Fakhrey H, Tiwari R, Johnston M, Al-Mathehaji YA.

[The Optimum Design of Location-Dependent Key Management Protocol for a WSN With a Random Selected Cell Reporter.](#)

***IEEE Sensors Journal* 2016, 16(19), 7217-7226**

Copyright:

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

DOI link to article:

<http://dx.doi.org/10.1109/JSEN.2016.2594591>

Date deposited:

09/11/2016

The Optimum Design of Location-Dependent Key Management Protocol for a WSN with a Random Selected Cell Reporter

Harith Fakhrey*, *Student Member, IEEE*, Rajesh Tiwari, *Member, IEEE*, Martin Johnston, *Member, IEEE*,
Yasir Al-Mathehaji*, *Student Member, IEEE*,

Abstract—Wireless sensor networks (WSNs) for smart city applications can be widely deployed, but they can also be vulnerable to be compromised by outsider attackers. It is therefore critical that the disseminated information is secured in terms of its confidentiality, availability and authenticity. Many key management schemes have been proposed to guarantee the aforementioned requirements and the most popular depend on the location of the sensor to generate the required credentials. This paper presents a new location-dependent key management protocol, called the location-dependent key management protocol with random selected cell reporters (LKMP-RSCR). In this protocol, the generated report of an event must contain a third level of endorsement handled by a set of cell reporters to be accepted by the sink. As a result, the adversary needs to compromise all endorsement nodes in addition to the entire set of cell reporters to generate a fraudulent report from a specific region. A mathematical analysis is presented to evaluate the LKMP-RSCR and in contrast to existing LEDS and MKMP schemes, our new scheme shows a significant improvement in terms of data confidentiality, authenticity, computation and communication cost. Revocation algorithms for compromised nodes and cells are also presented to overcome the possible consequences of node compromising.

Index Terms—Wireless Sensor Network (WSN), Smart Cities, Location-Dependent Key Management System, Optimization, End-to-End Security.

I. INTRODUCTION

WIRELESS sensor networks are well suited for monitoring small and low-cost sensors that can be deployed over a wide terrain where resources are limited [1]. Hence, data security is absolutely essential and so cryptographic techniques are employed to ensure security requirements, such as *confidentiality*, *availability* and *authenticity* [2]. Whatever technique is used, particular security credentials are required to be stored inside each node and a key management system (KMS) is needed in order to generate and distribute these credentials. However, the scarce resources of a WSN mean that most KMS approaches used in conventional wireless networks are infeasible [3]–[5]. To solve this problem, most proposed solutions are based on the pairwise construction of keys between each node and others in its vicinity after the completion of the deployment process

[6], [7]. As a result, hop-by-hop security is used to protect data secrecy in WSNs where the majority of communication occurs between each node and other nodes rather than direct correspondence between each node and the sink. On the other hand, for WSNs deployed in smart city applications, such as road surveillance [8], traffic management [9] fire detection and healthcare monitoring, this is not the case. In such applications, the dominant communication pattern occurs only between each node in the event region and the sink and communication-pattern-oriented node capture attacks might risk a hop-by-hop security design [10]. Consequently, the *confidentiality*, *availability* and *authenticity* of the data may be affected. To solve this problem, many schemes are proposed to secure end-to-end communication as a crucial requirement in recent WSN applications [11]–[13]. In [14], the authors proposed a location-dependent key management protocol for a WSN, where a set of cell reporters were randomly selected by the base station to improve cell security. However, the protocol in [14] is not optimal and in this paper, an improved design of the location-dependent key management protocol is presented, which addresses the following issues:

- 1) Computational cost and data confidentiality/authenticity
- 2) An investigation into the optimum number of cell reporters.
- 3) A revocation scheme for the compromised nodes/cells.
- 4) A mathematical analysis of the communication cost.

Hence, this protocol provides data protection inside a WSN used for smart city applications in terms of data confidentiality, availability and authenticity.

A. Related Work

When an event happens inside a monitored region, a group of n nodes around the event are assumed to be able to detect it and then generate an event report. A collaborative signature created by a threshold number of e ($1 \leq e \leq n$) nodes validate the generated report. Otherwise, the intermediate cells and sink block any report that does not contain a valid endorsement. Based on this scheme, a few approaches have been proposed in the last decade. First, an interleaved hop-by-hop authentication (IHA) scheme [15] has been proposed

*Harith Fakhrey and Yasir Al-Mathehaji are staff members with the University of Baghdad in Iraq and sponsored by the Iraqi MoHESR to study his Ph.D.

to verify the generated reports in a deterministic fashion. The statistical en-route filtering SEF [16] uses a probabilistic approach in order to filter out fake reports. However, both IHA and SEF suffer from the same drawback where the compromising of n nodes results in no protection. To overcome this problem, location-based resilient secrecy (LBRs) [17] adopts two techniques: location-binding key generation and location-guided key selection. As a result, the usage of endorsement keys is limited to the region where the event occurs, which leads to the thwarting of attacks that globally use the credentials of compromised nodes. However, LBRs does not satisfy the data authenticity requirement since the compromising of n nodes inside a particular area might enable the adversary to create a fake event in that area. Additionally, data availability is not guaranteed where it is vulnerable to selective forwarding and report disruption attacks.

As an enhancement to the aforementioned schemes, location-dependent end-to-end data security (LEDS) has been proposed [10]. In this scheme, the terrain served by a WSN is divided into several cells by a virtual grid. Within each cell, both a node key and cell key are derived by each node individually, depending on their location and cell center location respectively. All generated reports contain an endorsement generated by a threshold number of nodes and the bogus reports are filtered out. It is clear that LEDS has notable advantages such as guaranteeing data availability, limited influence of node-compromising attacks and end-to-end security. However, the WSN size is assumed to be fixed and without any revocation scheme to overcome problems caused by the compromised nodes. Moreover, the node localization technique is infeasible due to its dependence on robots which is not applicable [18].

The multi-BS key management protocol (MKMP) [18] is a recent scheme inspired by LEDS. It shows an enhancement in terms of coverage and data security, power consumption and storage cost. In addition, MKMP presents a distributed key revocation scheme to solve compromised node problems. However, MKMP has the same disadvantages as LEDS in terms of node-compromising consequences. In both approaches, compromising a threshold amount of nodes (e) leads to an increase of the entire cell capturing probability. Consequently, a fake report can easily be generated by the adversary inside that cell, which would then be accepted by the sink without being dropped by intermediate nodes. Also, both schemes are challenged by a high communication overhead caused by bidirectional multi-hop communication between each particular cell and the BS, in addition to the computational cost caused by the frequent derivation of authentication keys and route set-up.

B. Contribution

Our proposed scheme has two novel contributions:

- The presence of cell reporters enhances data security by increasing the difficulty of generating a malicious

report by an adversary and decreasing the impact of compromising a certain number of nodes in the network.

- The inclusion of a hybrid communication scheme in the WSN is assumed, so that data transmitted by the BS to a particular node in the network is accomplished by a single hop mechanism. Conversely, data sent in the reverse direction is assumed to be transmitted by a multi-hop mechanism. This criterion ensures:
 - A decrease in communication overhead, where the packets sent by the BS are transmitted directly to each node without flooding the entire network.
 - Load balance in computational cost between sensor nodes and the BS, where some credentials are generated by the BS then disseminated to each node. Moreover, the computations related to routing will be implemented by the BS rather than each group of nodes.

This paper is organized as follows: The LKMP-RSCR protocol is discussed in section II. A security analysis of our system is discussed in section III. The communication and computation cost is analysed in section IV and an optimization analysis for the number of cell reporters is given in section V. Finally, conclusions are given in section VI.

II. PROPOSED SCHEME: THE LKMP-RSCR PROTOCOL

A. System Assumptions

The LKMP-RSCR protocol is assumed to be employed over a wide area of a smart city of a predetermined size and shape, monitored using a large-scale WSN comprising N limited resource nodes and a sink with unlimited resources, hereafter known as a base station (BS). This unit is responsible for data collection, control of both report verification and en-route filtering and the origin of requests sent to all nodes. The BS is assumed to be able to cover most sensor nodes in the monitored region as shown in Fig. 1. The served region is represented as a virtual grid of N' cells. All cells are assumed to have a similar number of sensor nodes that are in communication coverage of all other sensor nodes and are able to estimate their positions using secure localization schemes, such as [19]–[21]. It is also assumed that all elements in the network (nodes and BS) have a unique public identity (name) and a private identity (ID).

B. Threat Model

The system is assumed to be secure during the bootstrapping interval, a short period after the deployment of all elements, then the attacker is assumed to be able to capture randomly selected nodes and compromise their security credentials. On the other hand, the same adversary has no opportunity to compromise the BS due to its rigid security which prevent its facilities from being compromised or cloned. When a node is compromised, the attacker is assumed to be able to inject, drop, eavesdrop, alter, or retransmit packets. However, the attacker has no access to the uncaptured nodes.

C. Notation and Terms

In our study, the following definitions are of significant importance:

- K : An initial master key used as a seed to derive other keys
- (x_0, y_0) : The BS location
- Δ : The side length of each cell
- t : The number of authentication cells $d_i : i = 1, 2, \dots, t$
- p : A prime number
- (x_c, y_c) : The center location of cell (c)
- (x_a, y_a) : The location of a node (a)
- $\parallel$: The operation of concatenation
- H : A Hash function
- ID_a : Identity of each particular node (a) which is known by the BS
- t_s : A recent time slot
- K_{Lcin} : An initial cell key
- K_a^B : A unique key shared between each node (a) and the BS
- $K_c^{d_i}$: An authentication key derived by the BS and shared between cell mates in cell c and cell mates in the authentication cell d_i
- K_{Lc} : The cell key
- $Enc_K\{M\}$: Encryption of a message M using key K
- $MAC_K\{M\}$: The message authentication code of a message M calculated over the key K
- e : A threshold number of endorsement nodes required to generate a legitimate report
- T : A predefined cell reporter validity
- N : Total nodes in the network
- N' : Number of cells in the network
- n : Number of nodes of each cell
- z : Number of cell reporters
- λ : Packet size
- $P_{C\{e|z\}}$: The probability of compromising a cell in terms of data confidentiality
- $P_{auth\{e|z\}}$: The probability of compromising a cell in terms of data authenticity
- **The report forward route** between a particular cell (c) and the BS contains all cells traversed by a virtual line between them as shown in Fig. 1, denoted as dark-grey cells. The highlighted sequence is listed based on the position according to the BS.
- **Report authentication cell**: A particular cell d_i belongs to the forward report path of cell (c) . Its location relative to (c) or the last authentication cell is $t+1$ cells as shown by the light-grey cells depicted in Fig. 1. However, there is no authentication cell in the case of a short report authentication route less than $t+1$ cells.

D. Setup Phase

Each particular node (a) is preloaded with the following parameters $\{K, (x_0, y_0), ID_a, \Delta, t, p\}$. Depending on their clocks and the application requirements, the setup phase time is divided into multiple identical time slots t_s to ensure

Figure 1: Illustration of system construction shows report forward route and authentication cells for a WSN with $n \simeq 3$ and $t = 2$ [14]

the freshness of security credentials derivations. Using its location and the BS location as a reference point, the node excludes the center location of its cell. Based on these data, the remaining credentials are derived as shown in Algorithm-1. All nodes in each particular cell create a list of its

Algorithm 1 Derivation of security credentials inside each node (a) during the setup phase

Input: $K, (x_0, y_0), t_s, \Delta, (x_a, y_a)$
 $K_{ts} \leftarrow K \parallel t_s$
 $x_c \leftarrow 0.5 \left\lfloor \frac{x_a - x_0}{\Delta} \right\rfloor, y_c \leftarrow 0.5 \left\lfloor \frac{y_a - y_0}{\Delta} \right\rfloor$
 $K_{Lcin} \leftarrow H(K_{ts} \parallel (x_c, y_c))$
return $K_{ts}, (x_c, y_c), K_{Lcin}$

neighboring nodes, hereafter known as cell-mates, which is sent by each sensor node to the BS. This process is illustrated in Algorithm 2.

The message $\{LIST\}_a$ is sent by each node (a) to the BS using a cell-by-cell method and the BS then follows the steps shown in Algorithm 3 to verify the received data. In this algorithm, receiving a message sent by different nodes in the cell (c) via multiple paths means c contains no malicious nodes. However, if a malicious node is detected, a revocation scheme is implemented by calling Algorithm (4) or Algorithm (5) to revoke nodes and cells listed in $SusbNodes, SubCells$ respectively as shown in section (II-F) to overcome these effects. All correspondence from the BS to any node (a) is implemented directly via a single hop scheme due to the wide coverage property of the BS. An

Algorithm 2 Creation and broadcasting of cell-mate list of each node (a)

Input: $ID_a, t_s, (x_a, y_a), (x_c, y_c), (x_0, y_0), K_{Lcin}$
 $a \rightarrow$ All cell-mates inside: $Enc_{K_{Lcin}}\{ID_a, t_s, (x_c, y_c)\}$
 $\{CellMateList\}_a \equiv \phi$
for all cell-mates **do**
 $ACK \rightarrow a$
if ACK is valid **then**
 Update $\{CellMateList\}_a$
end if
end for
 $K_a^{BS} \leftarrow H(K \| ID_a \| (x_0, y_0))$
 $\{LIST\}_a \leftarrow \{((x_c, y_c), \{CellMateList\}_a, K_{Lcin})\}$
 $a \rightarrow$ BS: $Enc_{K_a^{BS}}\{LIST\}_a$
return $K_a^{BS}, \{LIST\}_a, \{CellMateList\}_a$

en-route-filtering scheme presented in [22] is implemented by the authentication cells d_i belonging to each particular cell (c) to reduce the amount of fake reports arriving at the BS. This is implemented using the authentication keys $K_c^{d_i}$ derived by the BS. Hence, the computation cost is decreased dramatically in comparison with the cost of LEDS and MKMP, as shown in section (IV-B), where the nodes are responsible about the derivation of authentication keys.

The modified setup phase presented in this paper is designed to overcome the drawbacks in [14] such as the high computation cost, data confidentiality and data authenticity vulnerabilities and the absence of an appropriate revocation scheme. To address the first point, the computational cost required to calculate the cell centre can be reduced by uploading Δ , which removes the required calculations for grid integer determination shown in [14](2). In addition, K_a^{BS} is confidential when it is generated only by node a , where both its identity ID_a and location (x_a, y_a) are involved in this key derivation. The authenticity of the HELLO message shown in Algorithm 2 is guaranteed due to its encryption by K_{Lcin} , which thwarts any possible attack. Moreover, as the main part of a revocation scheme, the detection of suspicious nodes/cells is included. Finally, the setup phase explains the derivation of both the ID_c and the $K_c^{d_i}$ as the main part of en-route filtering. Hence, this paper presents a more realistic setup phase in comparison to that presented in [14].

E. Report Generation

The event report R , which includes the cell ID, event location and event type, is generated depending on all n sensor nodes inside the particular cell (c) using the signal strength strategy presented in [23]. Due to the packet encryption by K_{LC} , R , it is difficult for outside attackers to obtain R . However, the attackers could inject fabricated information or create a forged event, so some type of endorsement needs to be embedded inside the generated reports.

For LKMP-RSCR, three kinds of endorsement must be included in the report for verification: uniqueness of the

Algorithm 3 The BS verification of $\{LIST\}_a$ packages sent by each node

Input: $\{LIST\}_a : a = 1, 2, \dots, N$
 $COUNT = 0; SuspNodes \equiv \phi$
for $a = 1, 2, \dots, N$ **do**
 BS extracts ID_a from the header of $\{LIST\}_a$
 $K_a^{BS} \leftarrow H(K \| ID_a \| (x_0, y_0))$
 $Dec_{K_a^{BS}}\{LIST\}_a = \{((x_c, y_c), ID_{list}, K_{Lcin})\}$
 $ID_c = 10x_c + y_c$
 $Num = Length\{LIST\}_a$
if $\{LIST\}_a \equiv \{LIST\}_{a-1}$ **then**
 $COUNT + +$
else
 $SuspNodes \leftarrow a$
end if
end for
if $COUNT \geq \lfloor 0.5Num \rfloor + 1$ **then**
 Call Algorithm(4)
else
if $COUNT < \lfloor 0.5Num \rfloor + 1$ **then**
 $SuspCells \leftarrow ID_c$
 Call Algorithm(5)
end if
else
 $K_{Lc} \equiv K_{Lc_init}$
 $\{LIST\}_c = \{LIST\}_a \cup ID_a$
 $BS \rightarrow a : Enc_{K_a^{BS}}\{K_{Lc}, \{LIST\}_c\}$
if $BS \leftarrow a : ACK$ **then**
 $K_c^{d_i} = H(K_{Lc} \| K_{Ld_i} \| (x_{d_i}, y_{d_i}) \| (x_c, y_c))$
 $BS \rightarrow a \in c : \{d_i\} \cup \{K_c^{d_i}\}$
end if
end if
return $K_a^{BS}, SuspNodes, SuspCells, K_{Lc}, \{LIST\}_c, \{d_i\}, \{K_c^{d_i}\}$

individual data received from each node in the event cell, the message authentication code (MAC) generated by the authentication nodes and the cell reporter signature. The first endorsement is achieved by using the fundamentals of an (e, n) threshold linear secret sharing scheme (LSSS) [24]. The uniqueness of the K_a^{BS} key shared between each particular cell and the BS is used to derive a unique node share C_a from the encrypted event report $C = E_{K_{LC}}\{R\}$:

$$C_a = C \sum_{0 \leq i \leq e-1} (K_a^{BS})^i \bmod p, \quad (1)$$

Obviously, C_a is generated uniquely by node a depending on its unique key K_a^{BS} which is shared with the BS only. Moreover, each node inside cell (c) broadcasts its share, as a tuple $\{C_a, a\}$ to all its cell-mates. As a result, a total of $n - 1$ shares are collected by each node and concatenated:

$$C_{new} = C_1 \| C_2 \dots \| C_n \quad (2)$$

The second endorsement of the report consists of multiple MACs calculated over C_{new} . These MACs are derived

using the authentication keys dedicated and broadcast to each node by the BS (Algorithm 3). For example, in the case of dedicating two authentication cells $d1$ and $d2$ as intermediate cells between cell (c) and the BS, each node inside (c) broadcasts the following to its cell-mates.

$$MAC_{K_{Lc}}\{C, MAC_{K_c^{d1}}(C_{new}), MAC_{K_c^{d2}}(C_{new})\}$$

When node (a) receives different $i(n-1)$ MACs, where i refers to the number of authentication cells, it sends a synthesized report containing the ID of all cell-mates and the hosting cell, C_{new} and i MACs, shared with each authentication cell. A random timer scheme [17] is used to avoid report duplication. Based on the enclosed MACs, the en-route filtering scheme is implemented by each authentication cell. Due to space limitation, this scheme is not presented and the same procedure of [10] is followed. The third and the most important endorsement from LKMP-RSCR is the signature of the set of z cell reporters randomly selected out of a total of n cell nodes by the BS.

This set is changed every $\frac{1}{T}$ seconds, where T is a predefined cell reporter validity which may be changeable based on multiple parameters like data importance, estimated frequency of attacks and rhythm of event occurrences. The participation of the cell reporter signatures in the report generation overcomes the major security limitations in recent schemes. As described in section (I-A), the capture of a threshold number of nodes in a particular cell enables the adversary to generate fake reports that might deceive all verification process of the intermediate cell and the BS. Therefore, the received packet from the BS in LKMP-RSCR is accepted if and only if all cell reporters are involved in the report generation.

F. Key Revocation

In this section, the required key revocation schemes are explained to overcome the possible risk caused by nodes or cells listed in previously created *SusbNodes*, *SusbCells* lists. These schemes are crucial to prevent the compromised nodes/cells from starting any correspondences with other nodes in the network. Moreover, such a scheme is important to thwart colluding between compromised nodes. For all nodes $\in$ *SusbNodes* list, the BS call Algorithm (4) to revoke their credentials as shown below.

On the other hand, Algorithm (5) is called by the BS regarding each cell $sc \in$ *SusbCells*:

III. SECURITY ANALYSIS OF LKMP-RSCR

An analysis of the security of our system is presented based on three aspects. The first aspect is the system's capability of thwarting typical routing attacks while the second aspect is related to the role of the number of cell reporters (z) in the security of each particular cell. Finally, the third aspect is a revision of the wireless security requirements for data confidentiality, availability and authenticity derived in [14]. The outcome of the last two aspects is the cornerstone of section V, which investigates the optimum value of (z).

Algorithm 4 Revocation of suspicious node (s) located inside a cell (c) implemented by the BS

Input: $ID_s \in$ *SusbNodes*, $ID_c, K_{Lc}, K_c^{d_i} : i = 1, 2 \dots t$
 $\{LIST\}_{new} = \forall ID_a (ID_a \in \{LIST\}_c \wedge ID_a \neq ID_c)$
 $K_{Lc_new} \leftarrow H((x_c, y_c || ts || \{LIST\}_{new}))$
 $K_c^{d_i} = H(K_{Lc_new} || K_{Ld_i} || (x_{d_i}, y_{d_i}) || (x_c, y_c))$
for $\forall a(a \in \{LIST\}_{new})$ **do**
 $BS \rightarrow a : \{K_{Lc_new}, ID_c, K_c^{d_i} : i = 1, 2 \dots t\}$
end for
 $Remove K_c^{BS}$
return $\{LIST\}_{new}, K_{Lc_new}, K_c^{d_i} : i = 1, 2 \dots t\}$

Algorithm 5 Revocation of suspicious cell (sc) implemented by the BS

Input: $ID_{sc}, K_{Lsc}, K_{sc}^{d_i} : i = 1, 2 \dots t$
 $Remove K_{sc}^{BS}$
for $\forall d_i : i = 1, 2 \dots t$ **do**
 $Remove K_{sc}^{d_i}$
end for

A. System Robustness Against Routing Attacks

As presented earlier, LKMP-RSCR uses the node positions to derive the security credentials required to protect both cell-by-cell communication and the required correspondence between cell-mates. Hence, our scheme is secure enough to thwart almost all typical attacks such as node replication, HELLO flooding, black hole, wormhole, Sybil and selective forwarding attacks.

B. The impact of z value on the security of each particular cell

The assumption of our LKMP-RSCR protocol ensures the validity of any received report depending on three aspects: the endorsement of e sensor nodes, MAC authentication nodes and the participation of z cell reporters. While the importance of the first two aspects are well investigated in [14], which are reviewed in following sections, this section will investigate the importance of the value of z in determining the probability of compromised cells. The probability of an adversary compromising all z cell reporters inside a cell that contains n nodes can be calculated using the compromising strategy illustrated in the following example: Assume that $z = 3$ and $n = 10$:

The adversary has a:

- 1) probability of $P(E1) = \frac{3}{10}$ to compromise all z cell reporters in her 1st trial.
- 2) probability of $P(E2) = \frac{2}{9}$ to compromise the remaining 2 cell reporters from the remaining 9 nodes in the 2nd trial.
- 3) probability of $P(E3) = \frac{1}{8}$ to compromise the last cell reporter from the remaining 8 nodes in the 3rd trial.

The probability of compromising all 3 cell reporters in a row is:

$P = P(E1)P(E2)P(E3) = \frac{1}{120}$ This can be generalised as follows:

$$P_{z_comp} = \left(\frac{z}{n}\right)\left(\frac{z-1}{n-1}\right)\left(\frac{z-2}{n-2}\right)\dots\left(\frac{1}{n-z+1}\right)$$

$$P_{z_comp} = \frac{z(z-1)(z-2)\dots(1)}{n(n-1)(n-2)\dots(n-z+1)}$$

$$= \frac{(z!)(n-z)!}{n!} \quad (3)$$

The above mentioned equation is describing the relationship between the probability of compromising a particular cell and the number of its cell reporters (z). This equation has a vital importance while it is the cornerstone that is used to determine the optimum value of z as shown in section V.

C. Security Strength Regarding Data Confidentiality

The content of the generated report by an event cell is only revealed to its nodes because it is encrypted by the cell key K_{Lc} owned by them. This guarantees data confidentiality even when a number of intermediate nodes are compromised. However, if one of the nodes involved in report generation is compromised, the report contents could be revealed. In order to understand the security strength in this case, the effect of a random node capture attack (RNCA) on the entire system has been investigated by determining the probability of compromising all cells due to RNCA in [14].

$$P_{C\{e|z\}} = \left(1 - \frac{\binom{N-e}{x}}{\binom{N}{x}}\right) \left(1 - \frac{\binom{N-z}{x}}{\binom{N}{x}}\right) \quad (4)$$

For different values of z , the percentage of captured cells concerning data confidentiality in terms of the number of compromised nodes is shown in Fig. 2. The security of LKMP-RSCR in terms of data confidentiality is observed to decrease with increasing values of z . Moreover, for all z values, the MKMP-RSCR outperforms MKMP, which is significantly superior to LEDS in terms of confidentiality as presented in [14], [18]. This observed improvement varies according to the z value. In particular, for three values of z (1,2,3) the improvement is 95%, 90% and 85% respectively when 1000 nodes are compromised. This is due to the ability of the adversary to disclose event contents in the case of compromising one of the e endorsement nodes in MKMP and LEDS, whereas in our new scheme the data is disclosed if and only if the entire set of z cell reporters and all e endorsement nodes are captured. On the other hand, the improvement drops to 75%, 57% and 43% when the number of compromised nodes increased to 5000 due to the increase in the probability of compromising the entire set of cell reporters when more nodes are compromised.

D. Security Strength for Data Authenticity

Data authenticity in a particular cell is compromised if the attacker creates a forged report as a result of capturing some

Figure 2: Data confidentiality of LKMP-RSCR and MKMP under random capture attack

sensor nodes. To show the security strength of our scheme, a formula to calculate the probability of all compromised cells in terms of data authenticity was derived in [14]:

$$P_{auth\{e|z\}} = \sum_{j=1}^e \frac{\binom{e}{j} \binom{N-e}{x-j}}{\binom{N}{x}} \sum_{j=1}^z \frac{\binom{z}{j} \binom{N-z}{x-j}}{\binom{N}{x}} \quad (5)$$

The fraction of captured cells increase with the number of captured nodes as shown in Fig. 3, showing that LKMP-RSCR clearly outperforms MKMP in terms of data authenticity for all values of x . To be more precise, an enhancement of (49%, 24%, 12.5%) is gained using our approach with $z = 1, 2, 3$ respectively when half of all nodes are compromised. However, LKMP-RSCR shows an improvement in comparison to LEDS only for higher values of x . Hence, It outperform LEDS when ($x \geq 4000, x \geq 7000, x \geq 10000$) and $z = 1, 2, 3$ respectively. As a result, LKMP-RSCR is superior compared to the other schemes in terms of the fraction of compromised cells caused by RNCA thwarting data authenticity, especially when 50% of the nodes are captured.

IV. COST ANALYSIS

In this section, both the communication and computational cost of LKMP-RSCR are investigated and compared with other schemes. For communication cost, the LKMP-RSCR dissemination model has been compared with models of two-tier data dissemination TTDD and sink-oriented data dissemination SODD [25] in addition to LEDS and MKMP. Also, a brief description of the computation cost is given, which is lower than other schemes due to the assumption of the wide coverage of the BS.

A. Communication cost

A square terrain of size A is assumed, while nodes are uniformly distributed, number of nodes deployed on each

Figure 3: Probability of compromising all cells in terms of data authenticity of LKMP-RSCR, MKMP and LEDS under RNCA

side ($\approx \sqrt{N}$). The size of each packet constituting the generated report, acknowledgement message and bootstrapping correspondences are assumed to be of a fixed size (λ). The cost related to bootstrapping phase is considered for the entire network. Then a cost caused by a single cell report generation in a worst case scenario is assumed.

During the bootstrapping phase, all N nodes send their $\{LIST\}_a$ as described in Algorithm 2. The communication cost of this phase is $N\lambda$ which is similar to that for LEDS and MKMP. On the other hand, the communication cost related to BS acknowledgement is λ in LKMP-RSCR. In contrast, it is $N\lambda$ in LEDS and MKMP.

On the other hand, to generate a report by a particular cell, all n nodes are participating, so that, the communication cost is $n\lambda$. This is similar to the cost of LEDS and MKMP. In contrast, BS acknowledgement communication cost is λ due to single hop scheme followed in LKMP-RSCR. In contrast, the BS acknowledgement cost in LEDS and MKMP is $\sqrt{2N}\lambda$. As a result, communication cost might be expressed as following:

$$CC_{LKMP-RSCR} = \lambda(N + \sqrt{2N} + n + 1) \quad (6)$$

$$CC_{LEDS} = \lambda(2N + 2\sqrt{2N} + n) \quad (7)$$

$$CC_{MKMP} = k\lambda(2N + 2\sqrt{2N} + \frac{n}{k}) \quad (8)$$

Moreover, according to [25]:

$$CC_{SODD} = N\lambda + \frac{4N}{\sqrt{n}}\lambda + kmn\lambda + kc(m\lambda + d)\sqrt{2N} \quad (9)$$

$$CC_{TTDD} = kmN\lambda + kcd\sqrt{N} \quad (10)$$

Where, according to the specifications of systems shown in [18], [25] k is the number of BS, m is the number of cells traversed by each mobile BS, $0 \leq c \leq \sqrt{2}$ and d is the number of data packets received from a particular cell per time. To

Figure 4: Communication cost of LKMP-RSCR, LEDS, MKMP, TODD and SODD, $d=100$ packet, $l=1$ packet

keep comparison consistency between different schemes, a single stationary BS is assumed ($k, m = 1$).

B. Computational cost

In comparison to LEDS and MKMP, the security credential derivation and virtual grid construction is facilitated by the LKMP-RSCR assumption of the BS's wide coverage. This allow the BS to implement the mentioned derivation rather than their implementation by each particular node. In contrast, in the MKMP scheme, the authentication keys shared with the relevant authentication nodes are derived individually by each node. On the other hand, these derivations are implemented by a robot in LEDS which is an impractical assumption, especially in a harsh environment. Hence, less computation is

required to be implemented by each node when employing LKMP-RSCR. This is one of the major advantages of our scheme.

V. OPTIMUM NUMBER OF CELL REPORTERS

In order to guarantee the optimality of the proposed system, the following sections investigate the optimum number of cell reporters.

A. Mathematical Analysis

In this section, the optimum number of cell reporters z will be investigated based on the expressions $P_{C\{e|z\}}$ and $P_{auth\{e|z\}}$ given in (4) and (5) respectively. The calculated value must minimize (P_c) and (P_{auth}). Both expressions comprise two parts: the first part is constant in terms of z while the second part varies. Hence, the set of optimum values of z_{opt} can be expressed as:

$$z_{min} = \arg \min_z \left(\sum_{j=1}^z \frac{\binom{z}{j} \binom{N-z}{x-j}}{\binom{N}{x}} \right) \cap \arg \min_z \left(1 - \frac{\binom{N-z}{x}}{\binom{N}{x}} \right) \quad (11)$$

Obviously, $z = 1$ is the minimum value according to summation limits of the first part. However, the second part of the equation above shows that $z \leq N - x$. As a result, $z_{min} = 1$ and $z_{max} = N - x$ and this leads to

$$P_c|_{z_{min}} = \frac{x}{N} \left(1 - \frac{\binom{N-e}{x}}{\binom{N}{x}} \right) \quad (12)$$

$$P_c|_{z_{max}} = \frac{N! - x!(N-x)!}{N!} \left(1 - \frac{\binom{N-e}{x}}{\binom{N}{x}} \right) \quad (13)$$

$$P_{auth}|_{z_{min}} = \frac{x}{N} \sum_{j=1}^e \frac{\binom{e}{j} \binom{N-e}{x-j}}{\binom{N}{x}} \quad (14)$$

$$P_{auth}|_{z_{max}} = \sum_{j=1}^{N-x} \frac{\binom{N-x}{j} \binom{x}{x-j}}{\binom{N}{x}} \sum_{j=1}^e \frac{\binom{e}{j} \binom{N-e}{x-j}}{\binom{N}{x}} \quad (15)$$

For more clarification on the optimum number of cell reporters, P_c is plotted as a function of z as shown in Fig. 5. The range of z values, which maintain a value of P_c lower than 0.3%, are investigated. It is obvious that P_c significantly increases as the values of z increase. The increment rate is changed depending on the amount of compromised nodes x . When x is equal to one tenth of the total number of nodes inside the network, number of cell reporters can varies between 1 and 4, ($z \in [1, 4]$) while $P_c < 0.3\%$. On the other hand, when the number of compromised nodes is halved, the range of cell reporter numbers is wider, so that z can be selected to be 6 while the value of P_c is kept below 0.3%. The same analysis is shown in Fig. 6 where the value of P_{auth} is proportional to the number of cell reporters. Next the optimality of z in terms of cell capturing and its effect on the integrity of generated reports is investigated.

Figure 5: The relationship between number of cell reporters (z) and the percentage of compromised cells in terms of data confidentiality P_c

Figure 6: The relationship between number of cell reporters (z) and the percentage of compromised cells in terms of data authenticity P_{auth}

B. Optimal value of z in terms of the security of a particular cell

As shown in Fig. 8, P_{z_comp} approaches zero if z lies within a specific range. These values are referred to as Δz , which is varied based on the value of n . This figure proves two important aspects: the first aspect is related to the optimum value of z , which is clearly indicated. The second aspect is related to the relationship between the value of n and Δz . This proves that when n increases, the range of optimum values of z increases, which offers considerable flexibility for the system operator. For instance, when the number of nodes inside a cell is 10, the operator has to choose the number of cell reporters $z \in [3, 7]$ to ensure a very low probability of compromised cells. On the other hand, the operator can choose $z \in [2, 28]$ when $n = 30$ and be sure of the same compromised cell probability ($P \leq 0.01$). To be more precise about the optimal value of z , we investigate the minimum point for each graph in Fig. 7 and find the

optimum value of z :

$$\begin{aligned}
 z_{opt} &= \arg \min_z \left\{ \frac{(z!)(n-z)!}{n!} \right\} \\
 \frac{\partial}{\partial z} (z!)(n-z)! &= 0 \\
 (n-z)! \Gamma(z+1) \psi^0(z+1) \\
 &= z! \Gamma(n-z+1) \psi^0(n-z+1) \\
 \psi^0(z+1) &= \psi^0(n-z+1)
 \end{aligned} \tag{16}$$

While $\psi^m(x)$ is a monotonic function [26]:

$$z = \frac{n}{2} \tag{17}$$

It is clear that (17) represents a critical point for $P_{z_{comp}}$, so the first derivative test is applied. The selected points to implement this test are $(0 \in (-\infty, \frac{n}{2}))$ and $(n \in (\frac{n}{2}, \infty))$. In order to test these two points, the first derivative of $P_{z_{comp}}$ is calculated.

$$\begin{aligned}
 \frac{\partial}{\partial z} (P_{z_{comp}}) &= \frac{\partial}{\partial z} \frac{z!(n-z)!}{n!} \\
 &= \frac{1}{n!} [(n-z)! \Gamma(z+1) \psi^0(z+1) - \\
 &\quad z! \Gamma(n-z+1) \psi^0(n-z+1)]
 \end{aligned} \tag{18}$$

As a result, at $z = 0$

$$\begin{aligned}
 \frac{\partial}{\partial z} (z!)(n-z)!/n!|_{z=0} &= \frac{n! \Gamma(1) \psi^0(1)}{n!} - \frac{\Gamma(n+1) \psi^0(n+1)}{n!} \\
 &= \psi^0(1) - \psi^0(n+1)
 \end{aligned} \tag{19}$$

On the other hand, when $z = n$:

$$\begin{aligned}
 \frac{\partial}{\partial z} (z!)(n-z)!/n!|_{z=n} &= \frac{\Gamma(n+1) \psi^0(n+1)}{n!} - \frac{n! \Gamma(1) \psi^0(1)}{n!} \\
 &= \psi^0(n+1) - \Gamma(1) \psi^0(1)
 \end{aligned} \tag{20}$$

According to the proof shown in Appendix A, the expression of (20) is positive for all $n, x \in \mathbb{R}^+$. Moreover, the expression of (19) is negative for all $n, x \in \mathbb{R}^+$. Hence, the unique minimum point of $(P_{z_{comp}})$ is $z = \frac{n}{2}$. As an example, three different values of n are chosen as 10, 20 and 30. The resultant optimum value of z are obviously 5, 10 and 15 respectively. This is shown in Fig. 7.

VI. CONCLUSION

In this paper, a novel location-dependent key management protocol with a randomly selected cell reporter (LKMP-RSCR) is presented and is shown to achieve an improved performance compared to existing schemes. Each node has its unique credentials derived based on its position, which removes the influence of the capturing of a node on other sensor nodes in its vicinity. Each cell has a particular number of cell reporters z which are randomly chosen by the BS. The involvement of cell reporters in report generation is compulsory, otherwise, the received report at the BS side will be discarded. An extensive analysis was presented to evaluate

Figure 7: The relationship between number of cell reporters (z) and the probability of compromising all cell reporters inside a cell of 10, 20 and 30 nodes

this scheme, which shows a distinct robustness against a significant numbers of captured nodes. In contrast to other schemes, our system shows a considerable improvement in terms of data confidentiality (85%) and data authenticity (35%). Moreover, there is a notable decrease in the computational cost inside each node and communication cost inside the entire network. Finally, the optimum number of cell reporters was extensively investigated related to the security requirements, which was proven to be $z = \frac{n}{2}$.

APPENDIX A PROOF OF $\psi^m(n+1) \geq \psi^m(1)$

For all $n, x \in \mathbb{R}^+$:

$$\begin{aligned}
 x^n &\geq 1 \\
 x^n e^{-x} &\geq e^{-x}
 \end{aligned}$$

According to domination rule for definite integration [27]

$$\begin{aligned}
 \int_0^\infty x^n e^{-x} dx &\geq \int_0^\infty e^{-x} dx \\
 \int_0^\infty x^{(n+1)-1} e^{-x} dx &\geq \int_0^\infty x^0 e^{-x} dx \\
 \int_0^\infty x^{(n+1)-1} e^{-x} dx &\geq \int_0^\infty x^{1-1} e^{-x} dx \\
 \Gamma(n+1) &\geq \Gamma(1) \\
 \ln \Gamma(n+1) &\geq \ln \Gamma(1) \\
 \frac{\partial^m}{\partial z^m} \ln \Gamma(n+1) &\geq \frac{\partial^m}{\partial z^m} \ln \Gamma(1) \\
 \psi^m(n+1) &\geq \psi^m(1) \blacksquare
 \end{aligned} \tag{21}$$

ACKNOWLEDGMENT

The first and the fourth authors would like to thank the Iraqi Ministry of Higher Education and Scientific Research (MoHESR) and the University of Baghdad for financing their PhD study at the Newcastle University. Moreover, the authors would like to thank the anonymous reviewers for their valuable comments and help through this work.

REFERENCES

- [1] W. R. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. ACM, 1999, pp. 174–185.
- [2] D. Christin, A. Reinhardt, P. S. Mogre, and R. Steinmetz, "Wireless sensor networks and the internet of things: selected challenges," *Proceedings of the 8th GI/ITG KuVS Fachgespräch Drahtlose Sensornetze*, pp. 31–34, 2009.
- [3] W. Diffie and M. Hellman, "New directions in cryptography," *The IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, Nov 1976.
- [4] B. C. Neuman and T. Ts'o, "Kerberos: An authentication service for computer networks," *Communications Magazine, IEEE*, vol. 32, no. 9, pp. 33–38, 1994.
- [5] R. Rivest, S. A., and A. L. M., "A method for obtaining digital signature and public key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.
- [6] X. He, M. Niedermeier, and H. De Meer, "Dynamic key management in wireless sensor networks: A survey," *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 611–622, 2013.
- [7] S.-H. Seo, J. Won, S. Sultana, and E. Bertino, "Effective key management in dynamic wireless sensor networks," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 2, pp. 371–383, 2015.
- [8] W. Li, E. Chan, M. Hamdi, S. Lu, and D. Chen, "Communication cost minimization in wireless sensor and actor networks for road surveillance," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 2, pp. 618–631, Feb 2011.
- [9] S. Faye and C. Chaudet, "Characterizing the topology of an urban wireless sensor network for road traffic management," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [10] K. Ren, W. Lou, and Y. Zhang, "Leds: Providing location-aware end-to-end data security in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 585–598, 2008.
- [11] L. Harn, C. Hsu, O. Ruan, and M. Zhang, "Novel design of secure end-to-end routing protocol in wireless sensor networks," *Sensors Journal, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.
- [12] L. Harn and C.-F. Hsu, "Predistribution scheme for establishing group keys in wireless sensor networks," *Sensors Journal, IEEE*, vol. 15, no. 9, pp. 5103–5108, Sept 2015.
- [13] Y. Al-Nidawi and A. Kemp, "Mobility aware framework for timeslotted channel hopping IEEE 802.15.4e sensor networks," *Sensors Journal, IEEE*, vol. 15, no. 12, pp. 7112–7125, Dec 2015.
- [14] H. Fakhrey, S. Boussakta, R. Tiwari, Y. Al-Mathehaji, and A. Bystrov, "Location-dependent key management protocol for a wsn with a random selected cell reporter," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 6300–6305.
- [15] S. Zhu, S. Setia, S. Jajodia, and P. Ning, "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks," in *Security and Privacy, 2004. Proceedings. 2004 IEEE Symposium on*. IEEE, 2004, pp. 259–271.
- [16] F. Ye, H. Luo, S. Lu, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 4, pp. 839–850, 2005.
- [17] H. Yang, F. Ye, Y. Yuan, S. Lu, and W. Arbaugh, "Toward resilient security in wireless sensor networks," in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '05. New York, NY, USA: ACM, 2005, pp. 34–45. [Online]. Available: <http://doi.acm.org/10.1145/1062689.1062696>
- [18] H.-W. Ferng, J. Nurhakim, and S.-J. Horng, "Key management protocol with end-to-end data security and key revocation for a multi-bs wireless sensor network," *Wireless networks*, vol. 20, no. 4, pp. 625–637, 2014.
- [19] M. Wei, R. Aragues, C. Sagues, and G. Calafiore, "Noisy range network localization based on distributed multidimensional scaling," *Sensors Journal, IEEE*, vol. 15, no. 3, pp. 1872–1883, March 2015.
- [20] S. Tomic, M. Beko, and R. Dinis, "Rss-based localization in wireless sensor networks using convex relaxation: Noncooperative and cooperative schemes," *Vehicular Technology, IEEE Transactions on*, vol. 64, no. 5, pp. 2037–2050, May 2015.
- [21] O. Gungor, F. Chen, and C. Koksall, "Secret key generation via localization and mobility," *Vehicular Technology, IEEE Transactions on*, vol. 64, no. 6, pp. 2214–2230, June 2015.
- [22] X. Li, R. Lu, X. Liang, and X. Shen, "Side channel monitoring: Packet drop attack detection in wireless ad hoc networks," in *2011 IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–5.
- [23] F. Ye, G. Zhong, S. Lu, and L. Zhang, "A robust data delivery protocol for large scale sensor networks," in *Information Processing in Sensor Networks*. Springer, 2003, pp. 658–673.
- [24] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.
- [25] F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang, "A two-tier data dissemination model for large-scale wireless sensor networks," in *Proceedings of the 8th annual international conference on Mobile computing and networking*. ACM, 2002, pp. 148–159.
- [26] C.-P. Chen, "Complete monotonicity and logarithmically complete monotonicity properties for the gamma and psi functions," *Journal of Mathematical Analysis and Applications*, vol. 336, no. 2, pp. 812–822, 2007.
- [27] M. Weir, G. Thomas, and J. Hass, *Thomas' Calculus*, ser. Always Learning. Addison-Wesley, 2009. [Online]. Available: <https://books.google.co.uk/books?id=yavIPgAACAAJ>