CHAPTER 2

LITERATURE SURVEY

WSN is more complex than the conventional network due to resource constraint and low computational ability. WSN is prone to new attacks and WSN developers face great challenge to develop efficient security algorithms. A process called data aggregation collects the sensor data from all sensor nodes to reduce the computational complexity at the base station. Hence, WSN requires lightweight security algorithm to protect data from hackers. In this chapter, survey of relevant security issues and existing security solutions are discussed.

2.1 EXISTING SECURITY SCHEMES IN WSN

WSN has experienced tremendous growth in recent years because of its low cost and its capability to render new services. However, compared to traditional network, it is more vulnerable to security attacks due to communication in wireless media. There are two main approaches in use today namely Hop-by-Hop (HbH) and Encrypted Data Aggregation (EDA) for securing data transfer in WSN.

2.1.1 Hop-by-Hop (HbH) Scheme

The lifetime of the WSN is decreased when encryption, decryption and aggregation are performed in every intermediate node. Due to the encryption and decryption process at every intermediate node, the confidentiality of the sensed data is not preserved. The compromised intermediate node exposes an attacker to reveal the sensed data. However, this scheme leads to large overheads for resource constrained WSN (Zhu et al 2006, Castelluccia et al 2005, Yang et al 2008, Labraoui et al 2012, Lou et al 2012).

2.1.2 Encrypted Data Aggregation (EDA) Scheme

An encryption scheme, called homomorphic encryption is used to encrypt the sensor data. The encrypted data is aggregated without performing the decryption at intermediate sensor nodes. Sink node retrieves the data by decrypting the received aggregated data with the help of known shared secret key of the neighboring sensors. Here, the message is not decrypted at intermediate nodes, thereby ensuring confidentiality (Doyle et al 2006, Ozdemir & Xiao 2013, Lu et al 2014). The processed data of neighboring sensors is, often redundant or highly correlated. Therefore, transmission of redundant or highly correlated data causes over consumption of bandwidth.

HbH and EDA schemes are compared and evaluated based on the parameters such as data integrity, computation cost, vulnerability and security. The EDA scheme provides higher data integrity and high security than HbH, but it leads to high computational cost. However, EDA overcomes the passive attack problem in the network. But still EDA performs poor for an active attack.

2.2 EXISTING CRYPTOGRAPHY SOULTIONS FOR WSN

A number of basic cryptographic solutions exist to provide security for WSN (Chen et al 2009). Figure 2.1 shows the existing cryptographic techniques used to meet the security requirements of WSN. The objective of existing cryptographic techniques is to provide high level of security. In

WSN, these solutions are broadly classified into Random, Symmetric and Asymmetric Key Cryptosystem.

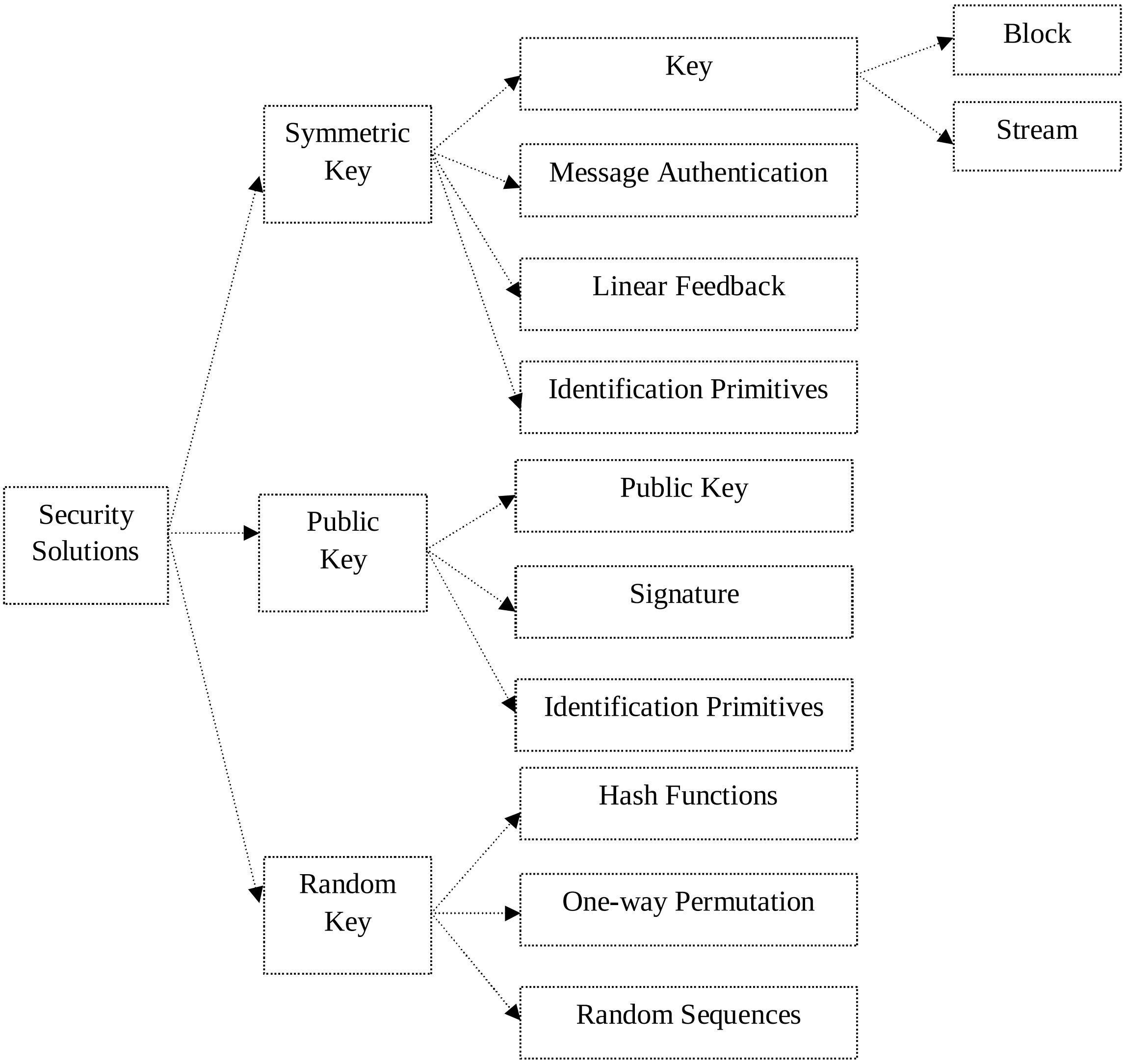

Figure 2.1 Existing cryptographic techniques

2.2.1 Symmetric Key Cryptosystems (SKC) for WSN

Data Encryption Standard (DES) algorithm is a standardized SKC based block cipher algorithm, developed in the year 1976 by NIST (Federal Information Processing Standard (FIPS)-46). Advanced Encryption Standard (AES) replaces the DES algorithm, due to various security breaches in DES. AES supports various key sizes such as 128, 192, and 256 bits (FIPS-197).

Other existing block ciphers for WSN applications are Skipjack (Brickell et al 1993), International Data Encryption Algorithm (IDEA) (Leong et al 2000), Tiny Encryption Algorithm (TEA) (Wheeler & Needham 1995) and Extented TEA (XTEA) (Moon et al 2002). TEA involves simple and fast operations which require less memory. The standardized WSN security protocol such as TinySec (Karlof et al 2004, Zigbee (Alliance 2006) and TEA uses Skipjack algorithm to ensure the authentication in various WSN applications. Hence, these algorithms are of worth consideration in special environments like embedded systems and WSNs. AES, RC6 and TEA algorithms require minimum computational resource. Hence, they are more suitable for resource constraint nodes.

Another category of SKC is stream cipher. The stream cipher uses XOR operation on bit-by-bit plaintext and key value for both encryption and decryption process. It requires less hardware resources and provides high throughput. RC4 stream cipher is most widely used security algorithm in Wired Equivalent Privacy (WEP) and Secured Socket Layer (SSL) protocol. However, it does not have withstanding capability in various potential attacks. Other existing stream ciphers such as A5/1, A5/2, Dragon, HC-256, LEX, Phelix, Py, Salsa20, SOSEMANUK and eSTREAM are implemented in software (Preneel & Rijmen 2008) for WSN application. Some of stream ciphers such as Trivium, Grain, Mickey-128, and Phelix are implemented in hardware (Good & Benaissa 2007). In order to provide high level of security in stream ciphers, there is a demand for efficient implementation techniques. The computational time of stream cipher is lesser than the block cipher, due to its ability of encrypting few bits. This kind of stream cipher significantly increases the throughput of an application.

2.2.1.1 Security analysis of SKC

A symmetric key approach requires revisiting various functionalities of existing cryptography primitives in order to identify a suitable algorithm that consumes less resources for the available sensor node. Most of the security algorithms are implemented in software rather than hardware due to its inherent resource constraints. The behavior of the nodes are dynamic in nature, hence the software implementation is more suitable for security than the hardware (Hankerson et al 2000). Though SKC provide security primitives and is easy to implement, it leads to various potential attacks due to its less complex algorithms. Hence, it is suggested to use Asymmetric Key Cryptosystem (AKC) algorithms than SKC to prevent the various attacks of WSN applications.

A well-known AKC encryption scheme known as Rivest, Shamir and Adleman (RSA) is invented in 1978 (Koc 1994, Koc 1995). The complexity of an algorithm depends on key generation, encryption and decryption function. The difficulty for an attacker lies in complexity of computing integer using factorization technique. However, it suffers from different kind of attacks such as factorization attack, low encryption exponent attack, low decryption exponent attack, etc (Forouzan 2007).

2.2.2 Asymmetric Key Cryptosystems (AKC) for WSN

Wireless channel in WSN has experienced tremendous vulnerability in hostile environment because it is more vulnerable to eavesdropping (Makin & Padha 2010). The complicated operations such as modular multiplications of large numbers are involved in encryption and decryption process. Exclusively in public key cryptosystems large resources of sensors are consumed (Arazi et al 2005). The three most widely used AKC algorithms are RSA, DSA and ECC.

Ronald et al (2004) described the design and implementation of public key-based protocols that allow authentication and key agreement between a sensor network and a third party as well as between two sensor networks. The author implemented the work in UC Berkeley MICA2 motes using the TinyOS development environment. The efficient RSA based public key cryptosystem used in design. If a node is compromised by the authentication, then the entire network will become unsafe (Yang et al 2015). However, RSA is computationally expensive operations for sensor network due to the large size of key (1024-bits) and more vulnerable for Discrete Logarithm Problem (DLP) than ECC based public key cryptosystem.

Wei-hong et al (2008) identified the major security problem faced by WSN such as confidentiality, node authentication, message integrity and freshness. Data link layer is responsible for encryption of data frame. Network layer and the application layer are responsible for key management and message exchanging process in WSN (Raymond & Midkiff 2008). The most extensive tool of cryptosystem is public-key cryptosystem used to solve the problems in information security. The most important existing cryptography scheme that provides highest security quality for each key bit in WSN application is Elliptic Curve Cryptography (ECC).

ECC provides high level of security among the existing cryptosystems. Besides security, ECC offers other features like small key size, minimum system parameter, less bandwidth, faster computation, easy implementation, low power, and low hardware requirements. The level of security offered by 160-bits of ECC key is equal to 1024-bits of RSA key. The challenges faced by AKC are high computational complexity, energy consumption, key size and more memory. Hence, it is difficult to implement

ECC based algorithms in resource-constrained devices. Energy consumption of ECC in AKC is accounted due to its large number of computation compared to SKC. The level of security obtained by AKC of 512-bit ECC key size is equal to the level of security achieved by SKC of 256-bit AES key size (Wei-hong et al 2008). As a result, large key size in AKC leads to more energy consumption by performing large number of computation. Hence, bandwidth requirement and computational requirement in AKC is large. Therefore, AKC is not suitable for resource-constrained WSN application compared to SKC.

Further, to reduce the computational complexity of ECC, various optimizations techniques are adapted to support resource-constrained environment. In this thesis, an efficient optimization technique is proposed in order to reduce the computational complexity of ECC.

2.2.2.1 Related works on ECC in WSN

ECC is more suitable for resource constraint applications such as embedded and WSN applications. Elliptic Curve Discrete Logarithmic Problem (ECDLP) decides the security level of ECC. The algebraic structure of elliptic curve performs arithmetic operations such as point addition and point multiplication. Hence, the time taken for multiplication process in ECC for generating the key, encryption, authentication and decryption is less than the RSA exponentiation process (Hankerson et al 2004). Figure 2.2 shows the taxonomy of ECC for WSN.

Boyle & Newe (2009) found that the ECC algorithm is based on the algebraic structure of finite fields over elliptic curves. It provides a reasonable computational load by smaller key sizes with equivalent security. Smaller key size in ECC reduces the size of message buffers and the implementation cost of the protocols. Key agreement schemes based on ECC provide a better security for the exchanged keys in cryptosystem. There are number of attacks exist in key agreement scheme when exchanging public key parameters from key generation. An authenticated public key based key agreement scheme is developed using key exchange mechanism. This is based on the Elliptic Curve Menezes-Qu-Vanstone (ECMQV) algorithm for WSN applications. It eliminates Man-In-The-Middle (MITM) attack that occurs in key exchanging mechanism than the conventional Diffie-Hellman key exchanging mechanism.

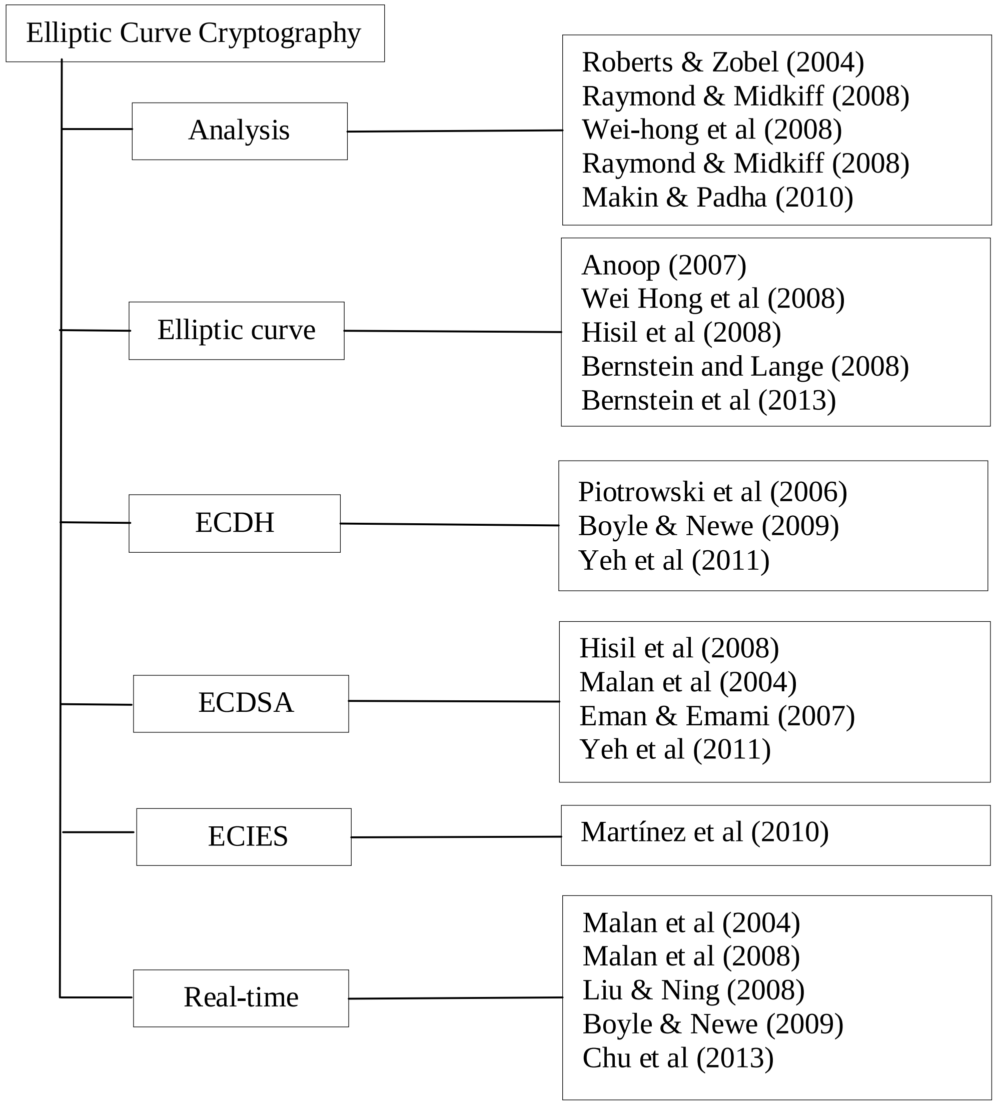

Figure 2.2 Taxonomy on Elliptic Curve Cryptography

Anoop (2007) identified an efficient implementation of ECC which is based on point multiplication or scalar multiplication and field arithmetic. Implementation of ECC using projective coordinates builds efficiency of ECC compared to implementation of ECC using affine coordinate. It is due to the elimination of multiplicative inverse operation in point addition and doubling in ECC. Otherwise, multiplicative inverse considers high cost by involving more processor cycles. A trinomial or pentanomial binary field of ECC such as GF (2^m) is chosen to implement ECC efficiently. A binary field called GF (2^m) over ECC makes efficient implementation of ECC than prime field GF(p). Specific domain parameters and irreducible polynomial (either trinomial or pentanomial) are given in the specification. Efficiency of ECC using the chosen polynomials for polynomial reduction in binary field is larger than ECC using modular reduction in prime field. Therefore, the chosen domain parameters and polynomials help to provide faster key generation in ECC.

Malan et al (2004) developed elliptic curve over GF (2^m) for WSN applications using 8-bit, 7.3828MHz, MICA2 mote. Public key cryptosystems are feasible key generation algorithms in TinySec MICA2. Liu & Ning (2008) suggested real time implementation of ECC on sensor nodes in TinyECC. There are number of optimization scheme in scalar multiplication schemes such as Non-Adjacent Form (NAF), width NAF (wNAF), sliding window NAF (swNAF), inline assembly, modular addition, modular subtraction and Sharmir’s trick that exists for reducing the complexity of ECC. In literature, other existing algorithms, reduces the computational complexity of public key cryptography (Malan et al 2008).

Piotrowski et al (2006) measured the power consumed by RSA algorithm and ECC based cryptographic algorithms. ECC includes signature generation and verification, confidentiality and authentication using common sensor platforms such as MICA2DOT, MICA2, MICAz and TelosB. From the experimental analysis, it is observed that ECC influences the network lifetime significantly and ECC is feasible to implement WSN applications on the devices. From the implementation of ECDSA based signature scheme in sensor network, it is observed that the amount of time consumed by ECDSA signature generation and verification is lesser than the amount of time taken by RSA signature generation in the sensor network.

Hisil et al (2008) introduced a normal form of elliptic curve with an exclusive addition law for adding two elliptic curve points. In addition to that, the elliptic curve over a non-binary field is birationally equivalent to the Edward curve over an extension field and extended coordinates of Twisted Edward. This provides a fast point addition algorithm for adding two elliptic curve points by computing 3-fold and 5-fold on an Edwards curve. Bernstein et al (2008) suggested the modification on Edward elliptic curve over finite field that includes more curves than the original Edward curve.

Malan et al (2008) identified that the amount of memory required for Elliptic curve module (EccM 1.0) is lesser than Discrete Logarithmic Problem (DLP) based Diffie-Hellman, which has minimum key and smaller key size. EccM 1.0 the amount of time consumed to generate a private key and a public key is 34.161 seconds for 100 iterations with a standard deviation of 0.921 seconds. The amount of time consumed for generating a shared secret key including private key generation and public key exchange is 34.173 seconds and on an average the number of iterations with standard deviation involves 0.934 sec. Malan suggested EccM 2.0 (as per NIST recommendation) of ECC over GF (2^m). The selected polynomial function for generating a prime ordered sub group over elliptic curve 'E' is $f(x) = x^{163} + x^7 + x^6 + x^3 + 1$. The adopted reduction polynomial for modulo operations of cryptography is $y^2 + xy \equiv x^3 + x^2 + 1$. EccM 2.0 selects a private key from a prime ordered group GF (2^m) using basic polynomial and computes public key using scalar multiplication using Koblitz curve and base point for a node. The public keys are generated from elliptic curve using Elliptic Curve Diffie Hellman (ECDH). A shared secret key is generated between the nodes.

Boyle & Newe (2009) addressed the implementation of ECC in Crossbow "MICAz" mote that consists of a battery, microprocessor (Atmega128L), RF transceiver, ADC, 128K bytes Program Flash Memory and 4K bytes EEPROM. Security services integrated in each node provide security. Any component of a network implemented without any security is prone to attacks. Therefore, this work lags high level of security in real time applications.

Yeh et al (2011) introduced ECC based authentication protocol using ECDSA for WSN applications to prevent clone attack in sensor nodes. However, the computational complexity of ECDSA is not suitable for resource-constrained applications. Hence, there is a need to reduce the computational complexity in ECDSA.

Bernstein et al (2013) proposed another form of elliptic curve called Twisted Edward curve by generalizing the addition and doubling formulae on actual Edward curve. It also shows that every elliptic curve is in Montgomery form and consumes lesser time for computing Twisted Edward Curves than the existing Edward curve. Twisted Edward curve of order 4 has been identified as bi-rationally equivalent to Montgomery curves. However, the Edward elliptic curve is vulnerable to side channel attacks. Also, it is prone to Simple Power Analysis (SPA) attack.

Implementation of Twisted Edward curves over prime field in 8-bit AVR processor is carried out by Chu et al (2013). The computational time is increased by 2.78 times and the energy consumption is reduced by one third than the conventional TinyECC in Mica motes. Hence, MICAz motes performs 1,74,000 number of ECDH key exchanges before running out of battery. It consumes 38.7 mJ per mote for the execution of ephemeral ECDH key exchange.

2.2.2.2 Security analysis of AKC

Roberts & Zobel (2004) measured the level of security offered by an ECC using the difficulty of ECDLP by the multiplicative group over an elliptic curve. A prime ordered cyclic subgroup causes large ECDLP value, thereby, the security of the elliptic curve is more than the existing multiplicative elliptic curve. The level of security of ECC depends on the difficulty of the ECDLP problem. The computationally infeasible ECDLP is decided based on the high ordered subgroups over elliptic curves. It is necessary to choose a higher ordered subgroup or a multiplicative group over elliptic curve for ECDLP. The selection of computationally infeasible ECDLP for higher ordered subgroups decides the difficulty to an attacker. The difficulty in key identification attack patterns increases by selecting higher ordered subgroups over elliptic curve. The computation of ECDLP provides higher level of security than the DLP of traditional integers.

Piotrowski et al (2006) estimated the consumed energy by ECC with 160-bit key generation and verification is 22.82 mWs and 45.09 mWs respectively. Janakiraman et al (2007) suggested a hybrid algorithm, which is a combination of the symmetric and asymmetric encryption techniques for WSN applications. Eman & Emami (2007) introduced a parallel digital signature algorithm using Secured Hash Algorithm-512 (SHA-512) to ensure authentication in WSN. This algorithm generates a digital signature of size 1024-bit using permutation combination between two 512-bit digital signatures. However, this digital signature is generated only after receiving the entire 1024-bit of input data. Potlapally et al (2003 & 2006) estimated the energy consumption of MD5 as 0.59µJ/byte.

In this thesis, an efficient implementation of Twisted Edward curve in TinyECC is proposed. In addition, implementation of ECC based cryptosystem such as ECDH, ECDSA and ECIES using Twisted Edward curve is carried out in MICAz motes to analyze the performance of the proposed scheme.

2.2.3 Random Key Cryptosystem (RKC) for WSN

RKC based approaches such as Rabin's scheme (Rabin 1979), NtruEncrypt (Hoffstein et al 1998) and Multivariate Quadratic (MQ) schemes (Wolf & Preneel 2005) are suggested for resource-constrained applications. Though it uses large key size, this approach consumes optimal execution time for producing the random key. Rabin (1979) suggested a high-speed factorization scheme for large prime numbers, and named as Rabin scheme. However, the security level of Rabin scheme is equal to RSA. Speed of encryption, signature and verification involving squaring operation decides the security level of this scheme. This scheme generates three various results in which two results are false and one is correct. The major drawback of this scheme is computational complexity and the larger key size.

Hoffstein et al (1998) suggested another approach based on a polynomial ring 'R' are NtruEncrypt and NtruSign for encryption and signature generation respectively. Strength of these algorithms is based on the complexity in solving Closest Vector Problem (CVP) and the Shortest Vector Problem (SVP). NtruEncrypt is identified as faster scheme than the existing asymmetric encryption schemes due to its simple primitive polynomial operations involved in encryption and signature generation.

Another most recent approach is multivariate public key generation named Multivariate Quadratic (MQ) scheme for random key generation. This scheme is faster in signature verification process. However, storage cost of this scheme is larger due to its large key size. In this scheme, the private key requires 879 bytes of storage and the public key require 8680 bytes of storage. Nevertheless, sensor nodes with low memory capabilities do not support this scheme. In addition to these, MQ scheme supports Encrypted Data Aggregation (EDA) that ensures confidentiality. However, it increases transmission overheads. In addition to increased overhead, several security attacks are identified for existing cryptographic algorithms.

In this thesis, a hybrid approach based on EDA named hybrid cryptography is proposed for key generation, encryption, decryption and authentication. The proposed scheme includes a padded message digest to prevent redundant and aggregated data transmission in WSN. AKC algorithm offers more security for WSN applications than SKC algorithms. A hybrid approach combines both SKC and AKC algorithms for providing high security in WSN applications. A combined approach in the proposed hybrid scheme gives further improvement in the level of security.

2.3 EXISTING KEY MANAGEMENT SCHEMES FOR WSN

Another cryptographic scheme to decide the secured communication over wireless media is key management scheme. It includes key generation, exchange, and storage of secret keys. Cryptographic key management plays a vital role in providing reliable, robust, and secure communication. Keys used in WSN are broadly classified into five categories as shown in Figure 2.3.

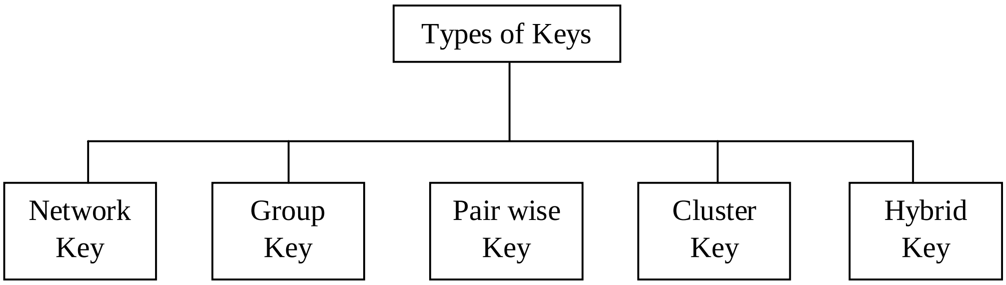

Figure 2.3 Types of keys used in WSN

- Network key: A unique key shared with the base station.

- Pair-wise key: A secret key is shared between two sensor nodes. It is used to maintain privacy of transmitted information.
- Cluster key: A secret key is shared with neighboring nodes to secure local broadcasts.
- Group key: A secret key is shared with all the nodes in the network group and it is used to send multicast and broadcast messages.
- Hybrid key: A combined keying mechanism of any two or more keys using the above-mentioned types.

2.3.1 Key Management Schemes

The key management schemes are broadly categorized into Pre-distributed, Self Enforcing and Arbitrated (Lee et al 2007). Figure 2.4 shows the main classification of key management scheme for WSN.

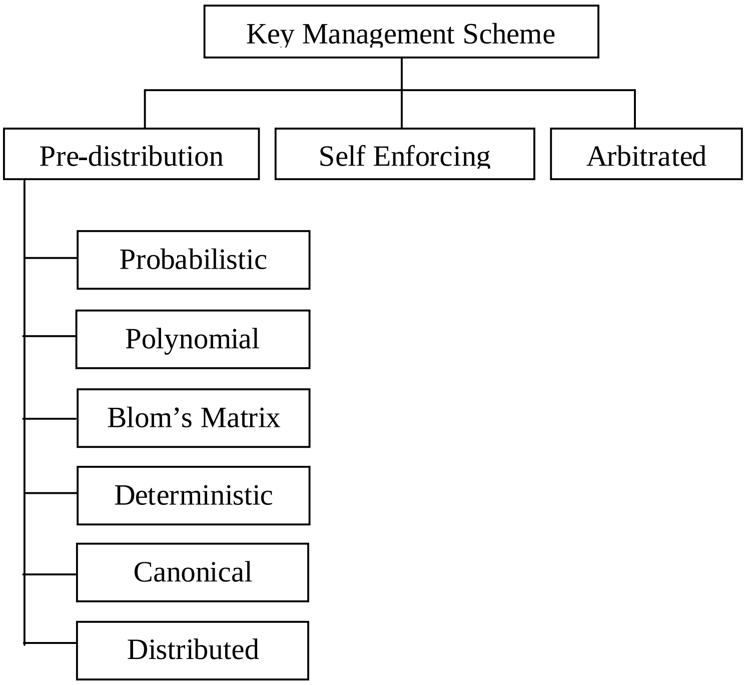

Figure 2.4 Classification of key management schemes for WSN

Hwang & Kim (2004), Zhang & Varadharajan (2010) suggested pre-distribution based key management scheme for sensors with preloaded set of keys. Hwang & Kim (2004) also introduced Bloom’s model for sharing the key between two nodes for WSN. They model consists of a private key matrix, which is stored in each of the nodes. A single row and column is exchanged between two nodes by key establishment process. These preloading processes of key distribution do not include any computation on key sharing (Xiao et al 2007). It involves no computation and only selection of keys is carried out from the preloaded key pool. There are number of methods such as probabilistic (Du & Chen 2004, Eltoweissy et al 2006), polynomial (Blundo et al 1993, Liu et al 2005), Blom’s matrix (Blom 1985, Hwang & Kim 2004, Zhang & Varadharajan 2010), deterministic (Sanchez & Baldus 2005, Lee & Stinson 2005, Zhang et al 2011), canonical (Chan & Perrig 2003, Tague & Poovendran 2007) and distributed (Das 2011) exist. Pre-distribution based keying mechanism provides less computational cost. However, it is unsecure and requires large memory to store key.

Eschenauer & Gligor (2002) scheme provides a key pool in each node. A key is selected from this key pool for secure data communication between two nodes. This method consumes less computations for key exchange process. However, it does not support scalability and needs more memory. Du et al (2009), Shen et al (2009), Sahingoz (2013) introduced the self-enforcing scheme that uses the asymmetric cryptographic algorithms such as RSA, ElGamal and ECC for key management. Though asymmetric key algorithm involves large number of computation, it provides high security for applications (Haque et al 2008). Sensors are limited by computational capability and energy constraints. Implementation of this type of key management is difficult. Lopez et al (2010) introduced the arbitrated keying scheme relying on trusted central point or trusted third party that ensures the authentication of a node. The advantage is that authentication is ensured if the third party is genuine. The drawback of this scheme is that to estimate trusted central point and it requires resource to serve as a third party. This is vulnerable to single point failure attack.

Zhu et al (2006) suggested a pre-distribution key mechanism named Localized Encryption and Authentication Protocol (LEAP) to exchange four different types of keys. However, it requires high storage because of four types of keys. This mechanism supports cluster, pair-wise and network wide operations based on TinySec and LEAP (Tobarra et al 2007).

A scheme named Scalable, Hierarchical, Efficient, Location aware and Light-weight (SHELL) is suggested by Younis et al (2006) to generate and manage the keys using distributed based key management scheme by combinatorial matrix. It provides support to add and replace nodes in the network. The overall energy cost of the system is larger because of the complex operations involved in the scheme.

Shaikh et al 2009 claimed that cluster based group key management protocol is more robust than network-wide keys. In multicast communication, a group of members shares a key called group key. Moreover, group key scheme supports scalability of network as it is capable of efficiently managing the addition and removal of nodes. A cluster based group key establishment protocol namely Hierarchical Key Agreement Protocol (HKAP) (Kodali & Chougule 2013) and Group Key Agreement Cluster Head (GKA-CH) (Klaoudatou et al 2011) exist for WSN applications.

Makin & Padha (2010) suggested a secure data aggregation based on the trust value of a node in WSNs to ensure data confidentiality and data integrity. Conti et al (2011) identified that distributed approach is not suitable for WSN as it consumes more energy and large memory. Klaoudatou et al (2011) introduced a key management protocol named Cluster-Based Group

Key Agreement (CBGKA) protocol for WSN. This scheme is not secure because the common key is shared with all members of the group. If the group key is compromised then all members are prone to various kinds of attacks.

From the literature survey, security solutions in WSN require employing and managing cryptographic keys for better security. The security level of WSN depends on the security level of cryptographic keys. Hence, an efficient key establishment and management is required in order to preserve the high-level security by providing secure cryptographic keys. The performance analysis of the cryptographic key is reviewed and evaluated in terms of simplicity, scalability, robustness and storage efficiency.

2.4 EXISTING DATA AGGREGATION PROTOCOLS FOR WSN

Aggregating data from a compromised node imposes a serious issue in decision-making applications. Hence, it is essential to ensure certain security requirements, such as confidentiality, integrity, authentication, availability, non-repudiation, authorization, freshness, forward secrecy and backward secrecy, etc. However, the vulnerability of a WSN to external and internal attacks makes it tedious to provide secured data aggregation. Moreover, the sensor network is highly constrained in terms of computational capabilities, memory, communication bandwidth and battery power. Hence, there is a need to provide a simplified security mechanism during data aggregation (Coleman et al 2004, Jha & Sharma 2011, Ozdemir & Xiao 2011).

In order to meet the security requirements, existing techniques perform complex encryption and authentication algorithms on the nodes. In addition to data security, the information related to security algorithms such as key generation and key exchange is essential to provide integrity to the data (Dutertre et al 2004, Karlof et al 2004, Du et al 2009, Das et al 2012, Galindo et al 2012). Along with data security, extraction of keys by the hackers must be made infeasible by ensuring the security of the keys. Thus, the key generation process prevents the attacks. Several data aggregation protocols play a vital role in increasing the performance of WSN. The choice of data aggregation protocol depends on the type of application in WSN. Security services are necessary to protect the data in WSN, especially in critical application like monitoring, healthcare, military applications, etc.

In this section, data aggregation protocols designed for various sensor network architectures such as Flat and Hierarchical Network are discussed. All sensor node possess similar equipment to perform similar task in Flat Network, whereas Hierarchical Network performs data fusion, data aggregation and control in selected node to reduce the number of messages transmitted to the sink. Therefore, the number of process executed on sink node is reduced and the communication involved in the network is significantly reduced. Hence, energy efficiency of the network is improved. Figure 2.5 shows the taxonomy of data aggregation schemes.

Westhoff et al (2006) introduced Concealed Data Aggregation (CDA) using public key based additive homomorphic encryption algorithm for data aggregation in WSN. It also concludes by providing the recommendations for selecting the most suitable public key schemes for different topologies and wireless sensor network scenarios.

Al-Karaki & Kamal (2004) introduced the WSN aggregated protocol named Sensor Protocol for Information via Negotiation (SPIN). It is based on Push based Diffusion Protocol (PDP). A source node floods the sensed data after an event is detected and initiates the diffusion of data to the sink node. SPIN consists of encryption and broadcast authentication process namely Secure Network Encryption (SNEP) and micro-Timed, Efficient, Streaming, Loss-Tolerant Authentication (µTESLA) for SmartDust motes. SPIN supports security negotiation and resource adaptation for successful data transfer in the network.

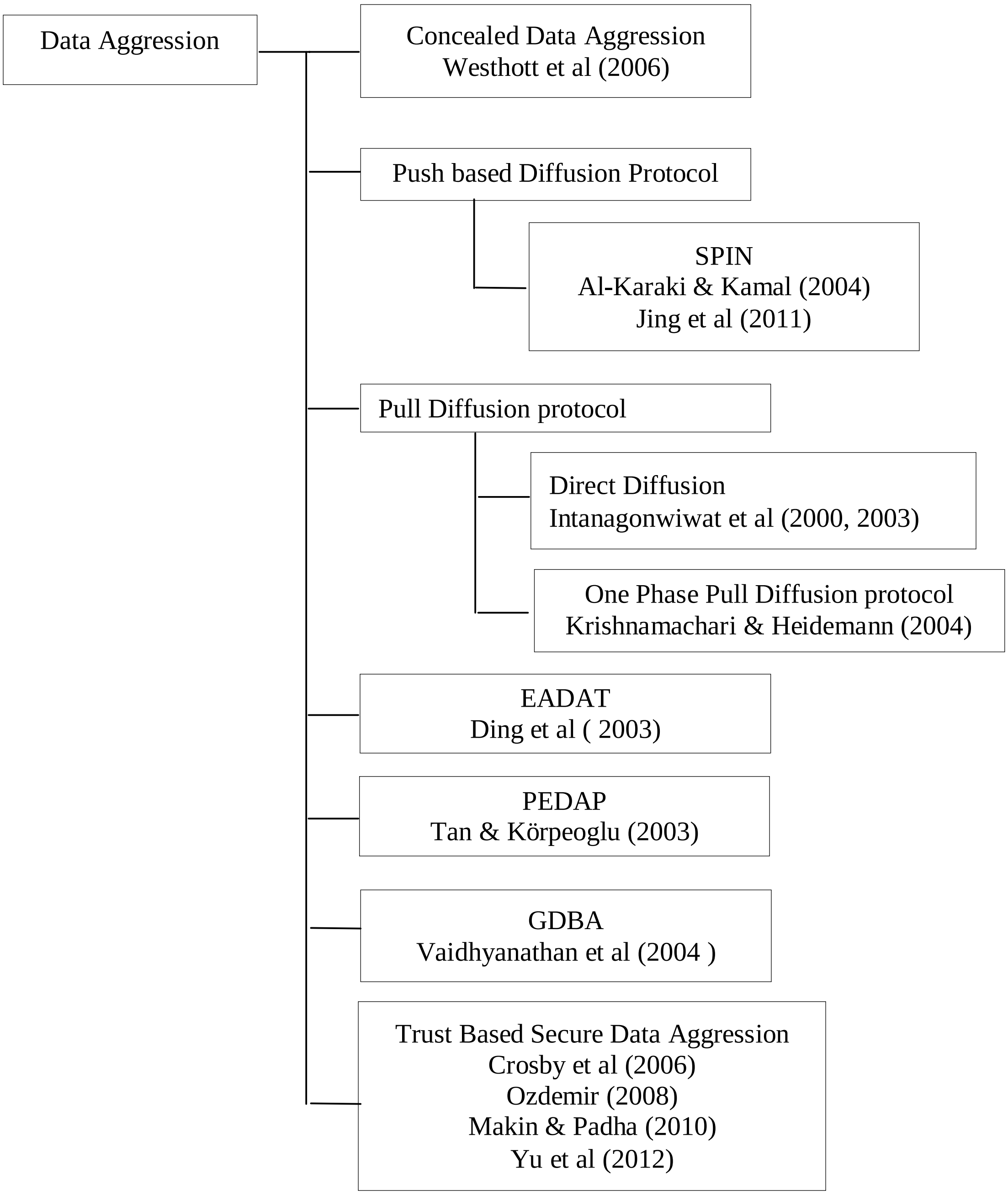

Figure 2.5 Taxonomy on data aggregation schemes

The salient feature of SPIN protocol is that the detail about the single hop neighbors is sufficient to adopt the topological changes in network. However, it does not support guaranteed data delivery. If the intermediate nodes are not interested on that particular data, then the packets are dropped. It leads to injection of an intruder in the network. SPIN also suffers with ‘blindly forward’ and ‘data inaccessible’. In order to overcome this issue Jing (2011) introduced energy saving routing algorithm named SPIN-1 based on SPIN protocol for WSN. This routing algorithm is introduced in order to achieve efficiency.

Intanagonwiwat et al (2003) developed an energy efficient data aggregation protocol called Directed Diffusion (DD) protocol. A routing protocol uses data centric scheme named Two Phase Pull Diffusion (TPPD). All the intermediate nodes in the network do the aggregation process and forward the data based on the threshold value using local rules to drive the data in the network. With large number of source nodes and few number of sink nodes, this algorithm performs energy efficient data aggregation in the network. DD protocol performs better than the omniscient multicast data aggregation scheme. This scheme provides high-energy efficiency by transmitting the data with shortest path in multicast tree to the sink node. The amount of energy consumed by DD is only 60% than the amount of energy consumed by an omniscient multicast scheme, whereas the average delay involved in both data aggregation schemes are equal.

Krishnamachari & Heidemann (2004) suggested one-phase pull diffusion scheme to eliminate flooding process in directed diffusion for reducing the overhead by considering large number of sources with few sink nodes in two phase pull diffusion scheme. This scheme propagates the interested message of sink through the network by finding the appropriate next-hop neighbor. Hence, the control overhead in this process is reduced by removing the unwanted data transmission in the network. However, the sources send only the absolute data. The delay involved in this process is very less. The source response is high, only when it satisfies the interest of sink node. The amount of excessive control overhead is decided based on the choice of diffusion mechanism. However, it increases 80% in control overhead compared to push diffusion protocol.

A flat network results in faster depletion of its battery power by excessive communication computations at sink node. Death of sink node breaks down the functionality of the network. There are several number of data aggregation approaches suggested for high scalability and energy efficiency.

The hierarchical data aggregation involves data fusion at special nodes to reduce the number of messages transmitted to the sink. Hence, the energy efficiency of the network is improved. Several literatures discussed the advantages and limitations of hierarchical data aggregation.

Klaoudatou et al (2009 & 2011) suggested a cluster based Group Key Agreement protocols for WSN. The deployment of infrastructure-based and infrastructure-less based architectures is examined for energy consumption to identify the impact of the key agreement protocol.

Sensor nodes are organized in tree structure to perform data aggregation at intermediate nodes. An aggregated data is transmitted to the root node. This tree based data aggregation is suitable for in-network data aggregation process to perform energy efficient data aggregation than the hierarchical data aggregation process. Makin & Padha (2010) suggested a Trust-Based Secure Data Aggregation Protocol to protect the sensitive data. The trust value is obtained using Bayesian technique to ensure data confidentiality and authentication of a node in WSN. Each node poses Combined Trust Values (CTVs) that provide trust evaluation factor, such as ID, sensing data and consistency to identify the compromised node and filter the data in the network. The base station eliminates the misbehaving aggregator nodes. The drawback is overhead on base station and high latency.

Ding et al (2003) suggested an Energy Aware Distributed Heuristic Aggregation Technique (EADAT) to provide data aggregation by constructing and maintaining tree based aggregation process in sensor network. Sink node broadcasts a control message to initiate the aggregation process and the root node performs the data aggregation process. A control message includes ID, parent, power, status and hop count. A next hop neighbor node is considered based on higher residual power and its ideal channel condition of the node. This algorithm provides an efficient data aggregation protocol by finding and arranging the root node and leaf node using residual power. A root node is decided based on the threshold value of residual power. A leaf-node identifies its root node using the next residual value of its neighbor. This leaf-node sends a hello message to join with the new root node. The advantage of this algorithm is that the average residual energy of alive sensors is decreased by reducing the number of hops to reach the root node. In addition to this, the network lifetime is increased in dense environment. It is very difficult to maximize the network lifetime, if every sensor node possess the data to transfer to the sink node (Ozdemir & Xiao 2009).

Tan & Körpeoglu (2003) suggested a Power Efficient Data gathering and Aggregation Protocol (PEDAP) for WSN. The goal of PEDAP is to maximize the lifetime of the network in terms of number of rounds. It is based on minimum spanning tree algorithm to improve the lifetime of the network. The residual energy of the nodes is considered to balance the load among the existing nodes in the WSN. It requires a prior knowledge of the sink node location. This protocol operates in a centralized manner where the sink node computes the routing information of the network. The time complexity of this protocol is $O(n^2)$ where n is the total number of sensor present in the network.

Vaidhyanathan et al (2004) suggested two data aggregation schemes namely Grid Based Data Aggregation (GBDA) and In-network Data Aggregation (IDA) by dividing the network region into grid. A set of sensors in fixed regions of the sensor network is assigned as a data aggregator in GBDA. A sensor in a network grid transmits the data directly to the data aggregator. Hence, the sensors in the grid do not communicate directly with each other. It is more suitable for mobile environments such as military surveillance and weather forecasting. It also supports dynamic changes in the network topology. Here all sensors directly transmit data to a predetermined grid aggregator.

In IDA, the aggregation process is similar to the GDBA aggregation process. However, it varies by allowing every sensor within a grid to communicate with its neighboring sensors and any sensor node within a grid act as a data aggregator. In IDA, the most critical information is aggregated and forwarded to the sink after performing the data fusion. The data aggregator is decided based on the highest signal strength possessed by a node. It is suitable for environments where the events are highly localized.

Ali et al (2008) presented various communication protocols designed for efficient utilization of energy resources for a sensor node. They also obtained the real time functionality and energy efficient routing for WSN. The performance analysis of designed protocols is tested for surveillance application in terms of latency, scalability and energy awareness. Girao et al (2006) have introduced additively homomorphic public-key encryption algorithms for data aggregation process in WSN. The authors tested public key cryptosystem under various topologies to support data aggregation process.

Makin & Padha (2010) identified security challenges and issues in data aggregation process of WSN. A trust based secure data aggregation protocol is suggested for WSN to ensure data confidentiality and data integrity using the trust value of sensor node. It protects the transmitted data from passive attack like eavesdropping using data confidentiality. Data integrity prevents altering the final aggregated data by compromised nodes or aggregator node.

Klaoudatou et al (2011) introduced a key agreement protocol named cluster-based Group Key Agreement protocols for WSN. Here, keys are shared within the cluster. The application environment of WSN is classified into two categories namely infrastructure-based and infrastructure-less. The performance analysis of the protocol is carried under various topologies and it is observed that less amount of energy is consumed for processing.

Ganeriwal et al (2008) developed a reputation-based framework for sensor network to maintain reputation of other nodes for evaluating the trustworthiness of sensor node. This algorithm supports scalable, diverse and generalized approach to identify the misbehaving node in the network. This approach also addresses the cluster formation, key generation and key sharing to ensure the secured topology. A malicious node is identified based on the calculated trust value. It also guides a clone node to display the clone's trust value. This data aggregation method provides fewer overheads.

Table 2.1 Comparison of existing security protocols in WSN

Security protocols	Algorithm Used	Key Size	Confidentiality	Integrity	Authentication	Freshness	Security Level	Implementation board
TinySec	RC5, Skipjack -CBC	80	√	-	√	-	Low	Mica, Mica2, Mica2Dot
SPINS	AES	128/160	√	√	-	-	Weak	Mica
MiniSEC	Skipjack -OBC	80	√	-	√	√	Low	TelosB
IEEE 802.15.4	AES, MD5	128	√	√	√	√	Weak	MicaZ, TelosB
TinyPK	RSA	1024	√	√	√	√	Weak	MicaZ
TinyECC	ECC	160/192/ 256/512	√	√	√	√	High	MicaZ, TelosB

Table 2.1 shows that the performance of 128-bit security level of symmetric key is equal to RSA-1024 and ECC-160. The performance comparison between the public key cryptography says that RSA is 10-times slower than ECC. RSA's key generation is slow compared to ECC key generation, with the RSA's being 100 to 1000 times slower. However, this may or may not be a significant consideration in systems that generate keys infrequently. It does matter for certain protocols or policies that require more frequent key generation. Public key signature validation is generally faster with RSA compared to ECC, which can provide a benefit. Typical RSA

implementations currently employ 1024 or 2048 bit keys, yet both are less secure than AES-128 and ECC-160.

From the literature survey, it is observed that the data aggregation process is essential for WSN applications to increase the network life time. Data aggregation process is done at intermediate node. The data aggregation process at intermediate node requires security of data at intermediate nodes. Efficient implementation of ECC is considered for key generation in WSN applications. AES is used for encryption and decryption, MD5 is used for authentication and integrity in WSN applications.

2.5 EXISTING CLONE DETECTION

Two methods namely centralized approach and distributed approach exist for clone detection. Figure 2.6 shows the taxonomy of existing clone detection methods.

Moniruzzaman et al (2009) suggested a clone detection scheme by broadcasting the location claim to its neighbors. The location claim is a combination of node id and x-y coordinates of its location. At least one of its neighbors forwards this location claim to the base station (BS). If the BS receives more than one location claim with the same identity (ID) but different locations (i.e., conflicting location claims), the BS detects node replication attack and then broadcasts a message to the whole network to remove the replicated node in the network. The centralized scheme achieves definite detection of cloned node. This solution has several drawbacks such as single point of failure (BS) or any compromise to BS, and high communication cost due to more number of exchanged messages. Furthermore, the nodes closest to the base station will drain soon. The protocol also delays clone detection, since the base station must wait for all of the location claims to be received, analyze them for conflicts and then flood message throughout the network.

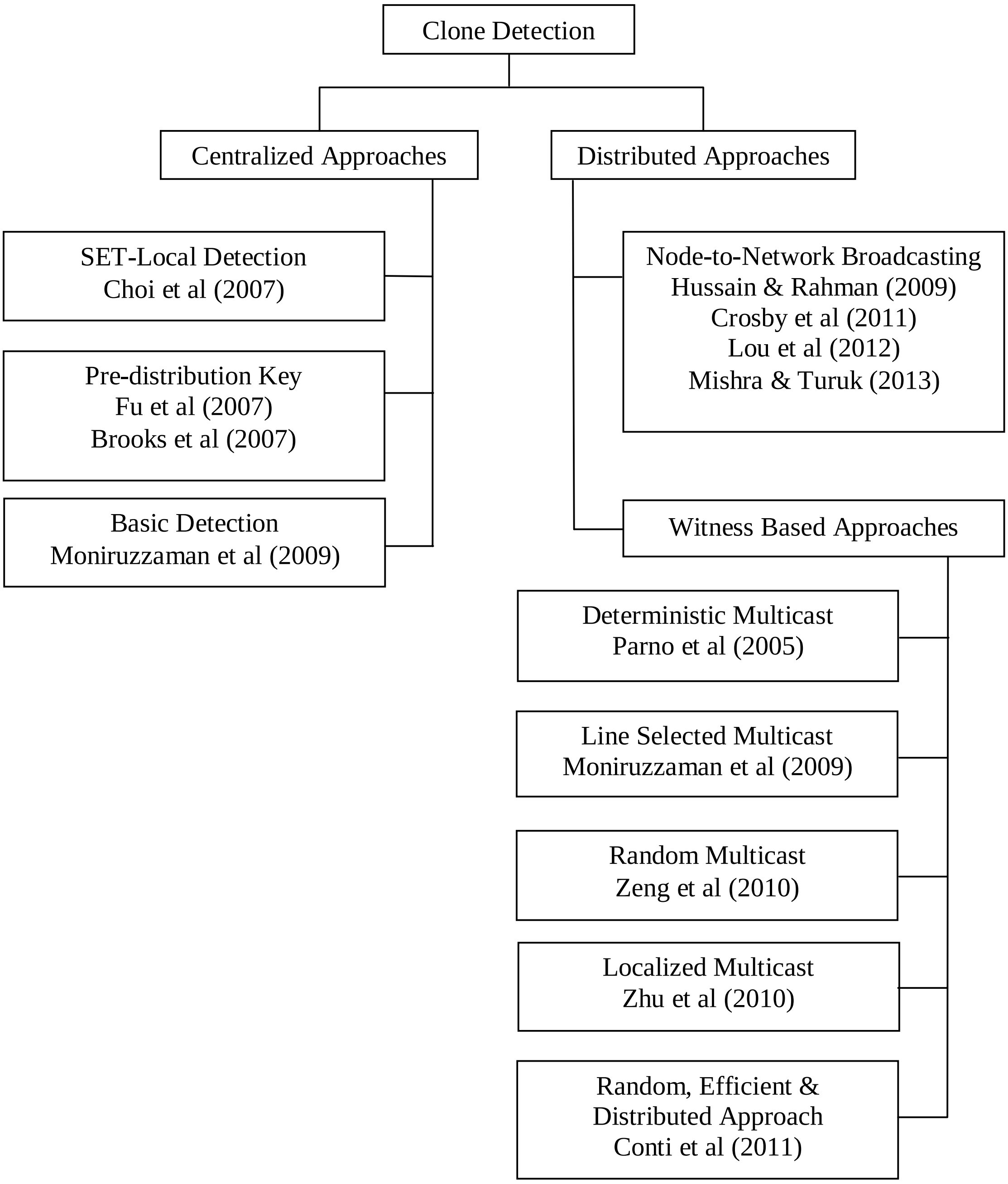

Figure 2.6 Taxonomy on clone detection schemes

Choi et al (2007) recommended clone detection technique named Exclusive Subset Maximal Independent Set (ESMIS) clustering to form exclusive unit subsets among one-hop neighbors by computing intersection and union operations to form a group. All nodes location identification and clone detection are carried out by leading node. It reduces the communication cost by eliminating the redundancy in the node location reports using the subset division procedure. An authenticated subset covering protocol is used to prevent malicious nodes in the ESMIS algorithm. It increases the communication delays and complicates the detection procedure. It also employs a tree structure to compute non-overlapped set operations and integrates interleaved authentication to prevent unauthorized falsification of subset information during forwarding.

Fu et al (2007) introduced a statistical approach to detect the clone node at base station. Keys are shared between neighbors to ensure authentication using Bloom's filter. Once the key exceeds the threshold, then the violating node is labeled as suspicious. Every node reports its keys to a base station and then the base station uses this approach to find cloned keys. The main challenge in this approach is high false negative and positive rates. Furthermore, honesty of the malicious nodes is uncertain while reporting their keys.

Crosby et al (2011) claimed the clone detection algorithm by detecting the duplicate location in distributed environment. Each node in the network uses an authenticated broadcast message to flood the network with its location information. Each node stores the location information of its neighbors and if it receives a conflicting claim, then the offending node is revoked. The drawback is that once the adversary node introduces the jamming attack, then authenticated broadcast message is not forwarded on the other part of the network. Node employs the redundant message or authenticated acknowledgment techniques, to prevent the clone attack. In this approach, each node stores location information about its '*d*' neighbors. A single node's location broadcast requires O(n) messages and the total communication cost for the protocol is O(n^2).

Parno et al (2005) introduced a method to detect the clone by broadcasting the location claim with a limited subset of deterministically chosen witness nodes. A node broadcasts its location along with claim message to its neighbors and forwards it to a subset of the nodes called witnesses. If the witness receives the claim from two different location claims with same node ID then the conflict on location claim becomes an evidence to trigger the revocation of the replicated node. In this approach, witness selection is based on deterministic scheme. The disadvantage is that if the adversary compromises all the witness nodes then unlimited replicas of a node can be created.

Zeng et al (2010) uses a random selection of witness node to detect the clone using location claim. Any node acts as a witness node to detect clones. A node announces its location and each of its neighbors sends a copy of the location claim to set of randomly selected witness nodes to detect the clone node. Since two set of witnesses exist in the path of replica node, it is not possible to detect the clone nodes. In a network of *n* nodes, if each location produces $\sqrt{n}$ witnesses, then the birthday paradox predicts at least one collision with high probability. The two conflicting locations claim form sufficient evidence is used to revoke the node, so the witness floods the pair of locations claims through the network, and each node can independently execute revocation procedure. The disadvantage of this scheme is high communication cost. Each neighbour has to send $O(\sqrt{n})$ claim message in order to detect the clone node.

Moniruzzaman et al (2009) introduced a Line-Selected Multicast (LSM) that selects witness by exploiting the routing topology of the network. Location claims of sensor nodes are transmitted through several intermediate nodes. If the intermediate nodes check replicas then clone is detected before the location claim reaches witness nodes. An intermediate node verifies the received claim using the signature of the claim. The received signature is then compared with the existing signature stored on the buffer at intermediate nodes. If the intermediate node finds the conflict in location information from the same ID then it floods the revocation message in the network. Each node chooses neighbors with probability *p* and *d* neighbors and forwards its location to '*g*' witness nodes through some intermediate nodes. In LSM, some nodes always have higher probability to act as witnesses and this weakens the detection itself. The attacker can take control of the node that will act as witness with highest probability. Furthermore, the protocol's overhead is not evenly distributed among the network nodes.

Zhu et al (2010) introduced protocols namely Single Deterministic Cell (SDC) and Parallel Multiple Probabilistic Cells (P-MPC) protocol to sensor network in the geographic grid as a unit, and each unit is named as cell. The protocol, uniquely maps the node ID to one of the cells in the grid. Each node broadcasts a location claim to its neighbor nodes by executing the detection procedure. Neighbor node forwards the location claim with a probability to a unique cell using a geographic hash function with the input of node's ID. After receiving the location claim by any one of the node in the destination cell, it floods the location claim to all node present in the cell. Then the destination cell stores the location claim with a probability. Therefore, the clone nodes are detected only with certain probability because the location claim is forwarded to the same cell. The number of destination cells is different in SDC and P-MPC. The location claim is forwarded to multiple deterministic cells with various probabilities by executing a geographic hash function with the input of node’s ID in P-MPC. The rest of procedure in P-MPC is similar to SDC. Hence, the clone nodes are detected with certain probability.

Conti et al (2011) introduced an algorithm named Random, Efficient and Distributed (RED), to forward the location claims of nodes to cluster heads. Cluster head randomly selects the witness nodes and forwards the location claim to witness node. This witness node verifies the signed claim and checks the coherency of the claim. If the claim is incoherent, then the network is flooded with clone detection message. The advantage of RED when compared to the other techniques is its higher detection probability. The drawback is that it requires more number of iterations to detect a clone in the network, because the intersection of the claim message is random. The other drawback is that, as any node in the network is selected as witness node, every node has the overhead of verifying the signature.

Hussain & Rahman (2009) suggested a mechanism to identify the replicated node using received signal strength. The RSSI value is measured at the receiver node to detect the node replication attack, as RSSI value is readily available for every message received by a node presented in the network. In this mechanism, communication overhead is high and it does not concentrate on energy efficiency.

Lou et al (2012) introduced a scheme namely the Single Hop Detection (SHD) protocol to detect the clone attack. This protocol is fully distributed, robust against node collusion that is designed for mobility based WSN. The drawback is every single hop neighbor needs to invoke the clone detection function.

Cho et al (2013) investigated the criteria for clone detection schemes by considering the type of device, detection methodology, deployment strategies, and detection of range. Simulation is conducted to compare the performance on grid-based deployment and this saves the energy by 94.44%. However, the error rate is ignored during mobility.

Transmitted data during the passive attacks such as eavesdropping, is the basic security issue. Data integrity prevents the compromised source nodes or aggregator nodes from significantly altering the final aggregation value.

2.6 SIMULATION TOOL

Automated Validation of Internet Security Protocol and Application (AVISPA) is a tool, which analyses the security goals of the protocol and its applications. It uses the descriptive programming in CASRUL (CAS+) and High Level Protocol Specification Language (HLPSL). CASRUL is a system for automatic verification of cryptographic protocols. It translates a protocol given in common abstract syntax into a rewrite system as HLSPL. This rewrite system can then be processed with a first order theorem to prove for equation logic for the automatic detection of flaws. This HLPSL specification is translated into the Intermediate Format (IF) a lower-level language than HLPSL and is read directly by the back-end of the AVISPA Tool. AVISPA comprises of four back-ends: On-the-fly Model-Checker (OFMC), CL-based Attack Searcher (CLAtSe), SAT-based Model-Checker (SATMC), and Tree Automata-based Protocol Analyser (TA4SP).

2.7 SUMMARY

As WSN is more prone to new kind of attacks, hence there is a need to develop efficient security algorithms. To protect WSN from attacks a lightweight security algorithm is desirable. In this chapter the security issues and existing security solutions is summarized.