MAGIK: An Efficient Key Extraction Mechanism Based on Dynamic Geomagnetic Field

Fudong Qiu, Zhengxian He, Linghe Kong, and Fan Wu
Shanghai Key Laboratory of Scalable Computing and Systems
Department of Computer Science and Engineering
Shanghai Jiao Tong University, China
fdqiu@sjtu.edu.cn, zx.he@foxmail.com, linghe.kong@sjtu.edu.cn, fwu@cs.sjtu.edu.cn

***Abstract*—Secret key establishment is a fundamental requirement for private communication between two wireless entities. An intriguing solution is to extract secret keys from the inherent randomness shared between them. Although several works have been done to extract secret keys from different kinds of mediums (*e.g.*, RSSI, CSI, CIR), the efficiency and security problems are not fully solved. In this paper, we consider the problem of secret key establishment for wireless devices, and propose MAGIK, a secure and efficient scheme based on dynamic geoMAGnetic field in Indoor environment for Key establishment. We carefully study the feasibility of utilizing indoor geomagnetic field for key extraction through extensive measurements. Our results demonstrate that geomagnetic field has several dynamic properties, including space-varying, time-varying, sensitive to measurement device, and correlative between two observed points in proximity. We also optimize the key extraction process and present two rotation-angle-based quantification methods, which can achieve faster key generation rates and lower bit mismatching ratios. Besides, we build a prototype on commodity mobile devices, and evaluate its performance by conducting real-word experiments in indoor scenarios. The experiment results confirm that our system is efficient, in terms of key extraction rate, and robust in secret key establishment without requiring additional overhead on mobile devices.**

I. Introduction

Secret key establishment is a fundamental requirement for private communication between two entities, especially in wireless scenarios. Because of the nature of wireless transmissions, it is easy for an eavesdropper to access the exchanged information. As a result, the development of robust techniques for ensuring the security of sensitive information has become an emphasis within the wireless communication research community. Symmetric encryption is the most popular, but it requires secure distribution of a unique secret key between two legitimate entities [1]. An alternate is the public key cryptography, which avoids key distribution by applying a pair of asymmetric keys. But this consumes significant amounts of computing resources, which might not be available in certain scenarios (*e.g.*, mobile devices, sensor networks) [1]. More importantly, concerns about the security of public keys in the future have spawned research on methods that do not use public keys [2].

An intriguing and promising solution to the problem of secret key establishment between wireless nodes is to generate/extract secret keys from the inherent randomness shared between them. Plenty of works have been done to demonstrate that it is possible to securely extract secret key from dynamic characteristics of their common environment [2]–[10]. For example, Hershey first proposed the idea of bit extraction for shared secret key in [3], then different kinds of channel measurement have been explored including angle of arrival [4], phase [5] and Received Signal Strength (RSS) [2] [6]. In addition to these one-dimensional measurements, Chanel State Information (CSI) [8] and Channel Impulse Response (CIR) [9] [10] have also been explored as the sources for shared secret keys extraction.

However, there are several problems unsolved in previous works. The most concern is the efficiency of key generation. For example, not all RSS-based measurements are created equal in terms of the number of bits it is possible to extract. And a wider channel bandwidth has a detrimental effect on the bit extraction rate (*e.g.*, in IEEE 802.11 based devices, the RSS is calculated for a signal over a bandwidth 4 times as wide as IEEE 802.15.4 based devices, so the channel gain is not as affected by narrowband fading). This reduces the number of bits that are possible to extract from RSS. Furthermore, the RSS and CIR, extracting key under a single sampling frequency, only provide coarse-grained channel information. Thus the effectiveness of key extraction is largely limited [11]. The second is the availability, which means whether the key extraction can be implemented in the commodity wireless devices or not. For example, although CSI can provide fine-grained physical layer information due to its multiple subcarriers property from orthogonal frequency-division multiplexing (OFDM), the measurement of CSI needs a special wireless card (Intel WiFi Link 5300), which is not available on commodity mobile devices. The last problem is the security issue, which is always the concern by mobile users. However, the extracted key may be not that secure if the sources (*e.g.*, AP in RSS-based method) are controlled by malicious third party.

In this paper, we consider the problem of secret key establishment for mobile wireless devices by utilizing the inherent randomness of geomagnetic field, and propose MAGIK, a secure and efficient scheme based on the dynamic geoMAGnetic field in Indoor environment for Key establishment. We first study the feasibility of utilizing indoor geomagnetic field for key extraction through extensive measurements in real-world scenarios. Fortunately, our observed dynamic properties, including space-varying, time-varying, sensitive to measurement device and correlative between two observed points in proximity, are beneficial for us to achieve a secure and efficient key extraction. Though time-varying, geomagnetic field is still in the extremely low-frequency (ELF) range (3-3000 Hz) [12],

This work was supported in part by the State Key Development Program for Basic Research of China (973 project 2014CB340303), in part by China NSF grant 61672348, 61672353, 61422208, 61472252, 61272443 and 61133006, in part by Shanghai Science and Technology fund 15220721300, in part by CCF-Tencent Open Fund, and in part by the Scientific Research Foundation for the Returned Overseas Chinese Scholars. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government.
Fan Wu is the corresponding author.

which means directly quantifying the magnitude of the sample may have a low key generation rate. So we optimize the key extraction process and present two rotation-angle-based quantification methods, which can achieve faster key generation rates and lower bit mismatching ratios. Besides, we also build a prototype on commodity mobile devices and evaluate it by conducting real-word experiments in indoor scenarios. The experimental results confirm that our system is efficient, in terms of key extraction rate, and robust in secret key establishment without additional overhead on mobile devices.

We summarize our main contributions as follows.

- We carefully study the feasibility of utilizing indoor geomagnetic field for key extraction through extensive measurements and observations in real-world scenarios. On one hand, the results demonstrate the correlation of geomagnetic field between two mobile devices in proximity, which makes the geomagnetism a possible candidate for key extraction. On the other hand, the observation shows that the indoor geomagnetic filed is space-varying, time-varying and also sensitive to the measurement device. These properties enhance the security of key extraction.
- We propose MAGIK. MAGIK can extract the identical secure key between two mobile devices under the presence of the malicious eavesdropper, and achieve higher authentication accuracy compared with existing channel based methods (*e.g.*, RSS-based).
- Instead of quantizing the magnitude of geomagnetic signal, we present a new quantification method targeting the time infomation, which corresponds to the largest rotational angle of mobile devices. We also propose an optimized quantification technique that can achieve both faster key extraction rates and lower mismatching ratios.
- We implement MAGIK directly on commodity mobile devices. More specifically, we build our system on Nexus 7 and validate its performance by conducting real-word experiments. Experiments results confirm that our system is efficient in key extraction rate and robust in key establishment without additional overhead on mobile devices.

The remainder of this paper is organized as follows. In section II, we introduce our system model and attack model. In section III, we claim some technical preliminaries and study the feasibility of using geomagnetic field to extract identical secret key between two mobile devices. Then, we describe our design details and security analysis in section IV and section V, respectively. In section VI, we implement MAGIK and present our evaluation results. In section VII, we briefly review the related work. Finally, we conclude our paper in section VIII.

II. System Model and Attack Model

In this section, we introduce the system model of our proposed geomagnetism-based key extraction scheme. Then we present the attack model considered in this work.

A. System Model

As shown in Figure 1, we assume that Alice and Bob are two legitimate wireless devices located nearby to each other, but they do not have any priori knowledge. A secure and authentic channel is required between them to prevent the adversary, Eve, from eavesdropping their communication. To set up the secure channel $C2$, a cryptographic key is needed to encrypt the channel. Alice and Bob extract the secret key via their measurement results for geomagnetic field (GMF) around them. We also assume that all communications between Alice and Bob are transmitted through public wireless channel $C1$ before the cryptographic key generated.

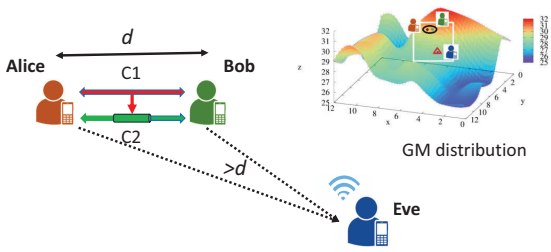


Fig. 1. System Model. The dynamic magnetic field can be utilized by two parties, Alice and Bob in physical proximity, to extract a cryptographic key through a public wireless channel $C1$. Then, they can communicate through a secure channel $C2$ and all the transmission can be encrypted by the extracted key. An adversary, Eve, cannot extract the same key since she is out of the proximity of Alice and Bob.

More specifically, Alice and Bob are close enough (distance is denoted as d) so that their measured GMFs are similar, even identical. Two devices should be moved together while they take GMF samples at a fixed frequency f. The sample data are denoted as $\mathbf{S}_u = \{\mathbf{s}_u^1, \mathbf{s}_u^2, \cdots, \mathbf{s}_u^i, \cdots, \mathbf{s}_u^p\}^T$, where $u =$ Alice/Bob/Eve represents different users and $\mathbf{s}_u^i = \{x_u^i, y_u^i, z_u^i\}$ represents the projection of GMF strength on three axis. Additionally, the exact sampling time point when the sample is taken will also be recorded as a time stamp of the vector.

There is an adversary Eve in our system, who tries to get the same secret key as Alice/Bob by eavesdropping or reproducing. We will talk about this in our attack model. Our goal is to defend against Eve while achieve secret key extraction with faster key generation rates and lower mismatching ratios.

B. Attack Model

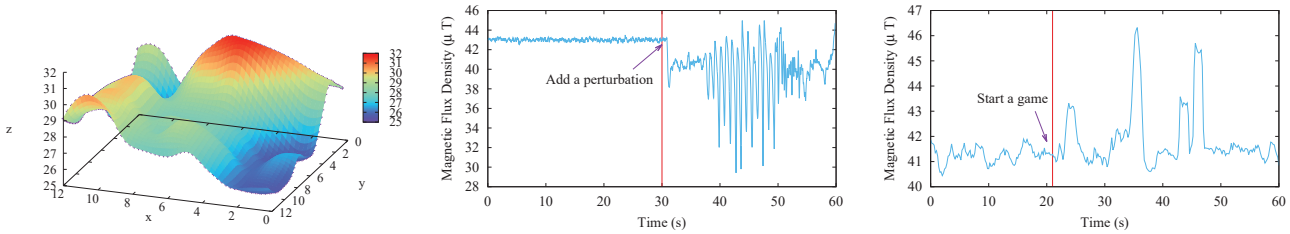
We assume that an adversary, Eve, exists all the time and tries to get the correct secret key between Alice and Bob. We mainly consider following three kinds of attacks in this work.

Eavesdropping: We first assume that Eve is able to overhear all the communication between Alice and Bob. Eve tries to get the correct secret key or any helpful information by analyzing the transcript between Alice and Bob.

Reproducing-simultaneously: We then consider that Eve tries to sample the GMF data and reproduce the secret key by herself while Alice and Bob are extracting their secret key. We assume that the key extraction algorithms and corresponding parameters' values are public. That is to say, Eve can reproduce the secret key correctly if she gets the same or similar measurement results (*i.e.*, GMF) as that of Alice/Bob. The main limitation for Eve is that she cannot be too close to either Alice or Bob while they are collecting data.

Reproducing-afterwards: We also consider the threat that Eve tries to sample after Alice and Bob leaving. Although we assume that Eve cannot get closer when Alice and Bob are sampling, she can sample anywhere after Alice/Bob leaving.

Besides, we think, essentially, Eve is more interested in the extracted secret key between Alice and Bob, but not disrupting their key establishment process. Therefore, we also assume that Eve can neither jam the communication channel between



(a) Space-varying distribution of GMF in indoor environment.

(b) Time-varying GMF due to the Influence introduced by perturbation source.

(c) Influence on GMF introduced by measurement device itself.

Fig. 2. Space-varying, Time-varying, and Sensitiveness to the measurement itself of geomagnetic field.

Alice and Bob nor modify the information exchanged between them.

III. Feasibility

In this section, we provide some preliminaries of geomagnetic field, and then discuss the feasibility of using geomagnetism for key extraction.

A. Geomagnetic Field Preliminaries

Our key extraction framework exploits the geomagnetic field (GMF), also known as the Earth's magnetic field. GMF is a kind of magnetic field which is always around us and has been extensively studied by researchers in astrophysics [13]. With the advancements in silicon chip fabrication, GMF can now be measured using low cost, low power and very small sensors commonly known as magnetometers [14]. Therefore, we can easily measure the GMF around us with the help of magnetometers embedded in commodity mobile devices.

We have observed that the GMF in indoor environment has a great space variation, which makes it an appropriate candidate to extract secret key. It is very different from the GMF in outdoor environment, where the magnitude and direction of GMF are always stable or changing gradually. Since the GMF can be regarded as the field generated by magnetic dipoles, the indoor GMF can be viewed as the composition of multiple magnetic dipoles sources. There are many sources, named *Magnetic Perturbation*, that can affect the distribution of GMF in indoor environment, such as electromagnetic devices or magnetization of manmade structures [15]. We can simply model the indoor GMF as

$$\mathbf{B} = \mathbf{B}_e(\mathbf{R}) + \mathbf{B}_p(\mathbf{d}), \tag{1}$$

where $\mathbf{B}_e(\mathbf{R})$ denotes the magnetic flux density of Earth's dipole, $\mathbf{B}_p(\mathbf{d})$ denotes the magnetic flux density of additional magnetic perturbation, $\mathbf{R}$ is the distance vector between the Earth's dipole and the observation point, and $\mathbf{d}$ is the distance vector between observation point and the perturbation.

B. Feasibility Study

In this subsection, we demonstrate the feasibility of using indoor GMF to extract secret key between two entities in proximity. We first show that the distribution of indoor GMF is *Space-varying*, *Time-varying*, and *Sensitive to measurement device* according our measurement results in real-world scenarios, and then display the correlation of GMF between two entities when they are in physical proximity.

1) Space-varying GMF: We can assume that any movement in our experiment has no influence on the measured magnetic field $\mathbf{B}_e(\mathbf{R})$ due to $|\mathbf{z}| \ll |\mathbf{R}|$, where $\mathbf{z}$ is the displacement of the new observation point. Therefore, we can assume that the former part in Equation (1) is a constant. The magnitude and the direction of the latter part are depended on the relative position between observation point and magnetic perturbation source. In other words, the magnetic field vector $\mathbf{B}$ of an observation point is varying with the change of distance between observation point and magnetic perturbation source. This property demonstrates the *Space-varying* distribution of GMF in indoor environment.

Apart from the analysis from theoretical perspective, we also prove the space-varying distribution of GMF through practical measurements in our working building. We measured the GMF distribution of a $12 \times 12m^2$ lab in 3rd floor, where many perturbation sources exists, using our self-developed measurement tool running on Nexus 7. As shown in Figure 2(a), the distribution of GMF in indoor environment is extremely irregular and space-varying.

2) Time-varying GMF: According to our measurements and observations, we find that the distribution of GMF is also time-varying in complex indoor environments. For example, the data shown in Figure 2(b) is the measurement result of a fixed observation point during 60 seconds. We opened a perturbation source (a laptop) near the observation point at the 30^{th} second. We can see that the corresponding magnetic flux density changed irregularly after we opened the laptop. Therefore, we can say that the irregular working status of perturbations will lead to time-varying distribution of GMF. We also observed that the moving objects can also contribute to the GMF variation. Here, we note that the adversary Eve can neither fully control the working statuses of all the perturbations nor the moving trajectories of all objects/people. This assumption makes sense because there are many examples like this in practical scenarios, such as the refrigerator works irregularly because it is controlled by embedded temperature sensor, the computer works irregularly because it is controlled by its owner, and the web server works irregularly because its workload is decided by remote requests, *etc.*

3) Sensitive to Measurement Device: Our another observation is that the magnetic field introduced by measurement device itself can also influence the measurement result of observing point. We measure the magnetic field using Nexus 7 in two scenarios: Nexus 7 works with light workload (CPU up to 20%) and with heavy workload (CPU up to 90%), respectively. The results are shown in Figure 2(c), where we run a large game on Nexus 7 at the 20^{th} second. This observation tells us

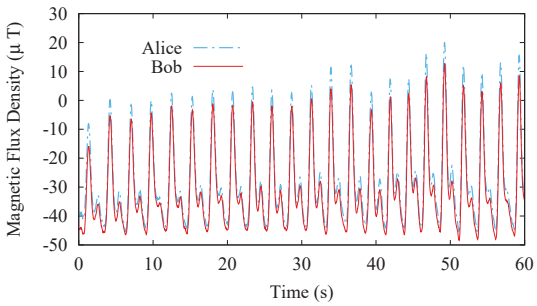


Fig. 3. 2D Correlation of two entities (Alice and Bob) in physical proximity. This picture shows the measurement results of Y axis.

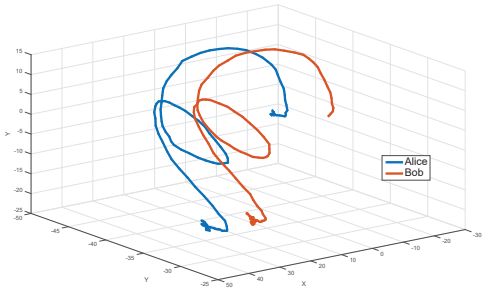


Fig. 4. 3D Correlation of two entities (Alice and Bob) in physical proximity. The results of this figure were collected when two devices moved together in the air along a trajectory "3".

that the magnetic field is sensitive to the measurement device and we can defend against the *reproducing-afterwards* threat by making full use of this property.

4) Correlation in Proximity: The above properties (*Space-varying*, *Time-varying*, and *Sensitive to Measurement Device*) are advantageous for the security of MAGIK. Here, we will show the correlation of GMF between two mobile devices when they are in physical proximity. We give two measurement results, shown as Figure 3 and 4. Figure 3 focuses on the correlation results of two entities in proximity from one-dimensional perspective, and we only show the result on Y-axis because the results on other two axises are similar with Y-axis. Figure 4 focuses on the correlation results from 3-dimensional perspective. We note that the results of this figure were collected when two devices moved together in the air along a trajectory "3". This kind of trajectory experiment is unrepeatable because we can not repeat the actual identical trajectories every time. From the above two figures, we can see that two entities can achieve close correlation if they are placed in physical proximity, and this kind of correlation is exactly what we need to extract identical secret key from GMF.

In summary, the properties of space-varying, time-varying, sensitive to measurement device and close correlation in proximity confirm our idea that it is feasible to adopt indoor geomagnetic field for key extraction.

IV. System Design

In this section, we present our basic idea and challenges of the proposed geomagnetic-based key extraction scheme and introduce the design details step by step.

A. Basic Idea and Challenges

We design MAGIK to translate the geomagnetic field measurement into secret key that is only shared between Alice and Bob in proximity. Finding information in common contained in each of their GMF measurements is the basic requirement in key extraction. Since adversary Eve can overhear all the messages they exchanged, leaking as little information as possible in their communication is rather important. In the following, we present our basic ideas, the challenges we met in the design and corresponding solutions at a high level.

First, we try to make the measurement results as similar as possible. As we mentioned in system model, Alice and Bob sample the GMF data at a fixed frequency simultaneously while they are moving together. However, we find out that it is hard to get similar resutls if we only measure the magnitude of GMF. The reasons may be that the magnetometer has high sensitivity, and the different magnetometers on different devices are hard to get exactly the same measurement result. Instead of using magnitude values, GMF's another parameter, *direction*, which indicates the magnetic field direction of observation point, seems to be a better choice. However, using *direction* may arouse another problem, that is, although we record each sample's orientation information, this is just the direction relative to the device itself. The relative position of Alice and Bob are not clear, so the samples' directions on two devices cannot be compared together. Fortunately, we find that although the coordinates of two devices may not be synchronized, the changes of their directions in a specific time period are highly identical, if the two devices move along exactly the same trajectory. Therefore, we decide to use the rotation angles of mobile device in a time period as the source of key extraction.

As it seems that the angles will not be that identical if we choose two pair of exact samples to calculate their angles, we seek the two corresponding time stamps when the most significant angle is generated during that period. It is highly possible that the biggest angle appears at a similar time period on both devices. However, there being only one biggest angle is far not enough to generate bits efficiently. We need to seek more *biggest* angles. This can be achieved by seeking the biggest angles of different time intervals.

We then try to exchange as few messages as possible while negotiating a shared secret key. After quantification and encoding, each of Alice and Bob has an n-bit sequence, which may differ at any position. Directly transmitting is clearly impractical. How will Alice tell Bob which bits in her sequence are different from him? We encode the sequences into golay code and send the checking message, which can not only correct the mismatching bits but also protect the original codes. The detailed realization is described in section IV-B4.

B. Design Description

We will illustrate the design details in this subsection, but focus on Quantification and Reconciliation.

1) Initialization: We synchronize the start time of two devices. Before sampling, Alice sends a signal to notify Bob that it should start sampling at an exact same time as Alice at t_A. On receiving the signal, Bob replies an acknowledge message at time t_B and starts sampling. There is a time difference between t_A and t_B because of the transmission and propagation delay. The *Round Trip Time*, denoted as RTT can be detected by Alice. RTT is used to estimate the time delay

caused by transmission and propagation. Alice then updates her start point $t_A + RTT/2 \approx t_B$. Here we do not require the very precise synchronization, since MAGIK does not simply quantize the time from the start point and is able to eliminate some minor error.

2) Sampling: As we have mentioned in system model, each device samples the GMF at a fixed frequency f. The chosen of f will be explored in evaluation part. Then the samples are stored into a series

$$\mathbf{S}_u = \{\mathbf{s}_u^1, \mathbf{s}_u^2, \cdots, \mathbf{s}_u^i, \cdots, \mathbf{s}_u^p\}^T, \tag{2}$$

where u = Alice/Bob/Eve. Each of the GMF measurement vector $\mathbf{s}_u^i$ is represented by three values, $\{x_u^i, y_u^i, z_u^i\}$, which are the projection of the GMF strength on x/y/z axis based on device's internal coordinate and indicate the direction of the field. The time stamp when the GMF is measured is also recorded in a time series

$$\mathbf{T}_u = \{t_u^1, t_u^2, \cdots, t_u^p\}. \tag{3}$$

We note that the vectors are all relative values to their own coordinate of each device. Fortunately, we have the starting point synchronized and $\mathbf{T}_u$ recorded, which helps us locate the corresponding samples from different devices.

3) Quantification: This is the process of translating the GMF measurements $\mathbf{S}_A$ and $\mathbf{S}_B$ into corresponding bits stream. At the beginning, we describe our design of new quantification algorithm to obtain bits from a given correlated quantity at Alice and Bob. Then an optimized quantification algorithm for better performance is presented.

Quantification on Angles: As we have mentioned in section IV-A, the measurement results will be different since the magnetometers on different devices cannot be exactly the same. Fortunately, the GMF measurements contain the orientation information of the geomagnetic field, which can be leveraged in our new designed quantification methods.

Our main idea is utilizing the rotation angles between measurements to achieve efficient quantification. Considering the fact that when the device moves in the air, the GMF it samples will change correspondingly. With the reference of the device itself, the field vector is spanning around it. Since the two devices move together, the spanning of their measured vectors are highly possible to be identical. Figure 4 shows the spanning. Therefore, the angles twisted of the field vector in a certain time period on each device are very likely to be identical. Although the exact magnitudes of the angles may not necessarily be the same or similar, the time stamps are highly alike. So we can find some representative angle of the two measurements on each device and pick up the time stamp as our quantification source.

The algorithm is shown in Algorithm 1. Given a sequence of GMF vector measurement $\mathbf{S}$ and the corresponding time series $\mathbf{T}$, the algorithm will return a time sequence $\mathbf{Q}$. Every iteration is to find the biggest angle between two measurements for a fixed interval. For example, in one iteration, the interval is τ, then we should look for the biggest angle in

$$\{\langle \mathbf{s}_A^1, \mathbf{s}_A^{1+\tau}\rangle, \langle \mathbf{s}_A^2, \mathbf{s}_A^{2+\tau}\rangle, \cdots \langle \mathbf{s}_A^{n-\tau}, \mathbf{s}_A^n\rangle\}, \tag{4}$$

where $\langle \alpha, \beta \rangle$ denotes the angle between vectors α and β. In each iteration, we will find a biggest angle $\langle \mathbf{s}_u^i, \mathbf{s}_u^j \rangle$, and we then put the time stamp pair (t_u^i, t_u^j) into the quantizing sequence. Changing the interval for each iteration, we can finally get the quantized sequence

$$\{(t_u^{p_1}, t_u^{q_1}), (t_u^{p_2}, t_u^{q_2}), \cdots\} (q_i, p_i \in \mathbb{N}). \tag{5}$$

Algorithm 1 Quantification on Angles

Input: Array of GMF vector measurement $\mathbf{S} = \{\mathbf{s}^1, \mathbf{s}^2, \cdots, \mathbf{s}^n\}$, time series $\mathbf{T} = \{t^1, t^2, \cdots, t^n\}$, total length n.
Output: Time sequence $\mathbf{Q}$.
1: **for** $k = 8 : 4 : n$ **do**
2: For all angle $\langle \mathbf{s}^i, \mathbf{s}^{i+k} \rangle$, find the biggest $\langle \mathbf{s}^p, \mathbf{s}^q \rangle$;
3: Make t^p, t^q as a tuple (t^p, t^q) and add (t^p, t^q) into $\mathbf{Q}$;
4: **end for**
5: **return** $\mathbf{Q}$;

Algorithm 2 Advanced Quantification on Angles

Input: Array of GMF vector measurement $\mathbf{S} = \{\mathbf{s}^1, \mathbf{s}^2, \cdots, \mathbf{s}^n\}$, time series $\mathbf{T} = \{t^1, t^2, \cdots, t^n\}$, total length n, threshold δ.
Output: Time sequence $\mathbf{Q}$.
1: **for** $k = 8 : 4 : n$ **do**
2: For all angle $\langle \mathbf{s}^i, \mathbf{s}^{i+k} \rangle$, find the biggest $\langle \mathbf{s}^p, \mathbf{s}^q \rangle$ and the second biggest $\langle \mathbf{s}^{p'}, \mathbf{s}^{q'} \rangle$;
3: **if** $|\langle \mathbf{s}^p, \mathbf{s}^q \rangle - \langle \mathbf{s}^{p'}, \mathbf{s}^{q'} \rangle| \geq \delta$ **then**
4: Make t^p, t^q as a tuple (t^p, t^q) and add (t^p, t^q) into $\mathbf{Q}$;
5: **end if**
6: **end for**
7: **return** $\mathbf{Q}$;

Advanced Quantification on Angles: The algorithm, running separately on Alice and Bob, will get time sequences $\mathbf{Q}_A$ and $\mathbf{Q}_B$, which are assumed to be identical. However, they may differ at any pair with a probability (ϵ) in reality. In order to get a better performance, we present an advanced algorithm to reduce the probability. First, we explore the source of the difference. In reality, the biggest angle may be slightly larger than the second biggest angle. Therefore, with some disturbance or other measure error, the second biggest angle may be measured to be bigger than the biggest one. This situation will occur often when the first and second biggest angles are very close. In Algorithm 2, we record the first and the second biggest angles and calculate their difference. If the difference is under a threshold δ, we keep the time stamp of the biggest angle. Otherwise, we just skip into the next iteration.

Now we have two time sequence $\mathbf{Q}_A$ and $\mathbf{Q}_B$, with a probability ϵ to be different on any time pair. Then we can translate the every integer time to its corresponding gray code [16]. Since gray code has the special characteristic that the bits only differ at one position in two adjacent code. According to the range of the time and the need of data reconciliation, we translate the time into 16 bit gray code then combine all the code into bit streams C_A and C_B.

4) Data Reconciliation: Previously, we have encoded time sequence into gray code C_A and C_B. Here we present the process of data reconciliation which means to correct or delete the different bits between C_A and C_B so that Alice and Bob can get an identical secret key K. How can Alice and Bob do reconciliation without actually sending C_A or C_B? As mentioned in subsection IV-A, one solution is to treat both C_A and C_B as distorted versions of 'some' n-bit codeword of an (n, k) error correcting code $\mathcal{C}$. A codeword is an n-bit string and a set of codeword that are very similar to each other may be encoded into one k-bit strings by a many-to-one encoding

function $E()$. The one-to-one decoding function $D()$ will map the k-bit string to one of the n-bit codeword. Since C_A and C_B are highly alike, they correspond to the same k-bit strings at a great possibility.

Here we adopt the binary Golay code [17] to achieve data reconciliation. Binary Golay code is a type of linear error-correcting code used in digital communications. The *extended binary Golay code*, G_{24}, encodes 12 bits of data in a 24-bit codeword in such a way that any 3-bits errors can be corrected or any 7-bit errors can be detected. The 24-bit codeword has two parts. The left 12-bit part is the original message and the right 12-bit part is the correcting bits. When receiver gets the codeword, it can decode the original message from the codeword. Once getting the codeword, any 3-bit error occurs on this codeword will not affect the message after decoding. That is to say, assuming that two messages, M_1 and M_2, have 3-bit error and encoded word $W_1 = E(M_1) = [M_1, H_1]$, where H_1 is the checking part, if we let $\overline{W}_2 = [M_2, H_1]$, the message decoded from $\overline{W}_2$ will also be M_1.

Alice and Bob first process the gray code locally. Taking Alice as example, encode every 12 bits of its binary gray code into $W_A = E(C_A) = [C_A, H_A]$. Then calculate the difference $P = C_A - H_A$. P is what we need to send from Alice to Bob. When Bob receives P, it adds P to C_B to get a codeword $\overline{W}_B = [C_B, C_B - P]$. Then Bob decode $\overline{W}_B$ to get $\overline{C}_A$, which is suppose to be the same as C_A. Checking the difference between $\overline{C}_A$ and C_B, if the difference is less than 3 bits, which means Bob correctly got Alice's code, Bob replaces C_B with $\overline{C}_A$. If the difference is more than 3 bits, drop this 12 bits and inform Alice to drop, too. After transmitting every 12-bit checking bits, Alice and Bob modify their codes to be the same. Now, Alice and Bob has the same key.

5) Privacy Amplification: Since the offset P is sent by Alice in plaintext, Eve obtains *partial information* about Alice's and Bob's shared key K. Although it is still hard for Eve to guess K from P alone, it knows something about K. Therefore, in order to make sure that Eve cannot even know *partially* about the key, Alice and Bob reduce the size of their bit-string by $n - k$ bits to obtain k-bit strings. We simply use the k-bit pre-image, *i.e.*, the left 12-bit part of the n-bit codeword, which Alice and Bob process as the generated key.

V. Analysis

We will analyze the security of our proposed MAGIK in this section based on the attacks mentioned in attack model.

A. Preventing Reproducing-simultaneously Attack

MAGIK prevents the adversary Eve from reproducing the identical secret key simultaneously, under the only assumption that Eve cannot be too close to either Alice and Bob while they are sampling data. As we mentioned in section III, the distribution of geomagnetic field in indoor environments is space-varying. Two reasons are provided here. One is that there are many magnetic perturbations, such as electromagnetic devices and magnetization of manmade structures, which can affect the distribution of GMF in indoor environments [15]. And another reason is that the magnetic field of a observation point is varying with the change of distance between observation point and magnetic perturbation source. The measurement results in real-world scenario, as shown in Figure 2(a), also demonstrates the space-varying property of GMF in indoor environments. This space-varying property guarantees that the sampling results of two entities (*e.g.*, Alice and Eve) at different observing points are different.

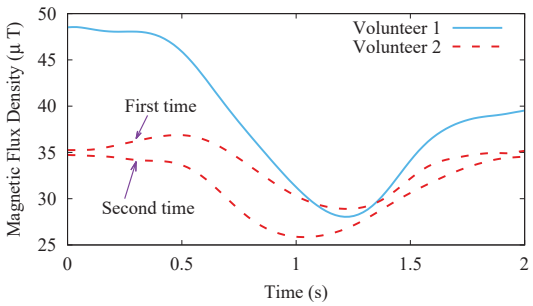


Fig. 5. Unrepeatable trajectory illustration. Same person can not get the same secret key in different time and different persons can not get the same secret key at the same place.

Another property, which is beneficial for MAGIK to prevent Eve from extracting the identical secret key, is that GMF is sensitive to the measurement device at observing point. As shown in Figure 2(c), we measure the magnetic field using Nexus 7: the device works regularly at the first 20 seconds with light workload (CPU utilization up to 20%) and a large game is turned on leading to a heavy workload (CPU up to 90%) at the 20^{th} second. From the figure, the GMF is seriously affected by measurement device's working status. Therefore, we can say GMF is sensitive to measurement device and it is closely related to the device's working status. That is to say, Eve can not collect the exact same samples as Alice/Bob because of their different measurement devices or different working statuses. This can also be used to resist the reproducing-afterwards attack that we will discuss next.

The properties of space-varying and sensitive to measurement device can guarantee that adversary Eve can not reproduce the identical secret if she is not too close to either Alice and Bob while they are sampling data.

B. Preventing Reproducing-afterwards Attack

MAGIK prevents the adversary Eve from reproducing the same secret key as Alice/Bob after they leaving. The properties of time-varying and sensitive to measurement device can help us to achieve this goal. We has claimed that the distribution of GMF is time-varying in complex indoor environments in section III. The observation in Figure 2(b) shows us the magnetic flux density will be changed irregularly after we open the laptop (as a perturbation source) during a period of time. Therefore, we can say that the irregular working status of perturbations will lead to time-varying distribution of GMF. Besides, the moving objects can also contribute to the GMF variation with time. Here we note that the adversary Eve can neither fully control the working statuses of all the perturbations nor the movements of all objects or people.

Though the geomagnetic field is time-varying, it is in the extremely low-frequency (ELF) range. Thus, Eve is still possible to measure the similar field after Alice and Bob leaving. To make our scheme more robust against this kind of attack, in our design, we require Alice and Bob to move together along a special trajectory, because the moving trajectory is not easy to repeat, which means Eve can not get exact identical moving trajectory as Alice/Bob. As shown in Figure 5, the same person

can not get the same secret key in different time, and different persons can not get the same secret key at the same place.

In summary, the properties of time-varying, sensitive to measurement device and hard to copy trajectory help us achieve the goal that the adversary Eve can not reproduce the same secret key as Alice/Bob after they leaving.

C. Preventing Eavesdropping

MAGIK can prevent the adversary, Eve, getting the secret key extracted between Alice and Bob by eavesdropping their communications. We analyze all the possible communications between Alice and Bob in the following contents to verify if there is any leakage about the secret key between Alice and Bob. The first interaction between them happens in initialization phase, where Alice sends a control signal to notify Bob to start sampling at an exact same time and Bob replies an acknowledge message once receiving the signal. There is no information about samples or keys disclosed in this phase. The second interaction happens in data reconciliation phase as described in section IV-B4. In that section, Alice calculates the difference P, and sends P to Bob. We can see that there is no information about secret key is transferred between Alice and Bob, except the difference P. Although the difference P only denotes the partial information about the secret key and it is hard for Eve to guess secret key from P, we utilize privacy amplification technique (IV-B5) to make sure that Eve can not even know partially about the secret key.

Therefore, we can say that MAGIK can prevent the adversary Eve eavesdropping the secret key by the well designed data reconciliation and privacy amplification schemes.

VI. Implementation and Evaluation

We have implemented MAGIK and evaluated its performance in practice.

A. Implementation

We have implemented an Android-based prototype and deployed it on mobile device (*e.g.*, ASUS Google Nexus 7 tablet in our prototype). In each device, magnetometer sensor is provided and managed by a *sensor manager*. We call some Android APIs to complete the measurement. We set up the magnetometer by setting default sensor as `Sensor.TYPE_MAGNETIC_FIELD`. And on a normal Android device, the frequency can be only set as `SENSOR_DELAY_FASTEST`, `SENSOR_DELAY_GAME`, `SENSOR_DELAY_UI` or `DELAY_NORMAL`, which correspond to the sampling frequency 50 Hz, 50 Hz, 16.7 Hz and 5 Hz, respectively. `SENSOR_DELAY_FASTEST` means the fastest frequency, while on our devices, it is the same as `SENSOR_DELAY_GAME` model. After setting the sampling frequency, the magnetometer sends a `SensorEvent`, which includes measured data, to main process periodically. The data is a vector that contains three values representing the geomagnetic field at three directions. Since our quantification algorithms do not require very precise start time on two devices, Alice can start sampling at the same time when she sends a start signal to Bob, and do not need to wait for the response from Bob. After sampling, the function `find_max_angle()`, which corresponds to the algorithm we present in subsection IV-B3, is called automatically and output an array of time pairs. Then the two devices will encode the time pairs into gray code and encode them by the golay code encoding function $E()$. Afterwards, reconciliation messages are exchanged between the two parties and, in the end, a shared key is generated.

B. Setup and Metrics

Our experimental devices are three ASUS Google Nexus 7 tablets, act as Alice, Bob, and Eve, respectively. Each tablet runs *Android 4.3* and is equipped with *1.5GHz quad-core Snapdragon S4 Pro Processor*, *2GB of RAM* and *magnetometer sensor*. We conduct our experiments in a $12 \times 12m^2$ room on the 3rd floor of our working building. There are varieties of electronic devices (*e.g.*, computers, servers, an air conditioner, a microwaver, a refrigerator, *etc.*) in the room. The distance between Alice and Bob is controlled by two flexible brackets. After initialization, we move the two devices together to measure the GMF while keeping a certain distance between them. We change the distance and the sampling frequency in different experiments to explore their influence on the performance. A time period T is set manually to make sure adequate bits are generated if two devices are close enough. After that time T, two devices stop to do the data processing for key extraction. Different methods (quantification algorithms, encoding methods) and different parameters (threshold δ) are explored in our experiments.

In this work, we employ following metrics to analyze the performance of our system.

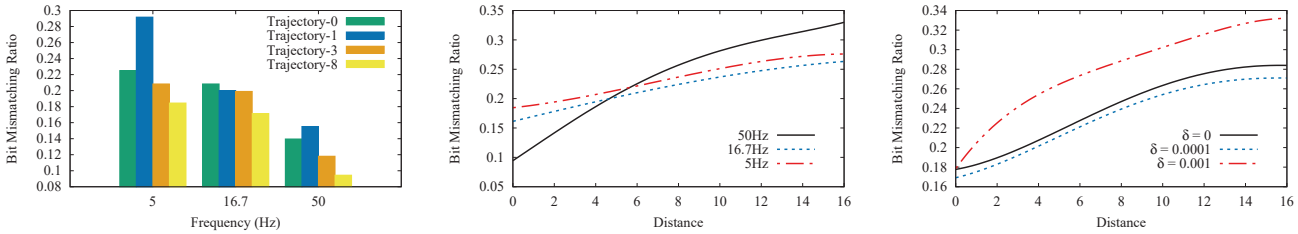
Key Bit Rate (KBR): KBR is the number of bits that are extracted per unit of time. It can indicate both the key extraction time and whether the two parties are located close enough. KBR depends on the distance between two devices, the quantification and coding method, the moving trajectory of two devices, and the sampling frequency.

Bits Mismatching Ratio (BMR): BMR is the number of bits mismatched divided by the total number of bits during a time period. It can help us judge the quality of the sampling, quantification and encoding. It is of high likelihood that two devices will generate exactly the same key if BMR is small enough.

C. Performance

We evaluate MAGIK's performance under different experimental settings, and display our evaluation results from two perspectives, BMR and KBR.

1) Bits mismatching ratio: We first consider the influence on BMR of sampling frequency and moving trajectory. We explored different sampling frequencies (5 Hz, 16.7 Hz and 50 Hz) and different moving trajectories (Holding the devices to write "0", "1", "3", "8" in the air) to extract secret keys. In these experiments, Alice and Bob overlap each other, and we can consider their distance as 0cm. And the sampling frequency is 50 Hz, and the threshold is 0.0001. The results are shown in Figure 6(a). We can figure out that, both the sampling frequency and the complexity of moving trajectory have influenced the BMR. More specifically, the BMR decreases with the increasing of sampling frequency, considering any moving trajectories. The reason is that a higher sampling frequency will collect more samples in a certain time and more samples may lead to a higher possibility to find more precise results, which can achieve a lower BMR. From another perspective, when the sampling frequency is fixed (*e.g.*, 50Hz), the biggest BMR is generated when the moving trajectory is "1", while the smallest by trajectory "8". This indicates that



(a) BMR under different sampling frequencies and different moving trajectories.

(b) BMR under different distances and different frequencies.

(c) BMR under different distances and different thresholds in advanced quantification algorithm.

Fig. 6. Bit Mismatching Ratio

the more complex the trajectory is, the lower the BMR will be. Therefore, in all the following experiments, we use the moving trajectory "8".

Then we change the distance from 0 to 16 cm on different sampling frequencies to evaluate their influences on BMR. As shown in Figure 6(b), it is intuitive that the BMR increases with the growing of distance between two devices. This is easy to understand, because the GMF is space-varying. Another observation is that the BMR increases as the distance increases, and more importantly, the rate of increase under low sampling frequency, such as 5 Hz and 16.7 Hz, is slower than that under high frequency (*i.e.*, 50 Hz). This is because the sampling under a low frequency is less sensitive to the subtle difference of the GMF, which will sometimes miss an important peak and lead to bit mismatch. However, this situation changes with the increasing of distance between two devices. The reason is that, the GMF around two devices become more and more different with the distance increasing, and the sampling under high frequency can detect more different peaks while the sampling with low frequency can not detect that much differences. Therefore, the BMR under 5 Hz and 16.7 Hz is growing slower than that under 50 Hz. This characteristic is essential to our system, since it makes sure the remote Eve would not generate enough bits in the key extraction.

In our quantification algorithm, the value of threshold δ is an important parameter and can directly influence the BMR. At first, we set the threshold $\delta = 0.0001$. Consequently, the BMR drops a little by about 0.02 as shown in Figure 6(c). We then increase the threshold to 0.001 intending to gain more reduction. However, on the contrary, the BMR increases 0.04 compared with the threshold $\delta = 0.0001$. Why the BMR increases while the threshold getting larger? We can categorize the causes of bit mismatching into two parts. The first part is the real difference, that is, the biggest angles of two devices do not come from the same position due to a difference in the GMF around two devices. The second part comes from the measurement errors. A proper threshold can reduce the second part but has no influence on the first part. If the threshold too large, *e.g.*, δ=0.001 in Figure 6(c), some right pairs of times will also be dropped so that the total BMR gets bigger.

2) Key bit rate: Obviously, a main factor that will influence the KMR is the sampling frequency, because with higher frequency, there will be more samples, and therefore, more bits will be extracted. If the sampling frequency and the sampling time are constant, the BMR can reflect the approximate level of the KBR. Intuitively, a higher BMR means that more bits are mismatched and deleted, consequently lowering the KBR. The measurement results demonstrate this as shown in Figure 6(c) and 7, where the sampling frequency is set to 50 Hz and the sampling times are all 3 seconds. In Figure 6(c), the black solid line ($\delta = 0$) becomes nearly parallel to horizontal axis when the distance is larger than 14 cm, and a same tendency of the line appears in Figure 7.

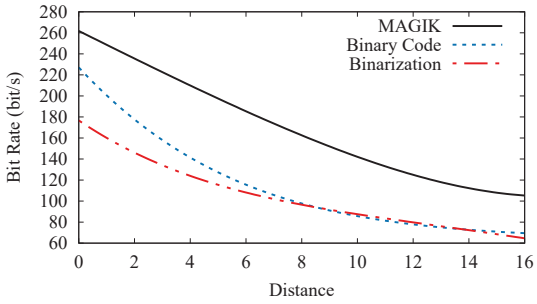


Fig. 7. Bit rate under different distances and different coding methods.

Finally, we compare MAGIK with the traditional way of quantification, binarization, and the binary code way in encoding, shown in Figure 7. Respectively. MAGIK can reach a high KBR of more than 260 bit/sec. From the quantification perspective, our maximum angle time quantification gains a nearly 1.5 times bigger KBR, when compared with the traditional binarization method. From coding methods perspective, our used gray code method gains an higher KBR than that of binary code. We proposed new ways of coding and quantification have dramatically increased the KBR.

Based on the KBR results, the secure distance d and the total extraction time T can be selected according to users' security requirements. If a 512 bits key is needed between two parties, their distance should be less than 8cm and T should be set to 3.5 seconds. Under this constraint, any key extraction of two parties 8cm away won't success. And it is impossible for Eve to to use the reconciliation message exchanged between Alice and Bob to adjust her on measurement to get the same key if Eve is 8cm away from Alice and Bob.

VII. Related Work

We review some of related works from two perspectives, secret key establishment and geomagnetic field applications in networking community.

The problem of secret key establishment has been studied for many years, and several protocols have been proposed to build the secure channel between two devices, especially in wireless scenarios. Bianchi *et al.* [18] discussed the key

exchange negotiation protocol for wireless sensor network and proposed negotiation based on the TLS handshake idea to obtain maximum flexibility. Zhang *et al.* [19] proposed a novel random perturbation-based scheme for pairwise key establishment without exposing any secret to others. This kind of work mainly utilizes the cryptical protocols (*e.g.*, RSA, ECC, *etc.*) and therefore, they may introduce heavy cost on the devices. The idea of proximity based authentication for wireless devices is proposed in [20], however, it only uses signal strength to authenticate user and the performance is not so good because of weak signals. Then, the idea of generating secret key from inherent randomness shared between two parties has been realized in amounts of works [2], [4]–[6], [21]–[25], and different sources have been explored, such as channel state information (CSI), received signal strength (RSS), channel impulse response (CIR), and *etc.* However, they are constrained for popular usage since they either require extra devices [23] or require long time to generate a long enough key. In addition, several prior works have generated the shared key by shaking devices equipped with accelerometers [26], [27]. Although the accelerometer is as common as the magnetometer on mobile devices nowadays, the accelerometer can only record the movement of its own while the magnetometer can sense the environment and the movement together.

The geomagnetic field is also studied by several works (not in astrophysics), but most of them are about the indoor localization or tracking using GMF. For example, Li *et al.* [28] studied the feasibility of using magnetic field alone for indoor positioning. Chung *et al.* [15] proposed an indoor positioning system that measures location using disturbances of the Earth's magnetic field caused by structural steel elements in a building. Afzal *et al.* [14] studied the assessment of indoor magnetic field anomalies using multiple magnetometers. These works leverage different characteristics of indoor geomagnetic field. However, most of them are not implemented on mobile devices, since they require precise measurement or several magnetometers. Lee *et al.* [29] presented a smartphone-based indoor pedestrian tracking using geo-magnetic observations but can only detect the pedestrian turns the corner at a right angle. Our work has successfully realized the system on Nexus 7 and can deal with the most indoor situations.

VIII. Conclusion

This paper has explored the problem of secret key establishment for mobile wireless devices and presented a secure and efficient key extraction mechanism based on the dynamic geomagnetic field in indoor environments. We have carefully studied and justified the feasibility of utilizing geomagnetic field for key extraction through extensive measurements in real-world scenarios. We then have optimized the key extraction process and proposed two new rotation-angle-based quantification methods, which can achieve faster key generation rate with even lower bit mismatching ratio. Besides, we have also implemented our design directly on popular commodity mobile devices and validated its performance by conducting real-word experiments. Experiments results confirm that our system is efficient in key extraction rate and robust in key establishment without additional overheads on mobile devices.

References

[1] W. Stallings, *Cryptography and Network Security: Principles and Practice, third ed.* Prentice Hall, 2003.

[2] S. Jana, N. Patwari, and S. V. Krishnamurthy, "On the effectiveness of secret key extraction from wireless signal strength in real environments," in *ACM MobiCom*, Sep. 2009.

[3] J. E. Hershey, A. A. Hassan, and R. Yarlagadda, "Unconventional cryptographic keying variable management," *IEEE Transactions on Communications*, vol. 43, no. 1, pp. 3–6, 1995.

[4] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, "Wireless secret key generation exploiting reactance-domain scalar response of multipath fading channels," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 11, pp. 3776–3784, Nov. 2005.

[5] A. Sayeed and A. Perrig, "Secure wireless communications: Secret keys through multipath," in *ICASSP*, Mar. 2008.

[6] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-telepathy: extracting a secret key from an unauthenticated wireless channel," in *ACM MobiCom*, Jun. 2008.

[7] C. Chen and M. A. Jensen, "Secret key establishment using temporally and spatially correlated wireless channel coefficients," *IEEE Transactions on Mobile Computing*, vol. 10, no. 2, pp. 205–201, Feb. 2011.

[8] H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, "Practical user authentication leveraging channel state information (CSI)," in *ASIACCS*, Jun. 2014.

[9] S. Mathur, R. Miller, A. Varshavsky, W. Trappe, and N. Mandayam, "Proximate: Proximity-based secure pairing using ambient wireless signals," in *ACM MobiSys*, Jun. 2011.

[10] A. Varshavsky, A. Scanned, A. LaMarca, and E. de Lara, "Amigo: Proximity-based authentication of mobile devices," in *ACM UbiComp*, Sep. 2007.

[11] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 5, pp. 2418–2434, 2010.

[12] J. Burnett and P. D. Yaping, "Mitigation of extremely low frequency magnetic fields from electrical installations in high-rise buildings," *Building & Environment*, vol. 37, no. 8-9, pp. 769–775, 2002.

[13] "Earth's magnetic field," https://www.wikipedia.org.

[14] M. H. Afzal, V. Renaudin, and G. Lachapelle, "Assessment of indoor magnetic field anomalies using multiple magnetometers," in *ION GNSS*, Sep. 2010.

[15] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai, and M. Wiseman, "Indoor location sensing using geo-magnetism," in *ACM MobiSys*, Jun. 2011.

[16] F. Gray, "Pulse code communication," in *U.S. Patent 2632058*, 1953.

[17] A. R. Calderbank, G. D. Forney, and A. vardy, "Minimal tail-biting trellises: The golay code and more," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1435–1455, Jul. 1999.

[18] G. Bianchi, A. T. Capossele, A. Mei, and C. Petrioli, "Flexible key exchange negotiation for wireless sensor networks," in *WiNTECH*, Sep. 2010.

[19] W. Zhang, M. Tran, S. Zhu, and G. Cao, "A random perturbation-based scheme for pairwise key establishment in sensor networks," in *ACM Mobihoc*, Sep. 2007.

[20] S. Banerjee and A. Mishra, "Mobicom poster: secure spaces: location-based secure wireless group communication," *Acm Sigmobile Mobile Computing and Communications Review*, vol. 7, no. 1, pp. 68–70, 2003.

[21] E. K. Lee, S. Y. Oh, and M. Gerla, "Frequency quorum rendezvous for fast and resilient key establishment under jamming attack." *Acm Sigmobile Mobile Computing & Communications Review*, vol. 14, no. 4, pp. 1–3, 2010.

[22] H. Liu, Y. Wang, J. Yang, and Y. Chen, "Fast and practical secret key extraction by exploiting channel response," in *INFOCOM*, Apr. 2013.

[23] K. Zeng, D. Wu, A. Chan, and P. Mohapatra, "Exploiting multiple-antenna diversity for shared secret key generation in wireless networks," in *IEEE INFOCOM*, Mar. 2010.

[24] H. Liu, J. Yang, Y. Wang, Y. Chen, and C. E. Koksal, "Group secret key generation via received signal strength: Protocols, achievable rates, and implementation," *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2820–2835, 2014.

[25] Y. Wei, K. Zeng, and P. Mohapatra, "Adaptive wireless channel probing for shared key generation," in *IEEE INFOCOM*, Apr. 2011.

[26] R. Mayrhofer and H. Gellersen, "Shake well before use: Intuitive and secure pairing of mobile devices," *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, pp. 792–806, 2009.

[27] D. Bichler, G. Stromberg, M. Huemer, and M. Löw, "Key generation based on acceleration data of shaking processes," in *ACM UbiComp*, Sep. 2007.

[28] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos, "How feasible is the use of magnetic field alone for indoor positioning?" in *IPIN*, 2012.

[29] S. Lee, Y. Chon, and H. Cha, "Smartphone-based indoor pedestrian tracking using geo-magnetic observations," *Mobile Information Systems, IOS Press*, pp. 124–137, Sep. 2013.