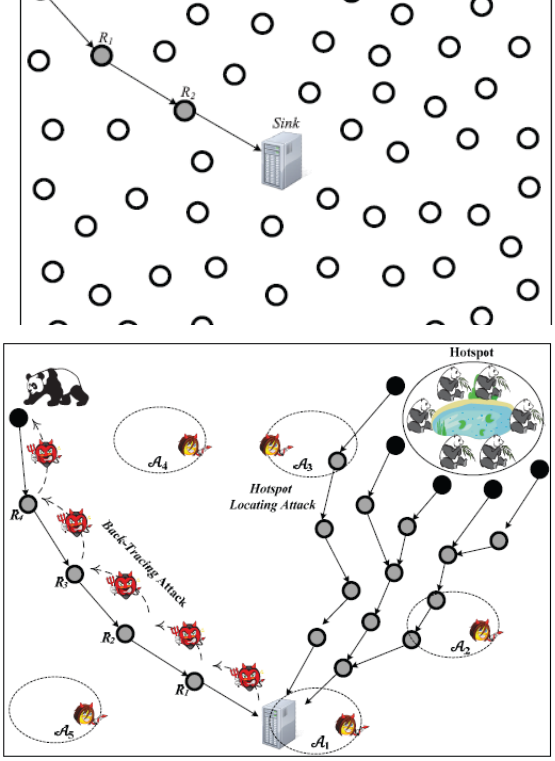
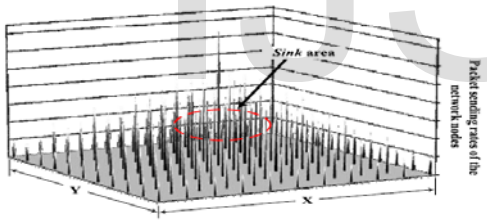


Fig.2. The Adversary Model

4 HOTSPOT-LOCATING ATTACK

4.1 Hotspot Phenomenon

A hotspot is formed when a large volume of packets are sent from the sensor nodes of a small area, causing an obvious inconsistency in the network traffic which may last for some time. The adversary attempts to make use of this traffic inconsistency to locate hotspots t… he-nomenon … en-sor node … ut-ing scher… the Sink thro… ffic pattern is … sor node and …



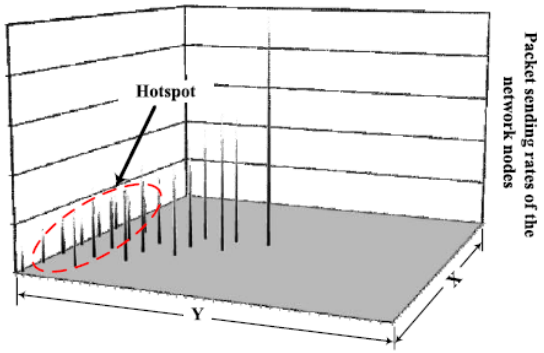


Fig. 4. The packet sending rate of each node with a hotspot.

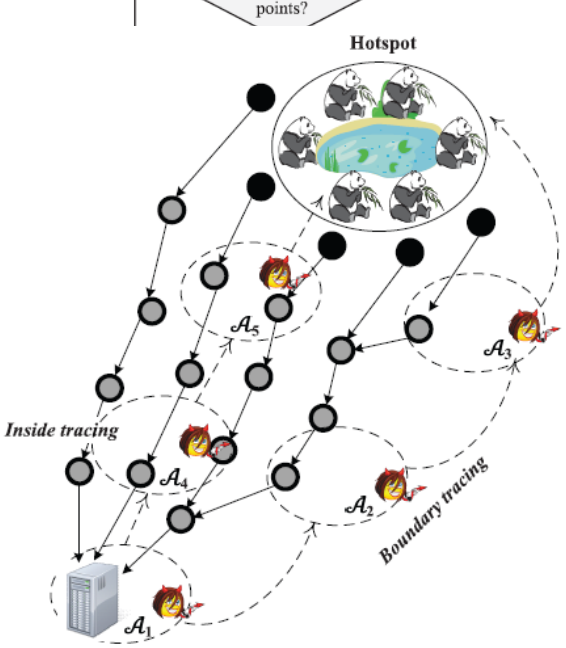
5 CLOUD-BASED PRIVACY-PRESERVING SCHEME

5.1 Predeployment Phase

Before deploying the network, each sensor node A is loaded with a unique identity ID_A, a shared key with the Sink K_A, and a secret key d_A that is used to compute a shared key with any sensor node using

Issue 6, J …

…of the …ensor … Sink …urce- … data … reless … dver-…erest, … these … attack …initial … Sink …buted …t the …ckets …lually …



…work, …. This …t the …; fake …nd 3) …g the … node …w the …nodes …es its …ch as …te.

Fig. 6. Inside and boundary back tracing for locating hotspots.

.

In order to assign fake source nodes, node A broadcasts Fake Nodes Request Packet (FREQ) that contains the maximum number of hops (h_{max}) the packet can be propagated. Each node adds its identity and broadcasts the packet if the number of hops is fewer than h_{max}; otherwise, it unicasts Fake Nodes Request Reply (FREP) packet to node A, containing the identities of the nodes in the route. Node A receives multiple FREP packets containing different routes with maximum number of hops of h_{max}. It chooses a group of nodes at different number of hops and unicasts the Fake Node Assignment Packets (FASS) to assign them as fake source nodes to its packets. For each FASS packet, node A adds the identities of the nodes in the route and a random value that will be used to generate pseudonyms shared between each two neighboring nodes in the route.

5.3 Event Transmission Phase

Privacy is the guarantee that information in its general sense is observable or decipherable by only those who are intentionally meant to observe or decipher it. According to Pfitzmann and Kohntopp anonymity is defined as the state of being unidentifiable within a set of objects called the anonymity set. The essence of our scheme is based on the principle that one of the best ways to avoid being identified is to mix with the crowd. Our scheme conceals a source node within a group of nodes with an irregular shape, called “cloud.” A source node is considered to have a complete anonymity if the adversary cannot identify it in the cloud, i.e., the adversary may be able to know that a node in a cloud sends an event packet, but he cannot identify this node.

5.3.1 Pseudonyms

If two nodes share a key, they can create a sequence of pseudonyms using a one-way keyed hash function byiteratively hashing a random value.nodes can generate different pseudonyms using the same key. This means that pseudonyms are not only used for identifying the sending and receiving nodes, but also for identifying routes by using different random values for different routes.However, if a node does not receive a pseudonym, e.g.,
due to packet drop, the two nodes may lose pseudonym synchroniza-

tion. To avoid this, each node should use a sliding window to match a received pseudonym against a window of expected pseudonyms. The expected pseudonym is number i, but the node matches a received pseudonym with a window of n expected pseudonyms.

5.3.2 Real—Fake Source Nodes' Route

When a source node (S) wants to send data to the Sink, it first picks up a fake source node (F) from its list of fake source nodes and a group (G_1) that contains F, and sends the following event packet:

The source node encrypts the messageM with the shared key with the Sink (K_S) to provide message confidentiality, authenticity, and integrity. In order to enable the Sink to know the location of the source node and the key it should use to decrypt the message.

5.3.3 Fake Packets

As an event packet is propagating from the real source node to the fake source, fake packets are sent to create a cloud of fake traffic. To send a fake packet, the node T chooses a group, e.g., G_4, and sends the following fake packet.

5.3.4 Fake Source Node—Sink Route

From the pseudonym , the fake source node can know that the packet has to be sent to the Sink. As shown in Fig. 8,the fake source node sends the packet to the first node in the route to the Sink. The packet contains a pseudonym shared with the next-hop node , and the message and the source node's pseudonym encrypted with the shared key with the Sink. Each relaying node re encrypts the packet with the shared key with the Sink and replaces the pseudonym with the one shared with the next node in the route. The purpose of adding an encryption layer at eachrelaying node is to make the packet look different as it propagates from the fake source node to the Sink to prevent packet correlation and make back tracing packets to the fake source node infeasible. As the packet propagates to the Sink,the neighboring nodes do not send fake packets because they cannot find the packet's pseudonym in their tables.When the Sink receives the

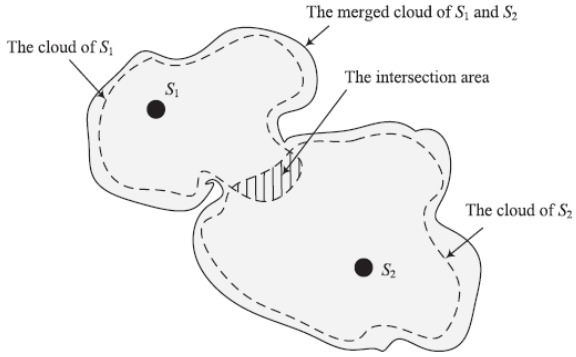


Fig .7. Merging clouds

This cloud merging is possible when S_1 and S_2 transmit events within a short time interval. Cloud merging has two main benefits: 1) lower energy cost: the nodes in the intersection areas do not send one fake packet for each cloud, e.g., if a node participates in n clouds, it sends only one fake packet instead of n, and thus the node can save n-1 fake packets; and 2) stronger privacy protection: a merged cloud has a larger anonymity set because it has more nodes than the individual clouds. Cloud merging property is especially important for hotspots because clouds are very likely intersected which can significantly reduce the number of fake packets and boost privacy protection.

6 EVALUATIONS

6.1 Privacy Preservation

6.1.1 Analysis

For Pseudonyms unlinkability, the adversary cannot link the pseudonyms of one sequence. The importance of this property lies in the fact that if an adversary could link a pseudonym to a node, he will not benefit from this conclusion in the future. In our scheme, generating or correlating pseudonyms is infeasible without knowing the secret key used in generating them. Even if there is only one transmission, fake packets can make pseudonyms linkability infeasible because the adversary cannot distinguish between event and fake packets. Pseudonym collision means that more than one node have the same pseudonym, because the hash function may generate the same hash value from hashing two different inputs.

For packet length correlation, the packets of one flow can be correlated if they are distinguishable from their lengths. To prevent this, all packets should have the same length, or random length by adding random-length padding bits at each relaying node. For packet sending time correlation, an adversary tries to deduce the forwarding path by observing the transmission time of a node and its neighbors. The adversary makes use of the fact that the nodes usually relay packets after short delay and based on first-received-first ransmitted basis. Changing packets' appearance at each hop cannot prevent this correlation because it depends on the packets' sending times and not the content. To obfuscate the temporal relationship between the transmissions of consecutive hops, each node can delay relaying the packets for a random amount of time and buffer/reorder packets. Moreover, fake packets can make time correlation nearly impossible because the random nature of the channel access in MACprotocols introduces randomness to the transmissions'times. Sending fake and real packets also confuses the adversary and makes time correlation infeasible even if there is only one event transmission.

For fake and real source nodes unlink ability, if an adversary could locate a fake source node, he should not gain any information about the location of the corresponding real source node. This link ability is infeasible because each real source node sends its packets through multiple fake sources, each fake source node serves different real sources, and the distance between a fake source node and the real source is random. If the distance between a fake source node and the real one is fixed or has a minimum number of hops (d_{min}), the adversary can figure out the relative location of the real source node or conclude that it cannot be in the fake source node's d_{min}-hop neighbors. What alsomakes this linkability infeasible is that the adversary observes all the transmissions of a cloud random because he cannot distinguish between fake and real packets.For cloud shape and source node unlinkability, if a strong adversary could trace a part of a cloud or all the cloud, he cannot infer any information about the source node's location. For example, if a cloud is circle shaped and the source node is located at the center, the adversary can gain some information about the source node's location by tracing a part of the cloud. In our scheme, this

linkability is infeasible because clouds are irregular and changeable, and some nodes may belong to multiple clouds at the same time, which creates an overlapped and complex merged cloud.

For merged-cloud splitting attack, the adversary tries to reduce the size of a merged cloud, e.g., to reduce the anonymity set. In our scheme, the traffic of individual clouds is indistinguishable because a cloud's packets do not have any data that refer to the cloud, and thus the adversary cannot split a merged cloud or even identify the boundaries of the individual clouds. Cloud merging can increase the anonymity set without extra overhead, e.g., if two clouds each with n_c nodes are merged, the anonymity sets of the individual clouds are n_c but the anonymity set of the merged cloud is $2n_c\_n_o$, where n_o is the number of nodes belonging to the two clouds.

For packet back tracing attack, it is unlikely that the adversary will continuously receive event packets from a source node because packets are sent through different fake source nodes which can be far from each other. What also complicates this attack is that event packets sent from a real or fake source node at different times are uncorrelated. Moreover, even if the adversary could capture the same packet at different relaying nodes, he cannot correlate the packets. Even if the adversary could trace back packets to a fake source node, he cannot locate the corresponding real source node due to the fake and real source nodes unlinkability.

For packet-replay attack, the adversary tries to replay old packets repeatedly in order to observe the traffic patterns of packet forwarding, e.g., to figure out the network topology to locate source nodes. This is infeasible because the adversary cannot compute fresh pseudonyms and the nodes drop packets if they cannot recognize their pseudonyms. For packet sending rate analysis, the adversary attempts to make
use of the fact that the nodes near of hotspots send more packets than the nodes far away to locate hotspots. Even with changing the packets' appearance at each hop, the adversary
can still analyze the packet sending rate. Our scheme uses fake packets to camouflage the nodes that are close to pandas with the other nodes in the cloud in such a way that makes this spot indistinguishable.

For event packets flow recognition attack, the adversary attempts to recognize the flow of real packets to identify the source node or at least a small area around it. For example, from Fig. 8, if the adversary could recognize the flow of the real packets from node B to F, he can deduce that the panda cannot be in the region between B and F and reduce the
anonymity set. In our scheme, the event and the fake packets are indistinguishable and the adversary cannot correlate an event packet as it is relayed from the real source node to the fake one.

Event unobservability means that the adversary cannot know whether pandas are sensed or not. This property is
more important in other applications such as military applications because the adversary can know whether the network operator (enemy) could observe his soldiers. However, this property is not important in habitat monitoring application especially when the network is large and exhaustive search for pandas is infeasible. Moreover, achieving this property requires extreme energy cost due to sending dummy packets periodically. In our scheme, the adversary may know that pandas are detected, but he cannot know the exact locations of the pandas or at least a small area where he can search for them.

Routing-based privacy-preserving schemes use privacy metric called safety period which is the number of packets the adversary has to capture in order to move from the Sink to a source node. Stronger privacy protection can be achieved with increasing the safety period. This metric is not accurate because it measures the best case when the adversary starts from the Sink, but if the adversary captures a packet at any relaying node, the safety period decreases.

6.1.2 Simulation Results

We have built up a discrete and event-based simulator to evaluate the effectiveness of the Hotspot-Locating attack and the privacy protection of our scheme and routing-based schemes. Four thousand nodes are uniformly randomly

TABLE 1
Simulation Parameters

Parameter	Value
Number of nodes	1000
Network size	3500m*3500m
Number of hotspot	1
Number of sensor nodes in hotspot	15
A sensor node`s transmission range	50m
Adversary`s hearing range	E * 50 m
Sink Location	Center
Sensor nodes and the hotspot	Uniformaly distributed
The number of monitoing device	N
Event transmission rate	1/30 seconds

TABLE 2
False Positive Probability

Scheme	N	4			8		
	E	1	2	4	1	2	4
Shortest Path		0.2	0.1	0.1	0.1	0.1	0
My Scheme		1	0.9	0.9	0.9	0.9	0.8

TABLE 3
Hotspot Detection Probability

Scheme	N	4			8		
	E	1	2	4	1	2	4
Shortest Path		0.7	0.7	0.9	0.8	0.9	1
My Scheme		0	0.4	0.1	0.5	0.1	0.2

The simulation results given in Tables 2 and 3 demonstrate that the false positive probability decreases and the detection probability increases when the monitoring devices overhearing radius increases. This is because the adversary can monitor more nodes and collect more accurate traffic information. This is also true when the number of monitoring devices increases. It can also be seen that the weak adversary who has few monitoring devices with small overhearing radius will very likely locate the hotspots in the shortest path and

Phantom schemes. This is because the shortest-path scheme does not preserve location privacy and the Phantom scheme cannot prevent packet correlation and conceal traffic analysis information. The slight improvement in the location privacy protection with increasing h_w is because of adding little randomness to the network traffic.

In our scheme, the powerful adversary who has a large number of monitoring devices with large overhearing radius will not locate hotspots. We found that in the runs that the adversary could be close to the cloud, he could not conclude information about the location or the direction of the hotspot in the cloud. The few times the adversary could locate the hotspot were random. Therefore, what an adversary can do is to exhaustively search the cloud.

6.2 Energy Cost

As we have discussed earlier, using cryptosystems is necessary to prevent packet correlation, and using fake packets can boost source nodes' location privacy preservation. To reduce the energy cost, our scheme uses energy efficient cryptosystems, including hash function and symmetric key cryptography, and avoids the extensively energy consuming asymmetric-key cryptography. From gives the consumed energy for sending/ receiving 1 bit and computing the cryptographic operations required for our scheme. We can see that the hashing and symmetric-key encryption/decryption operations consume low energy comparing to pairing operations. However, pairing operations are used only one time in the network lifetime because keys can be permanently stored after they are computed due to the static nature of the network topology. Since the Sink has more computational and energy capabilities than the sensor nodes, the nodes in the route
between a fake source node and the Sink encrypt the packets but the Sink removes the encryption layers instead of using encryption and decryption operations at each node. The overhead can be further reduced by encrypting the packets at some nodes instead of all the nodes in the route. Comparing to global-adversary-based schemes, our scheme uses fake packets much more efficiently by sending them only if there is an event instead of periodically. Moreover, fake packets are sent only in the active cloud instead of flooding the entire network, and cloud merging can reduce the number of fake packets. Although our scheme requires more cryptographic operations than global-adversary-based schemes, these operations consume much less energy than transmitting/receiving packets, as indicated in Table 3. the required energy for transmitting 1 KB of data over 100 m consumes as much energy as executing three million microprocessor instructions.

7 Conclusion and Future Work

In this paper, we have introduced a novel attackto locate source nodes in WSNs, called Hotspot-Locating, which uses a realistic adversary model. We have also proposed a source location privacy-preserving scheme that creates a cloud of fake packets around the source node, varies traffic routes, and changes the packets' appearance at each hop.We have shown that even if the adversary does not have a global view to the network traffic, he can locate hotspots using few monitoring devices and simple traffic analysis techniques. Our simulation and analytical results have demonstrated that routing-based schemes cannot preserve the location privacy of hotspots because they cannot conceal the traffic-analysis information. Moreover, our scheme can provide a strong protection against Hotspot-Locating attack with much less energy cost comparing to global-adversary-based schemes. In our future work, we will try sophisticated approaches to locate hotspots with low false-positive probability. We will use computer-based image recognition algorithms in addition to the proposed traffic-analysis techniques. In other words, we will use these algorithms to locate hotspots in the traffic-pattern image created by the traffic analysis techniques.

References

[1] K. Sohraby, D. Minoli, and T. Znati, Wireless Sensor Networks: Technology, Protocols and Applications. John Wiley & Sons, Inc., 2007.

[2] I. Akyildiz, W. Su, Y. Sankarasubramanian, and E. Cayirci, "Wireless Sensor Networks: A Survey," Computer Networks, vol. 38, pp. 393-422, 2002.

[3] A. Arora et al., "A Line in the Sand: A Wireless Sensor Network for Target Detection, Classification, and Tracking," Computer Networks, vol. 46, pp. 605-634, 2004.

[4] "WWWF-the Conservation Organization," http://www.panda.org/, 2012.

[5] Star News, Panda Poaching Gang Arrested, Shanghai Star Telegram, Apr. 2003.

[6] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing Source Location Privacy in Sensor Network Routing," Proc. IEEE Int'l Conf. Distributed Computing Systems (ICDCS '05), pp. 599-608, June 2005.

[7] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards Event Source Unobservability with Minimum Network Traffic in Sensor Networks," Proc. First ACM Conf. Wireless Network Security (WiSec '08), pp. 77-88, Apr. 2008.

[8] K. Pongaliur and L. Xiao, "Maintaining Source Privacy Under Eavesdropping and Node Compromise Attacks," Proc. IEEE INFOCOM, Apr. 2011.

[9] Y. Fan, Y. Jiang, H. Zhu, and X. Shen, "An Efficient Privacy-Preserving Scheme against Traffic Analysis Attacks in Network Coding," Proc. IEEE INFOCOM '09, Apr. 2009.