Integration of False Data Detection with Data Aggregation and Confidential Transmission in Wireless Sensor Networks

S. Ozdemir   H. Cam
IEEE

IEEE/ACM Transactions on Networking, 2009

Presented by Gowun Jeong

Outline

Outline

Security Vulnerability of Wireless Sensor Networks

- Security attacks
    - False Data Injection (FDI)
        - Compromised nodes (CNs) decrease data integrity.
    - Data Forgery
    - Eavesdropping

- Where FDI by CNs possibly occurs?

    - False data injection (1.x)
    - Data forwarding (1.x)

- False data transmission depletes

    - the constrained battery power, and
    - the bandwidth resource.

Security Vulnerability of Wireless Sensor Networks

- Security attacks
    - False Data Injection (FDI)
        - Compromised nodes (CNs) decrease data integrity.
    - Data Forgery
    - Eavesdropping
- Where FDI by CNs possibly occurs?
    - Data Aggregation (DA)
    - Data Forwarding (DF)
- False data transmission depletes
    - the constrained battery power; and
    - the bandwidth utilisation.

Security Vulnerability of Wireless Sensor Networks

- Security attacks
    - False Data Injection (FDI)
        - Compromised nodes (CNs) decrease data integrity.
    - Data Forgery
    - Eavesdropping
- Where FDI by CNs possibly occurs?
    - Data Aggregation (DA)
    - Data Forwarding (DF)
- False data transmission depletes
    - the constrained battery power; and
    - the bandwidth utilisation.

Security Vulnerability of Wireless Sensor Networks

- Security attacks
  - False Data Injection (FDI)
    - Compromised nodes (CNs) decrease data integrity.
  - Data Forgery
  - Eavesdropping
- Where FDI by CNs possibly occurs?
  - Data Aggregation (DA)
  - Data Forwarding (DF)
- False data transmission depletes
  - the constrained battery power; and
  - the bandwidth utilisation.

False Data Detection (FDD)

- Conventional work
  - Most discussed FDD during DF.

        Challenge! Any data change between two
                   communicating endpoints is considered as
                   FDI.

- Ozdemir and Cam's approach
  - attempts to correctly determine whether any data alteration
    is due to DA or FDI.
  - A Data Aggregation and Authentication protocol
    - against up to T CNs
    - over the encrypted data
    - for FDD both by a data aggregator and by a non-aggregating
      node

False Data Detection (FDD)

- Conventional work
    - Most discussed FDD during DF.
        Challenge! Any data change between two
                   communicating endpoints is considered as
                   FDI.
- Ozdemir and Cam's approach
    - attempts to correctly determine whether any data alteration
      is due to DA or FDI.
    - A Data Aggregation and Authentication protocol
        - against up to T CNs
        - over the encrypted data
        - for FDD both by a data aggregator and by a non-aggregating
          node

False Data Detection (FDD)

- Conventional work
  - Most discussed FDD during DF.

    Challenge! Any data change between two communicating endpoints is considered as FDI.

- Ozdemir and Cam's approach
  - attempts to correctly determine whether any data alteration is due to DA or FDI.
  - A Data Aggregation and Authentication protocol
    - against up to T CNs
    - over the encrypted data
    - for FDD both by a data aggregator and by a non-aggregating node

False Data Detection (FDD)

- Conventional work
    - Most discussed FDD during DF.
        Challenge! Any data change between two communicating endpoints is considered as FDI.
- Ozdemir and Cam's approach
    - attempts to correctly determine whether any data alteration is due to DA or FDI.
    - A Data Aggregation and Authentication protocol
        - against up to T CNs
        - over the encrypted data
        - for FDD both by a data aggregator and by a non-aggregating node

Outline

Basic Assumptions

- Network
  - A densely deployed sensor network of certain large size
- Sensor
  - Overlapping sensing ranges
  - Role change
    - Sensor nodes rotatively assumes the role of data aggregator.
  - Limited computation and communication capabilities
- Message
  - Time-stamped
  - Nonce used to prevent reply attacks
- Intrusion ways to compromise nodes
  - Physical capturing
  - Radio communication channel attack

Network Topology

- Data aggregators are chosen in such a way that
    1. there are at least T nodes, called **forwarding nodes**, on the path between any two consecutive data aggregators; and
    2. each data aggregator has at least T neighbouring nodes.
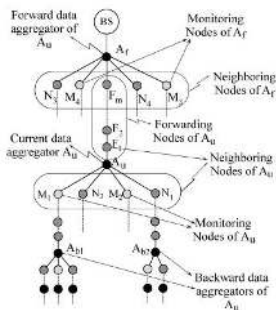
Generation of MACs

- Only data aggregators encrypt and decrypt the aggregated data.
- The forwarding nodes first verify data integrity using MACs and then relay the data if it is not false.
  - Two Full-size MACs (FMACs), each of which consisting of $T + 1$ subMACs, for a pair of plain and encrypted data
    - One computed by a data aggregator
    - T subMACs generated by its T monitoring nodes
  - The same Pseudo-Random Number Generator (PRNG), termed f
    - Random numbers between 1 and 32

Generation of MACs

- subMAC generation of data *D* by neighbouring node N_i of data aggregator A_u for its pairmate F_j
    1. Establish the shared key $K_{i,j}$ between N_i and F_j.
    2. Compute MAC(*D*) using $K_{i,j}$.
    3. Assuming that *S* denotes the size of MAC(*D*) in bits, selects $S/(T+1)$ bits to form subMAC(*D*) using its PRNG and $K_{i,j}$ as the seed.

- subMAC verification of *D* by F_j for its pairmate N_i
    1. Compute the MAC(*D*).
    2. Run its PRNG $S/(T+1)$ times to generate subMAC(*D*) with $K_{i,j}$ as the seed.
    3. Compare two subMAC(*D*)'s.

- PRNG synchronisation achieved by packet sequence numbers

Key Establishment

- Pairwise key establishment
    - Sybil attacks
        - A compromised node fakes multiple identities to establish pair relations with more than one monitoring nodes.
    - To prevent from Sybil attacks, a monitoring node can share a pairwise key with another node in multiple hops.
- Group key establishment
    - Group key K_{group}^u for data aggregator A_u and its neighbouring nodes is used to select the monitoring nodes and to protect data confidentiality while data transmitting.

Limitations

- The value of *T* depends strictly on several factors, such as geographical area conditions, modes of deployment, and so on.

- The pairwise key establishment between non-neighbouring nodes takes more time than that between direct neighbouring nodes.

- Compromising only one legitimate group member discloses not only some or all of the past group keys but also the current group key.

Outline

Introduction

Assumptions and Limitations

Data Aggregation and Authentication Protocol (DAA)
   Step 1: Monitoring Node Selection for an Aggregator
   Step 2: Sensor Node Pairing
   Step 3: Integration of Secure Data Aggregation and False
   Data Detection

Performance Analysis

Conclusion

Notations used in DAA

TABLE I
SUMMARY OF NOTATIONS

Notation	Explanation
A_u	Current data aggregator.
A_f	Forward data aggregator.
A_b	Backward data aggregator.
BS	Base Station.
N_i	Neighboring node i of A_u or A_f.
F_j	Forwarding node j of A_u.
M_k	Monitoring node k of A_u.
K_{group}^u	Group key of A_u and its neighbors.
$K_{i,j}$	Shared key between sensor nodes i and j.
$E_{K_{ij}}(D)$	Encryption of data D with key K_{ij}.
$MAC_{K_{ij}}(D)$	Message Authentication Code of data D calculated with key K_{ij}.
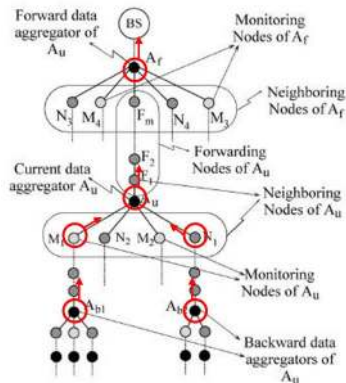
Outline

Introduction

Assumptions and Limitations

Data Aggregation and Authentication Protocol (DAA)
Step 1: Monitoring Node Selection for an Aggregator
Step 2: Sensor Node Pairing
Step 3: Integration of Secure Data Aggregation and False
Data Detection

Performance Analysis

Conclusion

Algorithm MNS (Monitoring Node Selection)

Table: Choose T monitoring nodes from n neighbouring nodes of A_u

1.	$A_u \Rightarrow$ all nodes	request two random numbers with node ID
2.	$N_i \rightarrow A_u$	R_a and R_b generated by $f(K_{u,i})$
		$\text{MAC}_{K_{u,i}}(R_a \mid R_b)$
3.	$A_u \Rightarrow$ all nodes	$\{N_1, \ldots, N_n\}$ in the receiving order
		$\{R_1, \ldots, R_{2n}\}$ labeled in an ascending order
		$\text{MAC}_{K_{group}^u}(R_1 \mid \cdots \mid R_{2n})$
4-1.	$N_i \rightarrow A_u$	(**verified**)$E_{K_{u,i}}(\text{MAC}_{K_{group}^u}(R_1 \mid \cdots \mid R_{2n}))$
4-2.	$N_i \rightarrow A_u, N_j$'s	(**unverified**)restart from 1.
5.	N_i	for $1 \leq k \leq T$, compute
		$I_k = [(\sum_{j=k}^{n-1+k} R_j + K_{group}^u) mod(n)] + 1$
		to determine T monitoring node ID's of A_u

Outline

Introduction

Assumptions and Limitations

Data Aggregation and Authentication Protocol (DAA)
   Step 1: Monitoring Node Selection for an Aggregator
   Step 2: Sensor Node Pairing
   Step 3: Integration of Secure Data Aggregation and False
   Data Detection

Performance Analysis

Conclusion

Three Types of Node Pairs

- $2T + 1$ node pairs are formed.
  AA-type pair One pair between A_u and A_f
  MF-type pair T pairs between M_k of A_u
                and F_j towards A_f
  MN-type pair T pairs between M_k of A_u
                and N_i of A_f

- T M_k's selected in Step 1 distinctly choose
  their own pairmates to form MF-type and
  MN-type pairs.
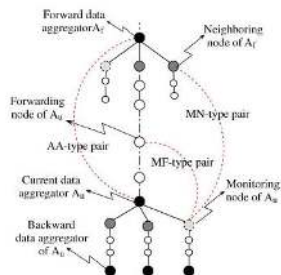
Pairmate Selection

1.	$A_f \rightarrow F_j \rightarrow A_u$	pairmate discovery message
		N_i's of A_f
		$\text{MAC}_{K_{f,u}}(N_i\text{'s})$
		F_j's IDs for $1 \leq j \leq h$
2.	$A_u \Rightarrow T\ M_k$'s	$\text{MAC}_{K_{group}^u}(F_1 \mid \cdots \mid F_h)$ for new, random
		forwarding node labeling
		$\text{MAC}_{K_{group}^u}(N_i\text{'s})$s
3.	$M_k \rightarrow A_u$	one forwarding node
		one neighbouring node
4.	$A_u \Rightarrow T\ M_k$'s	two pairmate lists of size T
5.	M_k	pairmate verification

Outline
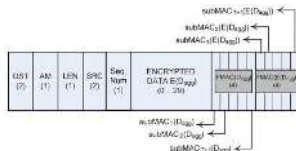
Data Confidentiality

- One pairmate computes a subMAC, and the other pairmate verifies the subMAC.
- subMACs for plain data are used for FDD during DA.
- subMACs for encrypted data are used for FDD during DF.
- Each data aggregator forms two FMACs as the following figure.

- A_u determines the order of subMACs and informs each forwarding node about its subMAC location individually.
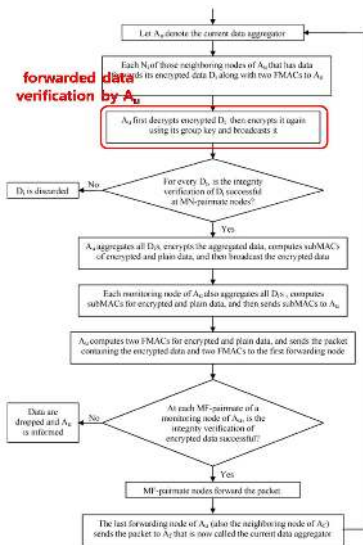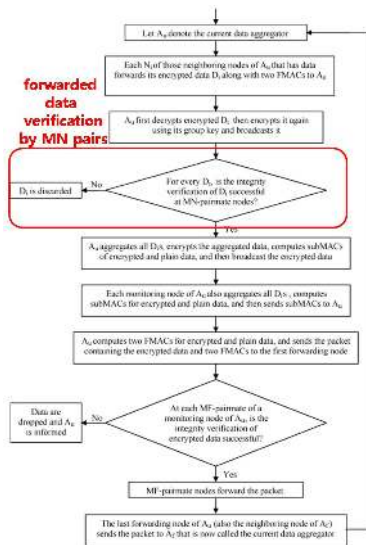 - probability of FDI at a forwarding node $= (1/2)^{32}$

Algorithm SDFC
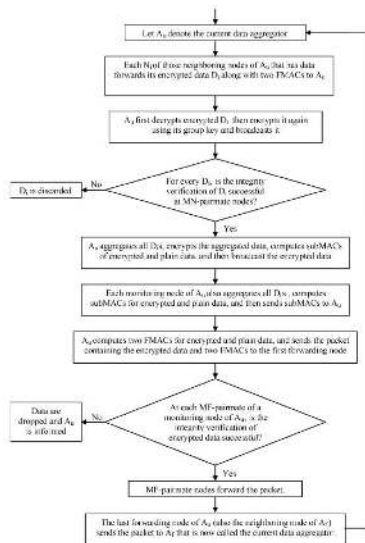
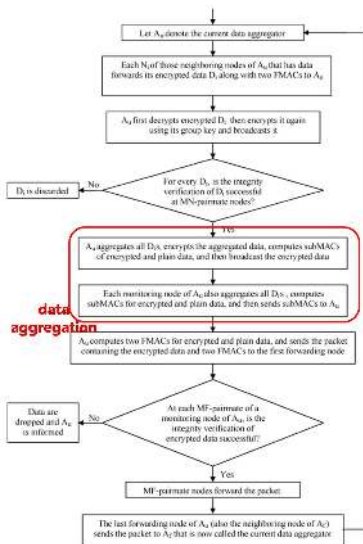# **Algorithm** SDFC

Algorithm SDFC

Algorithm SDFC
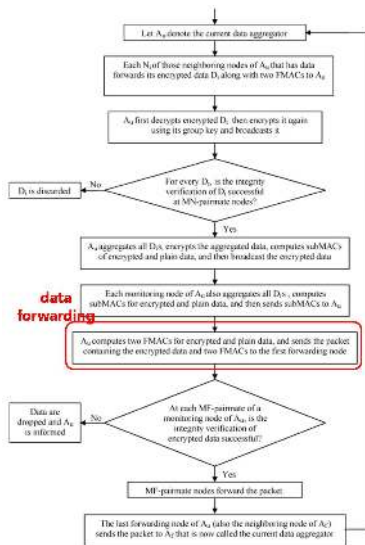
Algorithm SDFC
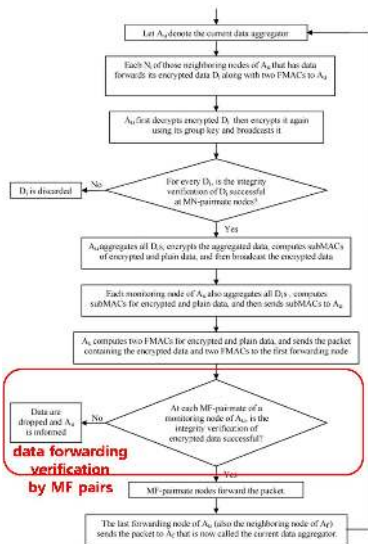
Algorithm SDFC

Algorithm SDFC

Algorithm SDFC
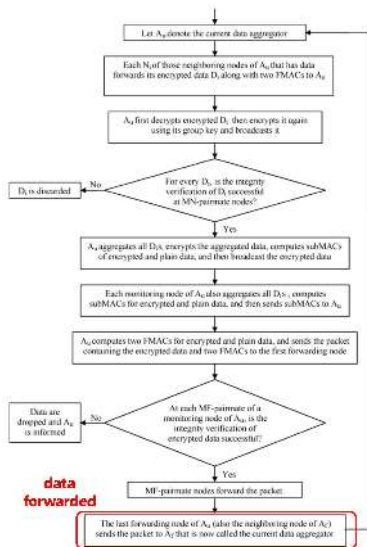
Outline

Security Analysis of Algorithm SDFC

Lemma 1
Assuming that A_u is compromised and there are additional at most $T - 1$ collaborating compromised nodes among the neighbouring nodes of A_u and A_f, any false data injected by A_u are detected by the A_f's neighbouring nodes only in SDFC.

- *Data verification by the monitoring nodes of A_u and the neighbouring nodes of A_f*

Lemma 2
Assuming that A_u and A_f are not compromised, any false data injected by any subset of A_u's forwarding nodes are detected by A_f in SDFC.

- *Data verification by A_f*

Security Analysis of FMAC and subMAC

- Changing the size of MAC
  - Security Level vs. Communication Overhead
- Probability of FDI at a node $= (1/2)^{32}$ for 4-byte FMACs
  - Probability of FDI into a subMAC $= (1/2)^{32/(T+1)}$
  - The size of FMAC $= T + 1$

Security Analysis of FMAC and subMAC

- Changing the size of MAC
  - Security Level vs. Communication Overhead
- Probability of FDI at a node $= (1/2)^{32}$ for 4-byte FMACs
  - Probability of FDI into a subMAC $= (1/2)^{32/(T+1)}$
  - The size of FMAC $= T + 1$

Computational Cost of Algorithm SDFC

Computation	Traditional Work	SDFC
MAC	1	$4(T + 1)$ $= (T + 1)$ subMACs $\times$ 2 FMACs $\times$ a pair
Aggregation	1	$T + 1$ $= 1$ by aggregator $+ \ T$ by monitors
Encryption/ **Decryption**	1	$T + 2$ $= 1$ encryption by A_u $+ \ T$ decryptions by monitors $+ \ 1$ decryption by A_f

- Only the first MAC computation consumes much resource.
- Data transmission requires much more energy than data computing in wireless sensor networks.

Communication Cost of Algorithm SDFC

D_{ADD}	the amount (in bytes) of data transmission using ADD of two FMACs
$D_{tradAuth}$	the amount (in bytes) of data transmission using the traditional scheme of a MAC
L_{tos}	the length (in bytes) of an authenticated and encrypted data packet
α	the number of data packets generated by legitimate nodes
β	the number of false data packets injected by up to T compromised nodes
H_d	the average number of hops between two consecutive data aggregators
H	the average number of hops that a data packet travels in the network

$$
\begin{aligned}
D_{ADD} &= (L_{tos} + 4)(\alpha H + \beta H_d) + T(L_{tos} + 4)(\alpha + \beta) + \frac{4T}{T+1}(\alpha + \beta) \\
D_{tradAuth} &= L_{tos} H(\alpha + \beta)
\end{aligned}
$$

- data transmission by a data aggregator

- data transmission by T monitors

- subMACs transmission by T monitors

Communication Cost of Algorithm SDFC

D_{ADD}	the amount (in bytes) of data transmission using ADD of two FMACs
$D_{tradAuth}$	the amount (in bytes) of data transmission using the traditional scheme of a MAC
L_{tos}	the length (in bytes) of an authenticated and encrypted data packet
α	the number of data packets generated by legitimate nodes
β	the number of false data packets injected by up to T compromised nodes
H_d	the average number of hops between two consecutive data aggregators
H	the average number of hops that a data packet travels in the network

$$
\begin{aligned}
D_{ADD} &= (L_{tos} + 4)(\alpha H + \beta H_d) + T(L_{tos} + 4)(\alpha + \beta) + \frac{4T}{T+1}(\alpha + \beta) \\
D_{tradAuth} &= L_{tos} H(\alpha + \beta)
\end{aligned}
$$

- data transmission by a data aggregator

- data transmission by T monitors

- subMACs transmission by T monitors

Communication Cost of Algorithm SDFC

D_{ADD}	the amount (in bytes) of data transmission using ADD of two FMACs
$D_{tradAuth}$	the amount (in bytes) of data transmission using the traditional scheme of a MAC
L_{tos}	the length (in bytes) of an authenticated and encrypted data packet
α	the number of data packets generated by legitimate nodes
β	the number of false data packets injected by up to T compromised nodes
H_d	the average number of hops between two consecutive data aggregators
H	the average number of hops that a data packet travels in the network

$$
\begin{aligned}
D_{ADD} &= (L_{tos} + 4)(\alpha H + \beta H_d) + T(L_{tos} + 4)(\alpha + \beta) + \frac{4T}{T+1}(\alpha + \beta) \\
D_{tradAuth} &= L_{tos}H(\alpha + \beta)
\end{aligned}
$$

- data transmission by a data aggregator

- data transmission by T monitors

- subMACs transmission by T monitors

Communication Cost of Algorithm SDFC

D_{ADD}	the amount (in bytes) of data transmission using ADD of two FMACs
$D_{tradAuth}$	the amount (in bytes) of data transmission using the traditional scheme of a MAC
L_{tos}	the length (in bytes) of an authenticated and encrypted data packet
α	the number of data packets generated by legitimate nodes
β	the number of false data packets injected by up to T compromised nodes
H_d	the average number of hops between two consecutive data aggregators
H	the average number of hops that a data packet travels in the network

$$
\begin{aligned}
D_{ADD} &= (L_{tos} + 4)(\alpha H + \beta H_d) + T(L_{tos} + 4)(\alpha + \beta) + \frac{4T}{T+1}(\alpha + \beta) \\
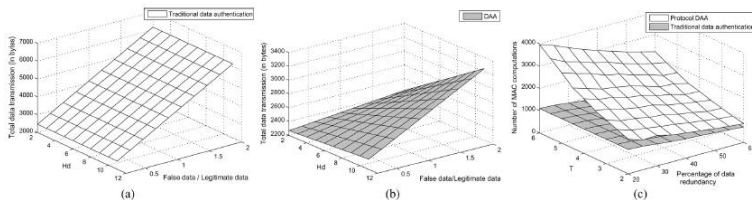D_{tradAuth} &= L_{tos} H(\alpha + \beta)
\end{aligned}
$$

- data transmission by a data aggregator

- data transmission by T monitors

- subMACs transmission by T monitors

$$\circ\circ$$
$$\circ\circ\circ$$
$$\circ\circ\circ\circ\circ\circ\circ\circ\circ\circ$$

Cost Comparison

- $L_{tos} = 41, H = 50, H_d \leq 12$ and $\beta/\alpha \geq 0.2$

(a) (b) (c)

- Comparing (a) and (b), D_{ADD} more mildly increases than $D_{tradAuth}$.
- (c) shows that the value of T trades off between security and computation overhead in the network.
- (c) also illustrates the impact of data aggregation.

Outline

Introduction

Assumptions and Limitations

Data Aggregation and Authentication Protocol (DAA)
Step 1: Monitoring Node Selection for an Aggregator
Step 2: Sensor Node Pairing
Step 3: Integration of Secure Data Aggregation and False
Data Detection

Performance Analysis

Conclusion

Contributions and Future Work

- Contributions
    - False data detection during data aggregation
    - Integration of data confidentiality and false data detection
    - Less communication overhead (by fixing the size of each FMAC)
- Future work
    - Security and efficiency improvement in networks where every sensor enables data forwarding and aggregation at the same time