IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 21, 2019, accepted February 12, 2019, date of publication February 22, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900390

Non-Interactive Group Key Pre-Distribution Scheme (GKPS) for End-to-End Routing in Wireless Sensor Networks

ASHWAG ALBAKRI[1,2], (Member, IEEE), AND LEIN HARN[1]
[1]Department of Computer Science Electrical Engineering, University of Missouri-Kansas City, Kansas City, MO 64110, USA
[2]Department of Computer Science, Jazan University, Jizan 45142, Saudi Arabia

Corresponding author: Ashwag Albakri (aoaz89@mail.umkc.edu)

ABSTRACT A novel design of secure end-to-end routing protocol in wireless sensor networks was recently proposed by Harn *et al.* Their design is based on a group key pre-distribution scheme (GKPS) using a multivariate polynomial. A group key also called a *path key*, is used to protect data transmitted in the entire routing path. Specifically, instead of using a link-to-link secure communication that uses multiple pairwise shared keys, it is an end-to-end secure communication that uses a single path key to protect data over the entire path. The problem with all polynomial-based key distribution schemes is that the security of these schemes, which are called *deterministic k-secure*, depends on the degree of the chosen polynomial. In other words, if the degree of the chosen polynomial is k, then capturing the $k + 1$ sensors (or more) can compromise the system's security. Although increasing the degree of the polynomial can improve the security, it increases the storage and computational requirements of the sensors. In this paper, we propose the first probabilistic polynomial-based GKPS, which is based on a multivariate polynomial. The security of our scheme is *probabilistic k-secure*, which means it is probabilistic to compromise the security of our GKPS after capturing the $k+1$ sensors. We show that the probability of a sensor capture attack can be significantly reduced.

INDEX TERMS Wireless sensor networks, network security, group keys, multivariate polynomial, cryptography.

I. INTRODUCTION

Wireless sensor networks (WSNs) have been deployed in numerous settings, such as in health/traffic monitoring [1], the military domain [2], and in hazardous environments for data acquisition purpose [3]. Due to the sensitivity of the data that are transferred in WSNs, these data need to be protected; otherwise, adversaries can easily capture the data and recover the sensitive information that is being exchanged among sensors. In order to fulfill security services such as data encryption and data authentication, the source and destination nodes must share a secret key prior to transmitting any data over the WSNs. Establishing secret keys among sensors is called the key distribution/establishment in WSNs. We assume that each sensor is randomly deployed into a geographical area so that their relative locations cannot be pre-determined. Furthermore, we acknowledge that sensors are limited-resource devices, which means they have limited memory space, computational power, and battery life. As a result, the design of a key distribution scheme in WSNs must take these limitations into account.

Most existing key distribution schemes in WSNs enable two sensors to establish a pairwise shared key. Tokens are generated and pre-loaded to the sensors by a key generation center (KGC) before deploying them into an area. Key discovery and establishment protocol are enabled to establish a pairwise key and utilizing it to encrypt and authenticate data transmitted between two sensors. In a communication path, which involves multiple links, the key establishment is executed repeatedly in every link to route encrypted data successfully. Recently, a novel design of a secure end-to-end routing protocol [4] has been proposed based on a group key pre-distribution scheme (GKPS). The group key, also called a path key, is used to protect data transmitted in the entire path. Thus, instead of using multiple pairwise shared keys in a link-to-link secure communication,

The associate editor coordinating the review of this manuscript and approving it for publication was Mingjun Dai.

it uses an end-to-end path key, protecting data over the entire path. It can be concluded that implementing the end-to-end protocol is far more efficient and secure than the link-to-link protocol [4]. There are other types of schemes to establish group keys. For example, Lee *et al.* [5] proposed a scheme recently, and its security is based on the bilinear computational Diffie-Hellman assumption. Lee *et al.* [6] proposed another three-party-authenticated key agreement scheme based on chaotic maps without a password table.

II. RELATED WORK

Most key establishment schemes in WSNs create a pairwise key between two sensors. We can classify these schemes into two types: the probabilistic and the deterministic key establishment schemes. Eschenauer and Gligor [7] proposed the first key pre-distribution scheme, called the *random key* distribution scheme. In a random key distribution scheme, a large pool of random keys is initially generated by the KGC. Then, a subset of random keys, called a *key ring*, is randomly selected and pre-loaded into each sensor. A pairwise shared key can be established between two sensors only if some overlapping keys in two key rings of sensors exist. Therefore, this scheme is considered a probabilistic key establishment scheme. The probability of key establishment can be adjusted by varying the sizes of random key pools and key rings. One weakness of this random key distribution scheme is that the secrecy of the random key pool will be compromised by an adversary if a sufficient number of key rings have been captured. Chan *et al.* [8] proposed a *Q-composite scheme* to improve the resilience of the random key scheme. In their scheme, only in the case of two sensors sharing at least Q keys, can they establish a link-to-link communication. Even though this scheme improves the resilience against sensor capture attacks, it degrades the network connectivity since it requires at least Q shared keys to establish secure communications. There are some other random key distribution schemes to improve the resilience against sensor capture attacks. For instance, Chan *et al.* [8] proposed a pairwise key pre-distribution scheme in which each captured sensor did not reveal any information about external links. Nonetheless, their scheme is not scalable. Du *et al.* [9] proposed a random scheme assuming that the location of the sensors was available before deployment. This assumption is considered impractical for most applications. Rasheed and Mahapatra [10] proposed two key pre-distribution schemes in which bivariate polynomials were used in generating the random key pool. However, their scheme requires the use of mobile sinks in order to insure secure communications. In 2013, Ruj *et al.* [11] proposed a triple key establishment scheme in which any three sensors could establish triple keys among them. Recently, Yağan and Makowski [12] investigated the resiliency of WSNs against sensor capture attacks where they based their scheme on the random pairwise key distribution scheme of Chan *et al.* [8]. Ding *et al.* [13] considered prior knowledge of network characteristics and application constraints in terms of communication needs between sensor nodes and proposed methods to design key pre-distribution schemes in order to provide better security and connectivity. In 2017, Gandino *et al.* [14] proposed a q-s-composite protocol in order to exploit the best features of random pre-distribution and to improve it with lower requirements. All in all, providing high connectivity and strong resiliency against sensor capture attacks are two principal design objectives in random pre-distribution schemes. In order to provide high connectivity, the size of the key ring of each sensor needs to be large so the probability of locating overlapped keys between two sensors is high. However, favoring these features weakens the resiliency against sensor capture attacks since the adversary can recover more keys from each captured sensor.

Blom [15] proposed the first deterministic pairwise key establishment scheme using a symmetric bivariate polynomial. Blundo *et al.* [16] further investigated the key establishment using a symmetric bivariate polynomial, $f(x, y) = a_{0,0} + a_{1,0}x + a_{0,1}y + a_{2,0}x^2 + a_{1,1}xy + a_{0,2}y^2 + \cdots + a_{t-1,0}x^{t-1} + a_{t-2,1}x^{t-2}y + \cdots + a_{t-1,t-1}x^{t-1}y^{t-1} \bmod p$, where $a_{i,j} \in GF(p)$, and $a_{i,j} = a_{j,i}, \forall i, j$. If the KGC selects a symmetric bivariate polynomial to generate shares, $f(ID_i, y)$, $i = 1, 2, \ldots, n$, where ID_i is the public information of each sensor, S_i, then each share, $f(ID_i, y)$, is a univariate polynomial. Since $f(x_i, x_j) = f(x_j, x_i), \forall i, j \in [0, t-1]$, a pairwise key can be shared between two sensors, S_i, and S_j. Blundo *et al.* [16] also proposed a non-interactive k-secure m-conference protocol based on a multivariate polynomial, $f(x_1, x_2, \ldots, x_m)$. Because each share, $f(ID_i, x_2, \ldots, x_m)$, is a polynomial involving $m-1$ variables with degree k, each sensor needs to store$(k+1)^{m-1}$ coefficients. The storage space of each sensor is exponentially proportional to the size of the conference that deems this protocol impractical. Khan *et al.* [17] proposed a pre-distribution scheme using a symmetric matrix and a generator matrix of maximum rank distance to establish pairwise keys for sensor nodes. Sheu and Cheng [18] proposed a hop by hop authentication scheme for path key establishment in WSN that enabled sensor nodes to identify malicious nodes and detected false data that were injected in the network. Recently, Harn and Gong [19] and Harn and Hsu [20] proposed group key establishment schemes using a special type of multivariate polynomials. The advantage in using this special type of polynomial for group key establishment is that the storage requirement of each sensor is fixed and is independent of the size of the WSNs. However, there is one problem associated with all polynomial-based key distribution schemes: the security of these schemes, called *deterministic k-secure*, depends on the degree of the chosen polynomial. More specifically, if the degree of the chosen polynomial is k, then capturing $k+1$ or more than $k+1$ sensors can compromise the network's security. In other words, after capturing $k+1$ or more than $k+1$ sensors, the attacker is able to obtain the secret polynomial used to generate tokens of all sensors. Although increasing the degree of polynomials can improve the security, it also

increases the storage and computational requirements of the sensors.

In this paper, we propose a novel design of GKPS, which is based on a multivariate polynomial, but the security of our scheme is *probabilistic k-secure*. It is probabilistic to compromise the security of our proposed GKPS after capturing $k + 1$ or more sensors. We show that the probability of sensor capture attacks can be significantly reduced. Furthermore, our GKPS is very flexible in establishing path keys in WSNs. We need to point out that if the token of each sensor is stored without any tamper-resistant technology, it is quite easy for the attacker to recover the token of the sensor. In other words, it is quite impossible to prevent the attacker from recovering the token of that captured sensor without employing any tamper-resistant technology. In this paper, our proposed scheme does not prevent such an attack since we do not employ any tamper-resistant hardware. On the other hand, since our scheme is *probabilistic k-secure*, then after capturing $k + 1$ or more than $k + 1$ sensors, the attacker has an extremely low probability of obtaining the secret polynomial used to generate tokens of all sensors. Our scheme can prevent the attacker from obtaining all tokens of sensors after capturing $k+1$ or more than $k+1$ sensors. Thus, our proposed scheme enhances the system's security. To summarize the contributions introduced in this paper:

- We propose two group key pre-distribution schemes: a deterministic scheme and a probabilistic scheme;
- Both schemes are based on a multivariate polynomial but with limited storage requirements;
- The security of the second GKPS is probabilistic.

The rest of this paper is organized as follows: in the next section, we introduce the model of our proposed GKPS, including a description of GKPS schemes and their performance. In section IV, we demonstrate the three schemes: the basic, the modified, and the proposed GKPS in detail. Performance is discussed in section V, and a comparison to other key pre-distribution schemes is given in section VI. Finally, we conclude in section VII.

III. MODEL

A. DESCRIPTION OF PROPOSED GKPS

In our proposed GKPS, sensors are divided into multiple classes. Each sensor has a unique token initially generated and pre-loaded by KGC. The storage space of each sensor is linearly proportional to the number of classes and is independent of the number of sensors. In addition, this scheme allows multiple sensors to establish a group key (i.e., also called "**path key**" in [4]) non-interactively. Fig. 1 shows different paths protected by different group keys to securely routing the data through the entire path from source to destination sensor nodes. By changing the number of sensors in a WSN, the probabilities of establishing path keys with different lengths change as well. Similarly, changing the number of classes in a WSN can also change the probability of connectivity. One unique feature of our proposed scheme is the fact that it is the first polynomial-based GKPS with probabilistic security.

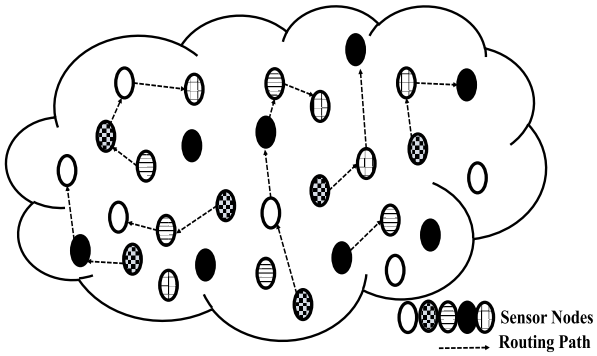

FIGURE 1. Group keys are utilized to securely routing data between sensor nodes.

B. PERFORMANCE

The following definitions will be used to evaluate the performance of the proposed GKPS.

Definition 1 (Probability of Capturing t Sensors Belonging to the Same Class P_t): After capturing t sensors, this is the probability that all captured sensors belong to the same class.

In Theorem 3, we show that this parameter can be used to determine the security strength of our GKPS.

Definition 2 (Deterministic k-Secure GKPS): A GKPS is said to be k-secure if GKPS can resist attacks that capture up to k sensors.

After deploying sensors in a WSN, attackers may try to capture sensors and recover secret tokens. The parameter k is used to evaluate the security strength of a GKPS and its ability to resist such attacks. In most polynomial-based key distribution schemes, adjusting the degree of the polynomial is the only way to defend against a sensor capture attack.

Definition 3 (Probabilistic k-Secure GKPS): A GKPS is said to be probabilistic k-secure if GKPS can resist an attack by capturing up to k sensors. Furthermore, after capturing k+1 sensors or more, it is probable that the adversary can successfully compromise the security.

Most existing polynomial-based key distribution schemes are deterministic k-secure schemes. To elaborate, capturing $k + 1$ or more than $k + 1$ sensors enables the adversary to successfully compromise the security of the network. Regardless, our GKPS is a probabilistic k-secure scheme. It is probabilistic that the adversary can successfully compromise the security. One of our design goals is to lower this probability in order to strengthen the security of the scheme. Our GKPS is very flexible since changing parameters of the GKPS can effectively lower this probability.

Definition 4 (j-Length GKPS): A GKPS is said to be j-length if GKPS can establish a group key among j+1 sensors.

Most key establishment schemes in WSNs can only establish a pairwise secret key between two sensors. One unique feature of our proposed GKPS is that it can establish a group

key among multiple sensors, so data can be protected by a path key [4]. A path key with length j involves $j+1$ sensors. So, the collected data in a WSN can be routed and protected by a path key. The parameter j is determined by many factors, such as the geographic size of the WSN, the total number of sensors, and the transmission distance of each sensor. In our proposed GKPS, we can adjust this parameter j, in order to facilitate an end-to-end secure communication.

Definition 5 (Connectivity): Sensors are said to be connected to each other if any two sensors share a common secret key.

Connectivity is a property of a WSN that determines whether information can be securely transmitted within a WSN. A deterministic key establishment scheme guarantees a shared pairwise key between any two arbitrary sensors. Thus, a deterministic key establishment ensures a connected network. On the other hand, probabilistic key establishment schemes, such as the random key scheme [7], does not guarantee a pairwise key between each pair of sensors within the network. As a result, such probabilistic key establishment schemes do not necessarily guarantee connectivity. When evaluating the schemes, the probability of connectivity is a parameter used to evaluate the performance of a probabilistic GKPS. In our proposed GKPS, we can increase the probability of connectivity by adjusting the parameters of the scheme.

Definition 6 (Probability of Connectivity With Path Length j P_j): The probability that any j+1 sensors can establish a group key (path length is j).

In most key establishment schemes, pairwise keys are used to protect transmitted data. The path length of these schemes is always restricted to equal one. However, our proposed GKPS can establish group keys with different path lengths. The parameter, P_j, is the probability of successfully establishing a path key involving $j+1$ sensors. In the performance section, we will discuss how to adjust this parameter.

Definition 7 (Probability of Connectivity P_c): The probability that any two sensors can establish a shared key.

This parameter is the probability that data can be protected and transmitted securely in a WSN.

IV. PROPOSED SCHEMES

In this section, we propose different schemes: the basic scheme, a modified scheme, and the proposed GKPS in detail.

A. BASIC SCHEME

We assume that there are l sensors, $S_i, i = 1, 2, \ldots, l$.

1) TOKEN GENERATION

The key generation center (KGC) needs to select l different polynomials, $f_i(x_i), i = 1, 2, \ldots, l$, and use them to generate tokens for sensors. Each polynomial is a univariate polynomial having $t-1$ degree. The token for each sensor, S_i, is $T_i = f_i(ID_i)\prod_{j=1,j\neq i}^{l} f_j(x_j) \bmod N$, where ID_i is the public information of a sensor, S_i, and N is the RSA modulus [21] which is the product of two large primes, p and q.

2) GROUP KEY ESTABLISHMENT

The group key, $K = \prod_{j=1}^{l} f_j(ID_j) \bmod N$, shared among l sensors, $S_i, i = 1, 2, \ldots, l$, can be computed by each sensor, S_i, using its secret token, T_i, and other sensors' IDs by computing $K = f_i(ID_i)\prod_{j=1,j\neq i}^{l} f_j(ID_j) \bmod N$.

Remark 1: The basic scheme can not only establish the group key for l sensors, $S_i, i = 1, 2, \ldots, l$, but can also establish group keys for any k (i.e., $2 \leq k \leq l$) sensors. For example, the group key among sensors, $S_i, i = 1, 2, \ldots, k$, is $K = \prod_{j=1}^{k} f_j(ID_j)\prod_{j=k+1}^{l} f_j(0) \bmod N$.

a: EXAMPLE 1

Assume that there are 3 sensors, S_1, S_2, and S_3. The KGC will select 3 polynomials, $f_1(x_1), f_2(x_2), f_3(x_3)$,and generate the tokens, $f_1(ID_1)f_2(x_2)f_3(x_3)$ for S_1, $f_1(x_1)f_2(ID_2)f_3(x_3)$ for S_2, and $f_1(x_1)f_2(x_2)f_3(ID_3)$ for S_3, where ID_i is the public information of S_i. Note that each polynomial evaluation is computed using a RSA modulus N. As a result, a group key, $f_1(ID_1)f_2(ID_2)f_3(ID_3)$, can be shared among the 3 sensors, S_1, S_2, S_3, and a pairwise key can also be shared between any two sensors. For example, the key,$f_1(ID_1)f_2(ID_2)f_3(0)$, can be shared between S_1, S_2.

b: SECURITY

We need to point out here that after capturing one sensor by the attacker, it is quite easy to recover the secret token of the sensor if the token is stored without using any tamper-resistant technology. In our proposed scheme, since we do not employ any tamper-resistant hardware, our scheme cannot prevent such an attack. On the other hand, our scheme is *probabilistic k-secure*, after capturing $k+1$ or more than $k+1$ sensors. In the following theorem, we demonstrate how attackers have an extremely low probability in obtaining the secret polynomial used to generate tokens of all sensors.

Theorem 1: The adversaries cannot obtain any information of secret polynomials selected by KGC.

Proof: In the analysis of the sensor capture attack, we classify the attacks into two types.

1) *Capturing one sensor-* It is obvious that by capturing any single sensor S_i, and obtaining the token $T_i = f_i(ID_i)\prod_{j=1,j\neq i}^{l} f_j(x_j) \bmod N$, the adversary cannot recover information of any individual polynomial, $f_i(x_i), i = 12, \ldots, l$, nor the product of all individual polynomials, $\prod_{j=1}^{l} f_j(x_j) \bmod N$.
2) *Capturing all sensors-* Assume that l sensors, $S_j \in c_j, j = 1, 2, \ldots, l$ (all sensors belong to different classes) with their public *ID*s, $ID_j \in$, $j = 1, 2, \ldots, l$, respectively, have been captured by an adversary. Then, multiplying their tokens $T_j = f_j(ID_j)\prod_{i=1,i\neq j}^{l} f_i(x_i) \bmod N, j = 1, 2, \ldots, l$, the adversary can obtain the product $\prod_{i=1}^{l} f_i(ID_i)(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$. Consequently, the adversary can remove $\prod_{i=1}^{l} f_i(ID_i)$ from the product and obtain $(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$. Next, the adversary may try to substitute $x_i = ID'_i, i = 1, 2, \ldots, l$, where ID_i's are

identities, into the polynomial $(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$, and gets $(\prod_{i=1}^{l} f_i(ID_i)')^{l-1} \bmod N = K^{l-1} \bmod N$. Based on the RSA assumption in [21], it is computationally infeasible to solve K. On the other hand, it is computationally impossible to solve the $(l-1)$-th root of $(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$ to obtain the secret product of polynomials,$(\prod_{i=1}^{l} f_i(x_i)) \bmod N$.

In the basic scheme, each sensor, S_i, needs to store a token, $T_i = f_i(ID_i) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$, which is a product polynomial of $l-1$ univariate polynomials where each individual polynomial has degree $t-1$. Each sensor stores t^{l-1} coefficients of the product polynomial in Z_N. The storage space is exponentially proportional to the number of sensors. The following modified scheme can be used to reduce storage from exponential complexity to linear complexity.

B. MODIFIED SCHEME

This section explains a modified version of the basic scheme aimed at reducing the storage requirements from exponential complexity to linear complexity.

1) TOKEN GENERATION

The KGC follows the same procedure to generate tokens for all sensors as described in the basic scheme. The token for each sensor S_i, is $T_i = f_i(ID_i) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$. In addition, for each token, the KGC will randomly selects $l-1$ secret integers $a_j \in Z_N, j = 1, 2, \ldots, l, j \neq i$, such that $f_i(ID_i) = a_1 a_2 \ldots a_{i-1} a_{i+1} \ldots a_l \bmod N$, and uses them to divide the token, $T_i = f_i(ID_i) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$, into $l-1$ sub-tokens, $s_{i,j} = a_j f_j(x_j), j = 1, 2, \ldots, l, j \neq i$. Note that the multiplication of all sub-tokens $\prod_{j=1, j\neq i}^{l} s_{i,j} = \prod_{j=1, j\neq i}^{l} a_j f_j(x_j)$, can recover the original token, $T_i = f_i(ID_i) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$. Each sensor is pre-loaded with sub-tokens, $s_{i,j} = a_j f_j(x_j), j = 1, 2, \ldots, l, j \neq i$.

Since each sub-token is a univariate polynomial, storage of each sensor is the coefficients of $l-1$ univariate polynomials. In other words, the total storage of this modified scheme is $t(l-1)$; which results in a linear complexity.

Theorem 2: The security of the modified scheme is the same as the basic scheme.

Proof: For each sensor S_i, it stores $l-1$ sub-tokens, $s_{i,j} = a_j f_j(x_j), j = 1, 2, \ldots, l, j \neq i$. Since $l-1$ integers, $a_j \in Z_N, j = 1, 2, \ldots, l, j \neq i$, are randomly selected by the KGC for every sensor, it is computationally impossible to recover any individual polynomial, $s_{i,j} = a_j f_j(x_j), j = 1, 2, \ldots, l, j \neq i$, from its sub-tokens. The only information available when capturing any sensor is obtaining the token that provides the same knowledge obtained when capturing a sensor in the basic scheme.

C. PROPOSED GKPS

In most sensor network applications, a large number of sensors has to be deployed in order to cover a wide geographical area. If the number of sensors, n, are too large, it is impractical to implement the above modified scheme since it requires a large storage space of each sensor (i.e., the storage is $t(n-1)$).

1) TOKEN GENERATION

The KGC evenly divides n sensors into l classes $c_i, i = 1, 2, \ldots, l$, and each class, c_i, is associated with a distinct polynomial, $f_i(x_i)$, with degree $t-1$ each. Tokens of sensors in the same class are generated by the KGC using the same formula but with different IDs. For example, for two sensors, $S_1$and $S_2 \in c_i$, with $ID_{i,1}$ and $ID_{i,2}$, respectively, the tokens are $f_i(ID_{i,1}) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$, and $f_i(ID_{i,2}) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod modN$, respectively. For token, $T_{i,1} = f_i(ID_{i,1}) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$, the KGC will randomly select $l-1$ integers, $a_j \in Z_N, j = 1, 2, \ldots, l, j \neq i$, such that $f_i(ID_{i,1}) = a_1 a_2 \ldots a_{i-1} a_{i+1} \ldots a_l \bmod N$, and use them to divide the token, $T_{i,1} = f_i(ID_{i,1}) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$, into $l-1$ sub-tokens, $s_{i,j} = a_j f_j(x_j), j = 1, 2, \ldots, l, j \neq i$. Note that the multiplication of all sub-tokens, $\prod_{j=1, j\neq i}^{l} s_{i,j} = \prod_{j=1, j\neq i}^{l} a_j f_j(x_j)$, enables the recovery of the original token, $T_{i,1} = f_i(ID_{i,1}) \prod_{j=1, j\neq i}^{l} f_j(x_j) \bmod N$. Sub-tokens, $s_{i,j}, j = 1, 2, \ldots, l, j \neq i$, are stored in sensor S_1.

2) GROUP KEY ESTABLISHMENT

Multiple sensors, which belonging to different classes can establish a group key as described in the basic scheme. However, sensors belonging to the same class cannot establish a group key. For example, we consider l sensors with their public IDs, $ID_j, j = 1, 2, \ldots, l$,respectively. If $S_j \in c_j, j = 1, 2, \ldots, l$, then the shared group key is $\prod_{i=1}^{l} f_i(ID_i)$, which can be computed by all sensors in the group. On the other hand, if there are only two sensors, S_{l-1} and S_l, in the subset of sensors, $\{S_j, j = 1, 2, \ldots, l\}$, belonging to the same class (i.e., $S_j \in c_j, j = 1, 2, \ldots, l-2$, and $S_{l-1}, S_l \in c_{l-1}$), then the shared group key among $l-2$ sensors belonging to different classes, $S_j \in c_j$, $j = 1, 2, \ldots, l-2$, is $f_{l-1}(0) f_l(0) \prod_{i=1}^{l-2} f_i(ID_i)$.

a: SECURITY ANALYSIS

This section demonstrates the security analysis of the proposed GKPS against sensor capture attacks.

Theorem 3: The proposed GKPS can resist attacks in capturing up to $t-1$ sensors in which all captured sensors should belong to the same class.

Proof: In the analysis of the sensor capture attacks, we classify the attacks into two types:

1) *All captured sensors belong to the same class-* Assume that t sensors, $S_j \in c_1, j = 1, 2, \ldots, t$, (all sensors belong to the same class c_1) with their public IDs, $ID_j, j = 1, 2, \ldots, t$, respectively, have been captured by an adversary. Then, following Lagrange interpolation on these tokens, $T_j = f_1(ID_j) \prod_{i=2}^{l} f_i(x_i) \bmod N, j = 1, 2, \ldots, t$, the adversary can obtain the product of secret polynomials, $(\sum_{j=1}^{t} f_1(ID_j) \prod_{i=1, i\neq j}^{t} \frac{x_1 - ID_j}{ID_i - ID_j}) \prod_{i=2}^{l} f_i(x_i) \bmod N = \prod_{i=1}^{l} f_i(x_i)$, necessary to break the security of our proposed GKPS. Note that the adversary can only obtain the product of all existing polynomials but

cannot obtain nor produce the individual polynomials. However, if the number of captured sensors is equal to or fewer than $t-1$, the adversary cannot obtain the product of all individual polynomials, $\prod_{i=1}^{l} f_i(x_i)$, since the degree of each individual polynomial is $t-1$. Furthermore, all captured sensors need to be in the same class in order for the Lagrange interpolation to work properly.

2) *All captured sensors belong to different classes-* Assume that l sensors, $S_j \in c_j, j = 1, 2, \ldots, l$ (all sensors belong to different classes) with their public *ID*s, $ID_j, j = 1, 2, \ldots, l$, respectively, have been captured by an adversary. Then, from Theorem 1, by multiplying their tokens, $T_j = f_j(ID_j)\prod_{i=1, i\neq j}^{l} f_i(x_i) \bmod N, j = 1, 2, \ldots, t$, the adversary can obtain the product, $\prod_{i=1}^{l} f_i(ID_i)(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$. By removing $\prod_{i=1}^{l} f_i(ID_i)$ from the product, the adversary can get $(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$. Regardless, it is computationally impossible to solve for the $(l-1)$-th root of $(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$ and obtain the secret product of the polynomials, $(\prod_{i=1}^{l} f_i(x_i)) \bmod N$. On the other hand, the adversary may try to substitute $ID'_i, i = 1, 2, \ldots, l$, where $\forall ID'_i \notin \{ID_i, i = 1, 2, \ldots, l\}$, into the polynomial, $(\prod_{i=1}^{l} f_i(x_i))^{l-1} \bmod N$, to get $(\prod_{i=1}^{l} f_i(ID'_i))^{l-1} \bmod N = K^{l-1} \bmod N$, where ID'_i are identities of the group key, K. Nonetheless, based on the RSA assumption [21], it is computational infeasible to get the key, K.

Remark 2: Note that the sensor capture attack as described in the aforementioned theorem can only be applied if all captured sensors are in the same class. This condition increases the difficulty of sensor capture attack since captured sensors are randomly distributed in WSNs. In summary, our proposed GKPS effectively reduces the risk of a sensor capture attack since this attack only works if the following two conditions are satisfied simultaneously: (a) having captured t or more than t sensors; and (b) among all captured sensors, there must exist at least t sensors belonging to the same class. In the performance analysis section, we prove that being able to capture sensors from the same class has a very low probability.

Remark 3: If we limit the number of sensors in each class to be less than or equal to the degree of each individual polynomial (i.e., $\lfloor \frac{n}{l} \rfloor \leq t-1$), then the sensor capture attack described in Theorem 3 can never endanger the security of our proposed GKPS.

Remark 4: The degree of each individual polynomial determines the competence of the GKPS in resisting the sensor capture attack. If the degree of each polynomial is $t-1$, then the WSN can resist attacks capturing up to $t-1$ sensors in which all captured sensors belong to the same class. Increasing the degree of the polynomials can strengthen the security; but that increases the storage and computational requirements of the sensors (will discuss this in the next section.)

b: PROPERTIES OF GROUP KEYS

1) *Non-interactive key establishment-* Sensors within a group can establish the group key using its secret token and all other sensors' public identities. After forming the group, there is no need to exchange any information among sensors.
2) *Secrecy of group keys-* In our proposed GKPS, sensors in different classes can establish a group key. The group key is a function of the individual polynomials associated with classes and sensors' identities. Any sensor not belonging to the group cannot obtain this group key. For example, we consider l sensors, with their public *ID*s, $ID_j, j = 1, 2, \ldots, l$, respectively. If $S_j \in c_j, j = 1, 2, \ldots, l$, then the shared group key is $\prod_{i=1}^{l} f_i(ID_i)$, which can be computed by all the sensors in the group. On the other hand, for any other sensor, $S'_1 \in c_1$, with ID'_1, computing the group key from its token, $f_1(ID'_1)\prod_{i=2}^{l} f_i(x_i)$ is not feasible.
3) *Key independence-* Each group key is a function of individual polynomials associated with classes and sensors' identities. Thus, each group key is independent of other group keys. However, if an attacker compromises t or more than t group keys belonging to a special subset, the attacker can recover all other group keys. The following theorem describes this type of known group key attack.

Theorem 4: Known group key attack- If t or more than t group keys in a special subset are compromised by the adversary, then adversary can use the compromised group keys to recover other group keys.

Proof: We use the following example to describe this special subset of compromised group keys. We assume that the attacker has compromised t group keys, $K_j = f_1(ID_{1,j})f_2(ID_2)\ldots f_l(ID_l), j = 1, 2, \ldots, t$. Note that in this special subset of group keys, there is only one identity, $ID_{1,j}, j = 1, 2, \ldots, t$, which is a variable that represents t sensors belonging to the same class, c_1, whereas the rest of the identities are fixed values. Using Lagrange interpolating formula, the attacker can obtain $\left\{\sum_{j=1}^{t} f_1(ID_{1,j})\prod_{i=1, i\neq j}^{t} \frac{x_1 - ID_{1,j}}{ID_{1,i} - ID_{1,j}})\right\}\left\{\prod_{i=2}^{l} f_i(ID_i)\right\} \bmod N = f_1(x_1)\prod_{i=2}^{l} f_i(ID_i)$. The attacker then can use this result to compute other group keys, $K_i = f_1(ID_{1,j})f_2(ID_2)\ldots f_l(ID_l), \forall j, j \neq 1, 2, \ldots, t$.

Remark 5: The probability of this type of known group key attack is extremely low since it requires all captured group keys to belong to a special subset of group keys. Furthermore, the usefulness of this attack is very limited since the recovered keys must belong to a special subset of group keys as well.

V. PERFORMANCE

In this section, we demonstrate the security analysis and performance analysis in terms of storage, computation, and connectivity of our proposed scheme. First, let us define the notations used in the section:

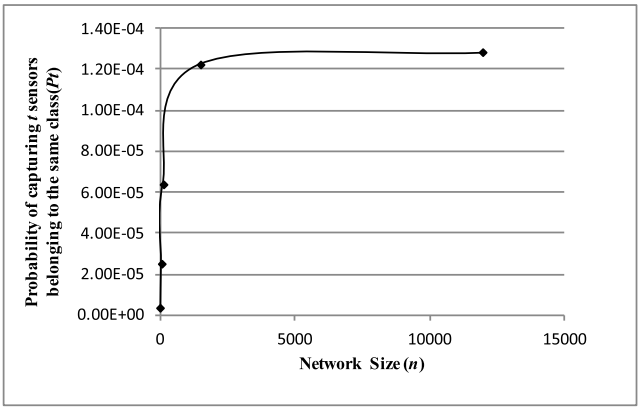

FIGURE 2. The probability of capturing t sensors that belong to the same class (P_t) with various number of sensors (for $t = 6$ and $l = 6$).

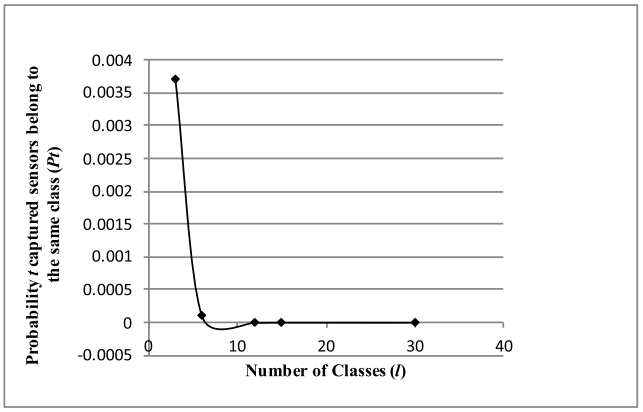

FIGURE 3. The probability of capturing t sensors belonging to the same class (P_t) while varying the number of classes (for $n = 300$ and $t = 6$).

n: number of sensors in WSN
l: number of classes of sensors
$t - 1$: degree of each individual polynomial
m: number of sensors in each class (i.e., $m = \lfloor \frac{n}{l} \rfloor$)

A. SECURITY

From Theorem 3, our proposed GKPS is a probabilistic $(t - 1)$-secure GKPS. In other words, our scheme can resist attacks of capturing up to $t - 1$ sensors. After t or more than t sensors are captured, if and only if all captured sensors belong to the same class can the adversary successfully compromise the GKPS. Concluding, the ability of an adversary to compromise the security of our scheme is proven probabilistic upon the preceding assertions. If t sensors within a network are captured, the probability that all captured sensors belong to the same class (P_t) is $P_t = \frac{C_i^m . l}{C_i^n}$. Fig. 2 exhibits the probabilities of P_t for varying numbers of sensors. In this analysis, it is proven that as network size increases, the probability of capturing t sensors belonging to the same class increases. However, the increases in probability are quite small (i.e., $P_t = 0.00006367$ for $n = 120$) and can almost be disregarded. Thus, a sensor capture attack will not affect security if the network size is increased. Fig. 3 shows the probabilities of P_t for a different number of classes within a network. The figure exhibits that increasing the number of classes can significantly decrease the probability of capturing t sensors belonging to the same class. In Fig. 4, the probability P_t is sharply decreased with large thresholds (i.e., $t \geq 4$). From these results, it is proven that increasing either the number of classes, l, or the threshold value, t, can effectively lower the probability P_t. This result demonstrates that our GKPS is very flexible to enhance the security of the polynomial-based key distribution scheme. The design objective is to lower this probability P_t as much as we can.

B. STORAGE REQUIREMENTS

In the proposed GKPS, each sensor needs to store $l - 1$ sub-tokens, $s_{i,j} = a_j f_j(x_j), j = 1, 2, \ldots, l, j \neq i$, where each $f_j(x_j)$ is a univariate polynomial having degree $t - 1$ with

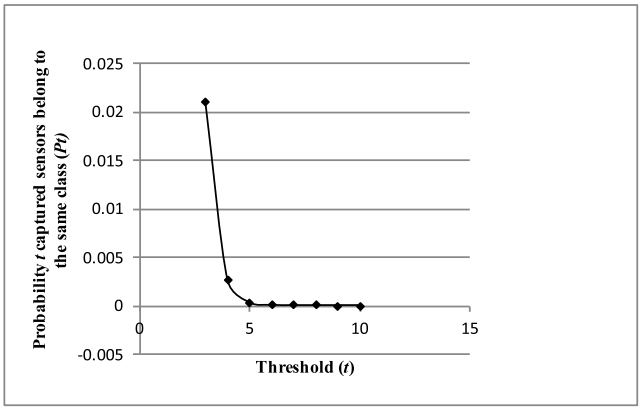

FIGURE 4. The probability distribution of all sensors belonging to the same class (P_t) when t sensors are captured; with various threshold values (for $n = 60$ and $l = 6$).

coefficients in Z_N. In other words, the storage requirement is $(l - 1)t$ coefficients in Z_N, which is a linear complexity.

C. COMPUTATIONAL REQUIREMENTS

We evaluate the computational effort to establish a path key with full length $l - 1$. Each sensor, S_i, must use its sub-tokens to compute $\prod_{j=1, j\neq i}^{l} a_j f_j(ID_j)$, where ID_j, $j = 1, 2, \ldots, l, j \neq i$. In other words, each sensors needs to evaluate $l - 1$ univariate polynomials with the degree $t - 1$. Each polynomial evaluation can follow Horner's rule [22] which requires $t - 1$ multiplications and t additions. In total, each sensor needs to compute $(l - 1)(t - 1)$ multiplications in Z_N, which is a linear complexity.

D. CONNECTIVITY EVALUATION

We have proposed a probabilistic $(t - 1)$-secure $(l - 1)$-length GKPS. In general, parameters in our proposed GKPS are determined in the following manner. From the geographic size of a WSN and the communication distance of each sensor, the maximal length,$l - 1$, of a communication path in the WSN is determined first. Then, from the security requirement, the degree of each individual polynomial, $t - 1$, is determined. According to storage requirements of each sensor, we need to select sensors capable of storing at

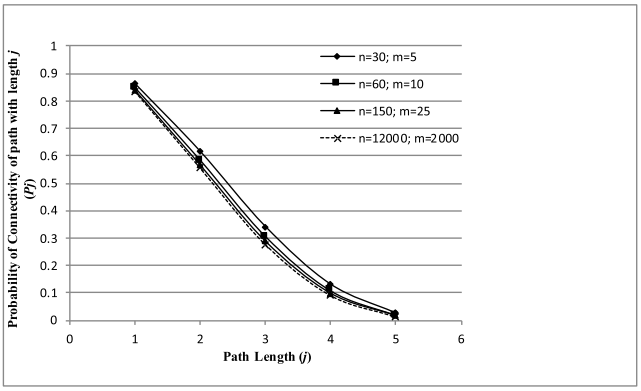

FIGURE 5. Probability distribution of connectivity with path length j.

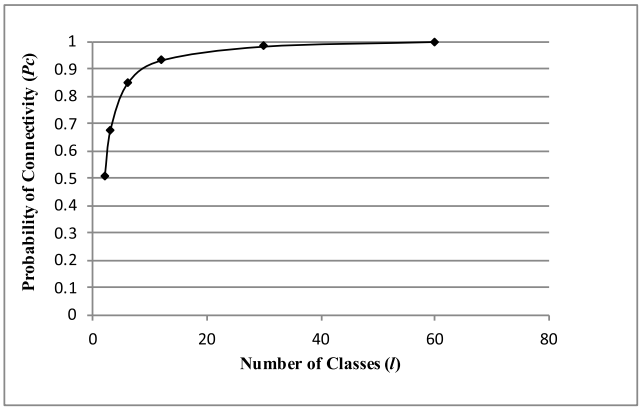

FIGURE 6. The probability distribution of network connectivity with different number of classes (for $n = 60$).

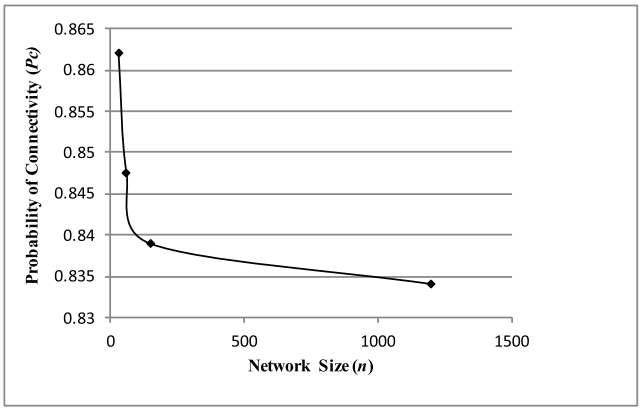

FIGURE 7. Probability distribution of network connectivity with different number of sensors.

least $(l-1)t$ integers in Z_N. Finally, the number of sensors, n, can be properly determined in order to provide adequate connectivity and satisfactory probability for establishing a path with a certain length.

1) *Probability of connectivity with path length j-* A path with length j must involve $j+1$ sensors. In our proposed GKPS, these $j+1$ sensors must all belong to different classes so that a group key can be established. The property P_j can be computed using the following formula, $P_j = \frac{C_{j+1}^{l} \cdot m^{j+1}}{C_{j+1}^{n}}$.
Fig. 5 shows the probabilities of $P_j, j = 1, 2, 3, 4, 5$, for the total number of classes, $l = 6$. The ability of our GKPS to establish varying-length path keys is a unique feature in our GKPS that distinguishes it from most pairwise key establishment schemes, in which paths are bounded by the length 1. Fig. 5 shows the probabilities P_j for different numbers of n. When the length of the path increases, the probability is gradually decreases. Moreover, the probability of connectivity between any two sensors is very high (i.e. 83%) and gradually decreases as the path length is increases. In addition, increasing the number of sensors in the network can only slightly affect the probability of connectivity. Therefore, we are able to increase the size of the network covering the entire geographical area and almost get the same connectivity as that of a smaller network. Note that increasing n will not affect the storage requirements of sensors.
2) *Probability of connectivity-* In our proposed GKPS having l classes, if two sensors belong to different classes, these two sensors are connected; otherwise, they are disconnected. The probability of dis-connectivity (P'_c) is: $P'_c = \frac{l.C_2^m}{C_2^n}$; and the probability of connectivity (P_c) is: $P_c = \frac{m^2.C_2^l}{C_2^n} = 1 - P'_c$.

One possible way to increase the probability, P_c, is to increase the number of classes in the WSNs. If the number of sensors, n, is fixed, increasing the number of classes, l, causes the number of sensors in each class to decrease. As a result, the probability that two sensors belong to the same class decreases, which in turn increases the probability of connectivity. Fig. 6 shows this probability P_c for different numbers of classes, l.

Note that if $l = n$, then each sensor belongs to a unique class. This situation is identical to the basic scheme, which is a deterministic key establishment scheme with $P_c = 1$. Notably, increasing the number of classes can increase both storage and computational costs of each sensor. This observation explains the motivation of our proposed GKPS, which is a flexible scheme since it can adjust parameters properly to balance the needs of high connectivity, strong security, and proper storage requirement for each sensor.

In case we need to deploy a large number of sensors to cover a large WSN, Fig. 7 shows this probability P_c for different numbers of sensors. Although this result shows that increasing the size of a network can slightly decrease the probability of connectivity, the probability, P_c, remains very high for large number of sensors (e.g. $P_c = 0.83, n = 1200$). This result demonstrates the merits of our proposed GKPS. It can provide a high probability of connectivity for a wide range of sensors.

VI. COMPARISON

In this paper, we proposed two GKPSs to establish "path keys" among sensors in order to facilitate secure

TABLE 1. Comparison among different key establishment schemes.

Schemes	Security assumption	Type of keys	Type of key establishment	Sensor capture attack
Random Key [7]	No assumption	Pairwise shared key	Probabilistic	Probabilistic secure
Group Key [19], [20]	RSA	Group keys	Deterministic	Deterministic k-secure
Basic scheme	RSA	Group keys	Deterministic	Deterministic k-secure
Proposed GKPS	RSA	Group keys	Probabilistic	Probabilistic k-secure

communications within sensor networks. Data transmitted over a communication path do not need to be protected using multiple pairwise shared keys in a link-to-link transmission. Instead, data can be protected by a single path key using an end-to-end transmission. Our proposed GKPS is a probabilistic $(t-1)$-secure$(l-1)$-length GKPS. Thus, our scheme's security is effectively strengthened compared to the deterministic schemes proposed in [19] and [20]. We summarize the comparison with other key establishment schemes in Table 1.

VII. CONCLUSION

A novel design of a group key pre-distribution scheme is introduced in this paper. The basic scheme is a deterministic key establishment scheme in which group keys are established among multiple sensors. The other GKPS is a probabilistic scheme in which group keys with different lengths can be established but with different probabilities. The security of our proposed GKPSs is based on the RSA assumption. For a large-sized WSN, our second scheme is a probabilistic $(t-1)$-secure $(l-1)$-length GKPS. One unique feature of our proposed GKPS is that it strengthens the security of polynomial-based schemes significantly. The storage and computational requirements of each sensor are very efficient. Finally, performance analysis is included.

REFERENCES

[1] T. Gao, D. Greenspan, M. Welsh, R. R. Juang, and A. Alm, "Vital signs monitoring and patient tracking over a wireless network," in *Proc. IEEE 27th Annu. Conf. Eng. Med. Biol.*, Jan. 2006, pp. 102–105.

[2] A. Milenković, C. Otto, and E. Jovanov, "Wireless sensor networks for personal health monitoring: Issues and an implementation," *Comput. Commun.*, vol. 29, pp. 2521–2533, Aug. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366406000508

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[4] L. Harn, C.-F. Hsu, O. Ruan, and M.-Y. Zhang, "Novel design of secure end-to-end routing protocol in wireless sensor networks," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1779–1785, Mar. 2016.

[5] C.-C. Lee, T.-H. Lin, and C.-S. Tsai, "A new authenticated group key agreement in a mobile environment," *Ann. Telecommun.-Annales Télécommun.*, vol. 64, no. 11, p. 735, 2009. doi: 10.1007/s12243-009-0096-z.

[6] C.-C. Lee, C.-T. Li, S.-T. Chiu, and Y.-M. Lai, "A new three-party-authenticated key agreement scheme based on chaotic maps without password table," *Nonlinear Dyn.*, vol. 79, no. 4, pp. 2485–2495. doi: 10.1007/s11071-014-1827-x.

[7] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *Proc. 9th ACM Conf. Comput. Commun. Secur. (CCS)*, 2002, pp. 41–47. doi: 10.1145/586110.586117.

[8] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," in *Proc. IEEE Symp. Secur. Privacy (SP)*, 2003, p. 197. [Online]. Available: http://dl.acm.org/citation.cfm?id=829515.830566

[9] W. Du, J. Deng, Y. S. Han, S. Chen, and P. K. Varshney, "A key management scheme for wireless sensor networks using deployment knowledge," in *Proc. INFOCOM*, vol. 1, 2004, p. 597.

[10] A. Rasheed and R. Mahapatra, "Key predistribution schemes for establishing pairwise keys with a mobile sink in sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 1, pp. 176–184, Jan. 2011.

[11] S. Ruj, A. Nayak, and I. Stojmenovic, "Pairwise and triple key distribution in wireless sensor networks with applications," *IEEE Trans. Comput.*, vol. 62, no. 11, pp. 2224–2237, Nov. 2013.

[12] O. Yağan and A. M. Makowski, "Wireless sensor networks under the random pairwise key predistribution scheme: Can resiliency be achieved with small key rings?" *IEEE/ACM Trans. Netw.*, vol. 24, no. 6, pp. 3383–3396, Dec. 2016. doi: 10.1109/TNET.2016.2527742.

[13] J. Ding, A. Bouabdallah, and V. Tarokh, "Key Pre-Distributions From Graph-Based Block Designs," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1842–1850, Mar. 2016.

[14] F. Gandino, R. Ferrero, and M. Rebaudengo, "A key distribution scheme for mobile wireless sensor networks: q -s -composite," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 34–47, Jan. 2017.

[15] R. Blom, "Non-public key distribution," in *Advances in Cryptology*. Boston, MA, USA: Springer, pp. 231–236, 1983. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4757-0602-4_22

[16] C. Blundo, A. De Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, "Perfectly-secure key distribution for dynamic conferences," in *Advances in Cryptology—CRYPTO* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 1992, pp. 471–486. [Online]. Available: https://link.springer.com/chapter/10.1007/3-540-48071-433

[17] E. Khan, E. Gabidulin, B. Honary, and H. Ahmed, "Matrix-based memory efficient symmetric key generation and pre-distribution scheme for wireless sensor networks," *IET Wireless Sensor Syst.*, vol. 2, no. 2, pp. 108–114, 2012.

[18] J.-P. Sheu and J.-C. Cheng, "Pair-wise path key establishment in wireless sensor networks," *Comput. Commun.*, vol. 30, nos. 11–12, pp. 2365–2374. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366407001764

[19] L. Harn and G. Gong, "Conference key establishment protocol using a multivariate polynomial and its applications," *Secur. Commun. Netw.*, vol. 8, no. 9, pp. 1794–1800, 2015. doi: 10.1002/sec.1143.

[20] L. Harn and C. F. Hsu, "Predistribution scheme for establishing group keys in wireless sensor networks," *IEEE Sensors J.*, vol. 15, no. 9, pp. 5103–5108, Sep. 2015.

[21] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978. doi: 10.1145/359340.359342.

[22] D. E. Knuth, *Art of Computer Programming: Seminumerical Algorithms*, vol. 2, 3rd ed. Reading, MA, USA: Addison-Wesley, 1997.

ASHWAG ALBAKRI received the B.S. degree in computer science from King Abdulaziz University, in 2010, and the M.S. degree in information science, specialized in security assured information systems, from the University of Pittsburgh, PA, USA, in 2015. She is currently pursuing the Ph.D. degree in computer science with the University of Missouri-Kansas City. Her research interests include cryptography, information security, and network security. She is also a Lecturer with Jazan University, Saudi Arabia.

LEIN HARN received the B.S. degree from National Taiwan University, in 1977, the M.S. degree from the State University of New York at Stony Brook, in 1980, and the Ph.D. degree from the University of Minnesota, in 1984, all in electrical engineering. He joined the Department of Electrical and Computer Engineering, University of Missouri-Columbia, in 1984, as an Assistant Professor, and in 1986, he moved to the Computer Science and Telecommunication Program, University of Missouri-Kansas City, where he was on development leave to work with the Racal Data Group in Florida for a year.