Security and Privacy in Large-Scale RFID Systems

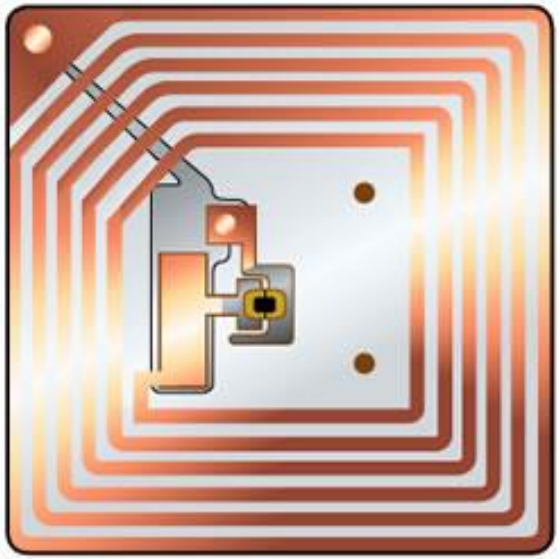
Min-Te Sun, Ph.D.
Professor
Computer Science and Information Engineering
National Central University
September 27, 2017

Outline

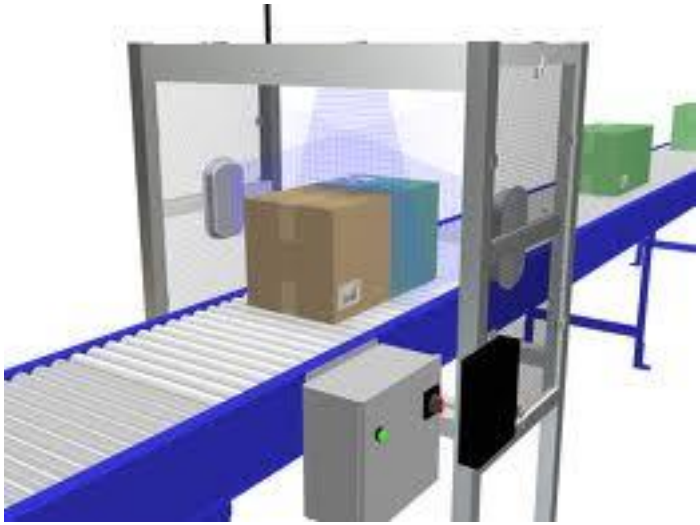
1. Background of RFID Systems
    1. Introduction to RFID Technologies
    2. Security and Privacy Issues in RFID Systems
2. Encryption-Based Private Authentication
3. Non-Encryption-Based Private Authentication
4. Conclusion

Radio Frequency Identification

- **RFID** (Radio Frequency Identification)
  - Is an electronic identification technology
  - Consists of an RF reader and RF tags
  - Tags are attached to objects
  - A reader automatically identify all objects by reading tags



An RF tag



An RF reader

RFID v.s. Barcode Technologies

RFID	Barcode
No line-of-sight	Labels must be seen by a reader
Long read range (~1m)	Short read range
Automatic singulation	Labels are scanned individually
Read and write capability (~512 bits)	No write operation

Applications

- Applications
  - Supermarkets
  - Supply chain managements
  - Book stores or library
  - Natural habitat monitoring
  - Transportation payment
  - Smart cards

Passive RF Tags

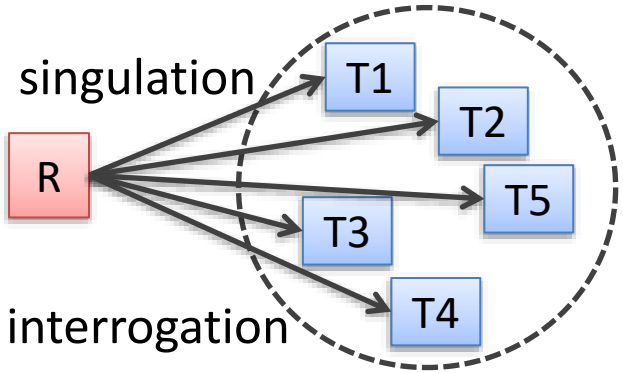
- There are two kinds of tags, active and **passive tags**
  - Active tags are more like sensors
- Passive tags
  - Have no power supply, and a tag is energized by signal
  - Are computationally weak devices
  - Are very cheap ($0.1 in 2011)

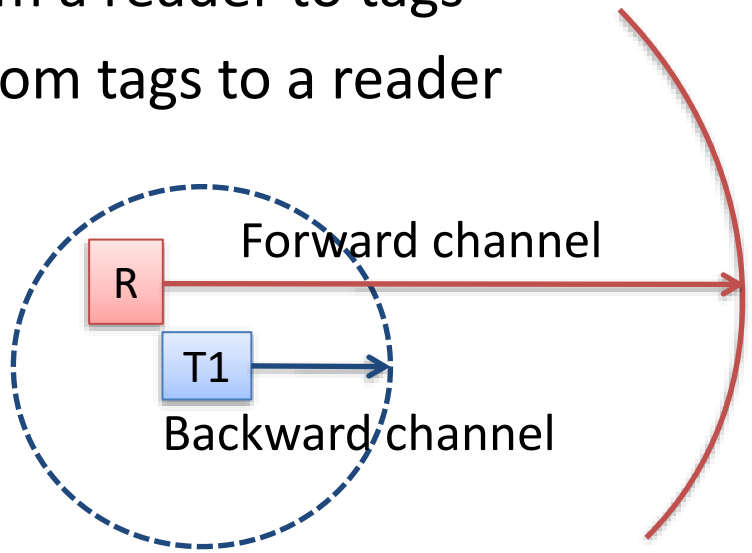
Frequency	868-956 MHz.
Memory	512 bits
Transmission range	~ 1 meter
ID length	96 bits
Passwords	32 bits

Functions
XOR, concatenation, 16-bit pseudo random generator, Collision resistance hash function, etc.

Objects Identification

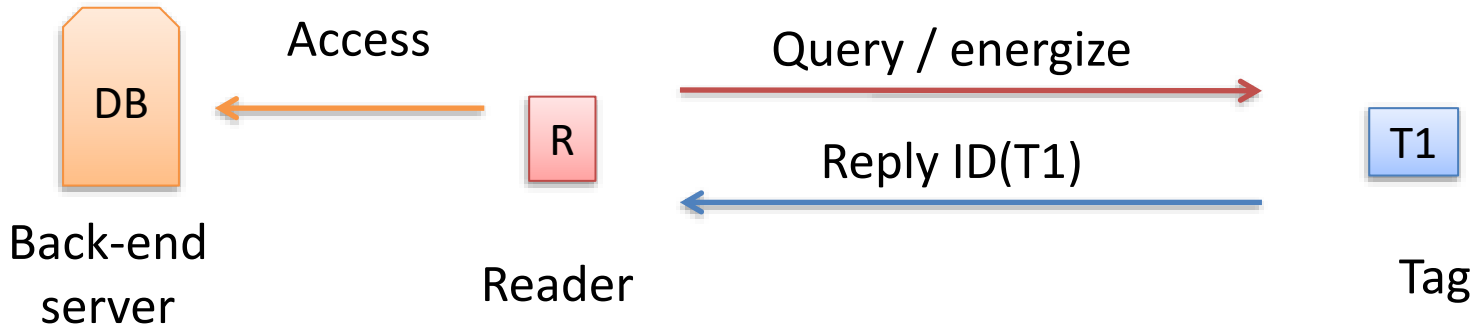
- Terms and definitions
  - **Singulation** – the process by which a reader identifies individual tags
  - **Interrogation** – the cycle by which a reader identify all tags in its reading region
- Singulation by the query-and-response
  - **Forward channel** – the signal from a reader to tags
  - **Backward channel** – the signal from tags to a reader





Object Identification (Cont.)

- An RF reader is connected to the back-end server
- A tag's ID is used as a **pointer** to the data entry in the server
  - Database contains objects' information
  - Or object status (e.g., Object 1 is at LA, Chicago, NY)

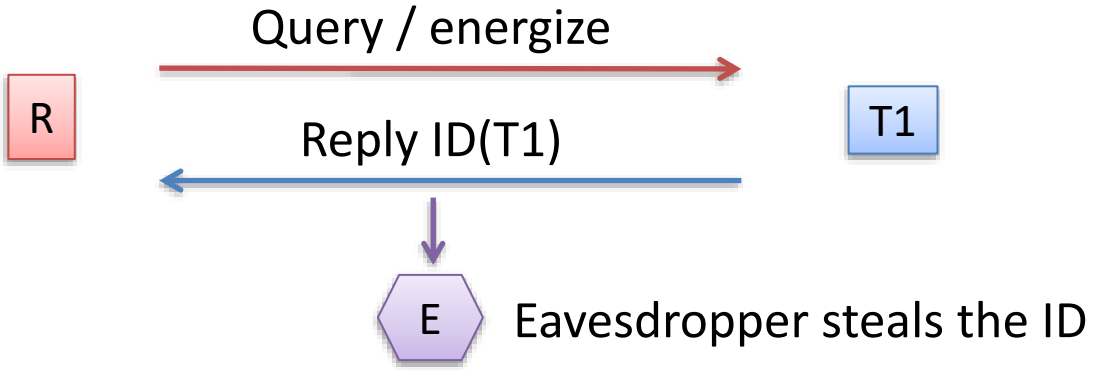


Private Authentication Problem

- Tag's ID itself is **private** information
- During a singulation process, tags' ID must be protected from adversaries

Private tag authentication problem

An RF reader securely accesses tags without disclosing tags' content to any third parties (e.g., eavesdroppers)



Private Authentication Problem

- **Encryption-based** approach
  - Used for large-scale RFID systems
    - e.g., Inventory management such as book store
  - A secret key is assigned to each tag before deployment
  - Low-cost cryptographic operations are assumed
- **Non-encryption-based** approach
  - Used for the RFID systems, in which common secrets are not possible
    - e.g., smart cards, toll collections, etc.
    - Public/private key operation is not possible
  - Relies on the physical layer technologies, e.g., jamming

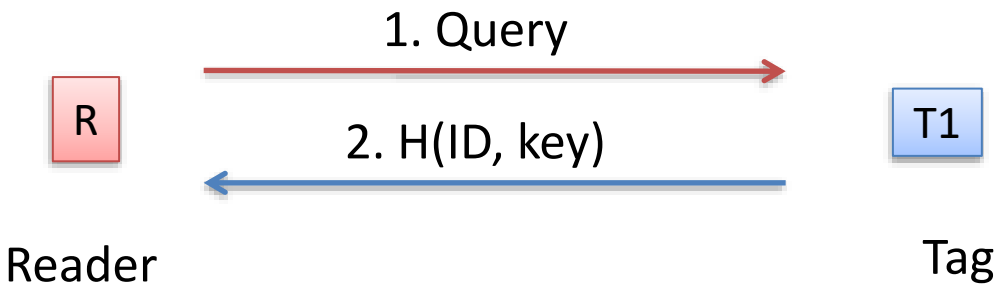
Proposed Work in This Tutorial

- Private authentication
  - Encryption-based authentication protocol
  - Two Non-encryption-based authentication protocols
- System Architecture
  - Trusted Masking Device (non-encryption-based authentication)
  - Distributed RFID sensing (non-encryption-based authentication)

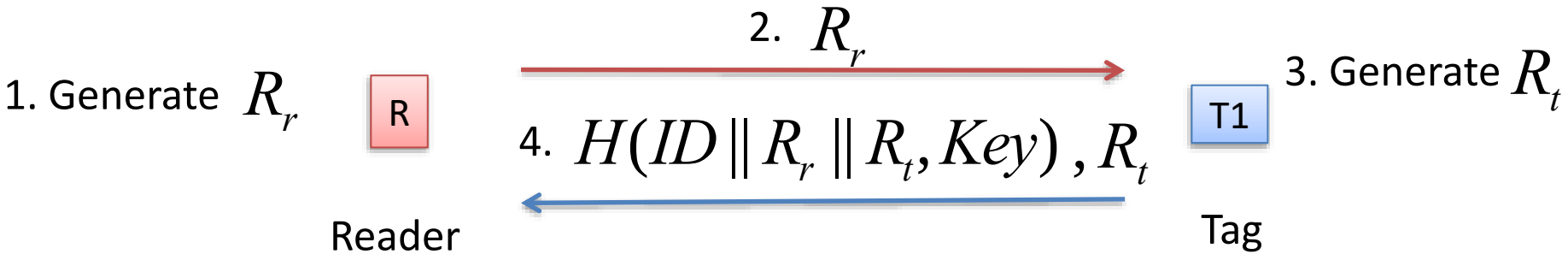
1. Background of RFID Systems
    1. Introduction to RFID Technologies
    2. Security and Privacy Issues in RFID Systems
2. Encryption-Based Private Authentication
3. Non-Encryption-Based Private Authentication
4. Conclusion

- The system consists of **N** tags
  - A secret key is assigned to each tag in the system
  - Each tag has its **ID** and **key**
  - The server has (ID, key) for all tags
- Goals
  - Protect tag's content from adversaries
  - An authentication protocol should satisfy security requirements
  - High performance in term of authentication speed

- A naive approach
  - 1. A reader sends a query
  - 2. A tag replies with **Hash(ID, key)**
  - 3. A reader scan all keys to find the tag (ID', key') s.t. **H(ID', key')** = **H(ID, key)**
- **Not secure** for some attacks
- Poor performance, **O(N)**

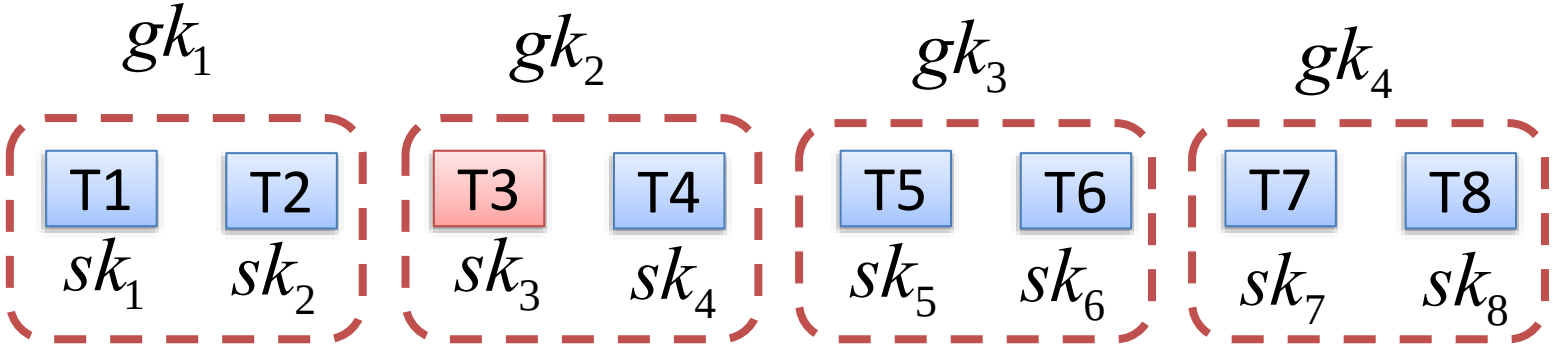


- **Privacy**
  - The tag's content must be protected
  - Privacy of tags can be protected by the use of secret keys
  - Reply = $H(ID, Key)$
- **Untraceability**
  - An attacker cannot trace a tag from its replies
  - A solution is the **use of nonce**, i.e., a random number, R
  - Reply = $H(ID \| R_r \| R_t, Key)$ and R_t

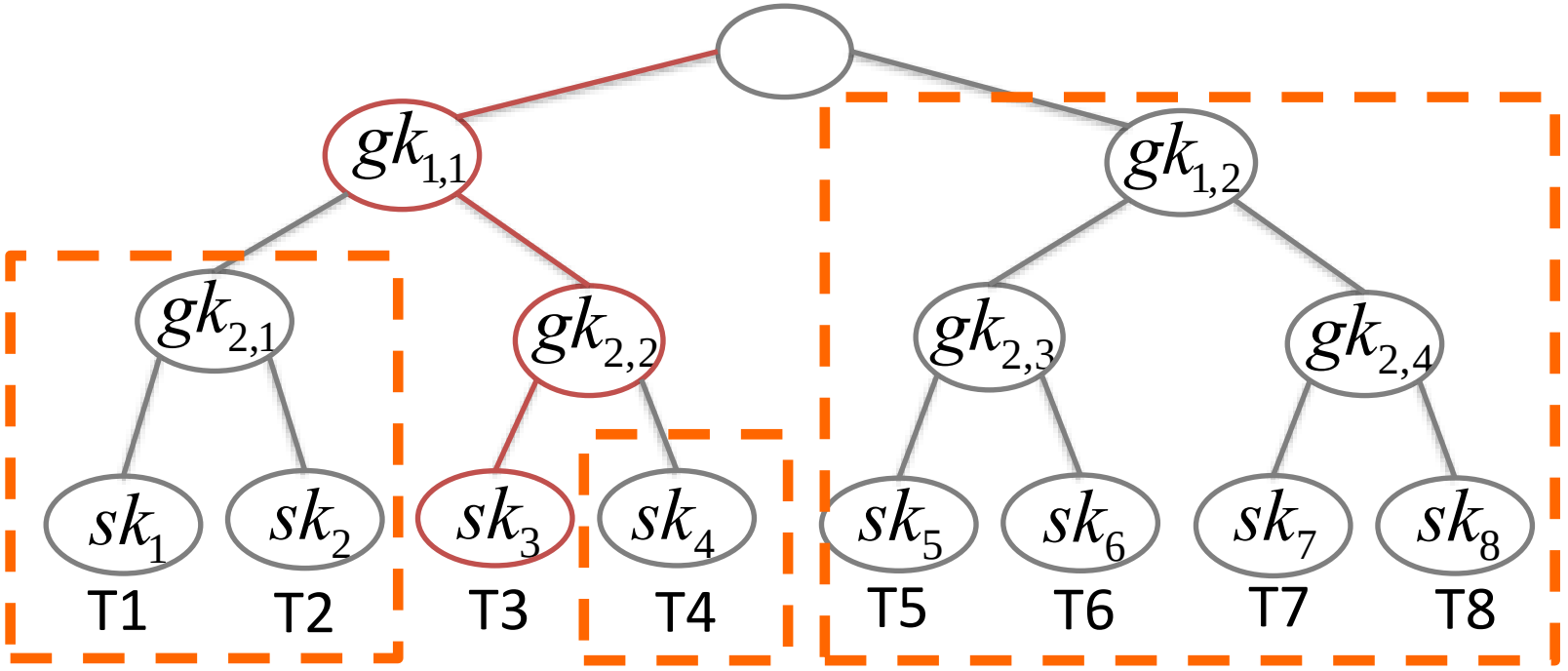


- **Cloning attack resistance**
  - An attacker cannot counterfeit a legitimate tag by cloning a tag's reply
  - The use of nonce avoid cloning attacks
- **Forward security**
  - An attacker cannot obtains information in the previous communications by the key of compromised tags
  - **Key updating** mechanism must be addressed
    - $H(ID \| R_1, Key_1)$ and R_1' $Key_2 = H(Key_1, R_1)$
    - $H(ID \| R_2, Key_2)$ and R_2' $Key_3 = H(Key_2, R_2)$

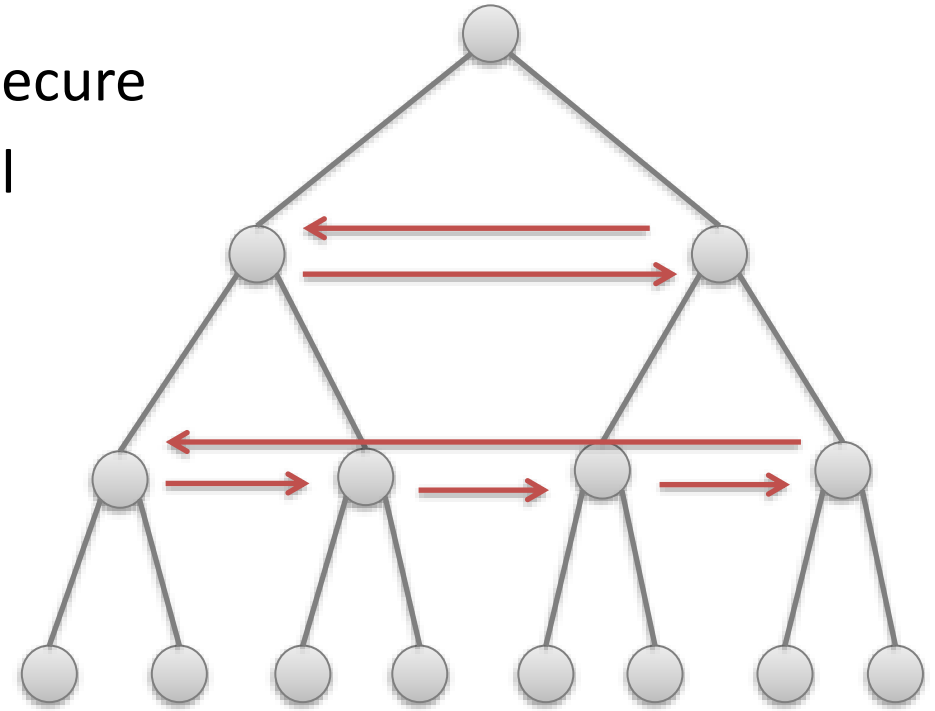
- Unstructured
  - Reply = $H(ID \| R_r \| R_t, Key)$ and R_t
    - Where R_r and R_t are random numbers
  - This approach must scan all keys in the server
  - So, it is **very slow**, **O(N)** where N is the number of tags
- Protocols with a structured key management
  - A set of shared keys and a unique key are assigned to tag,
  - There are **group-based** and **tree-based** protocols



- **Compromise Attacks**
  - Should tags be physically compromised, an adversary obtains all keys from the compromised tags
  - Other tags are divided into disjoint sets (**anonymous sets**)
    - T1 is identified by 1/2, T5 is identified by 1/4 , etc.

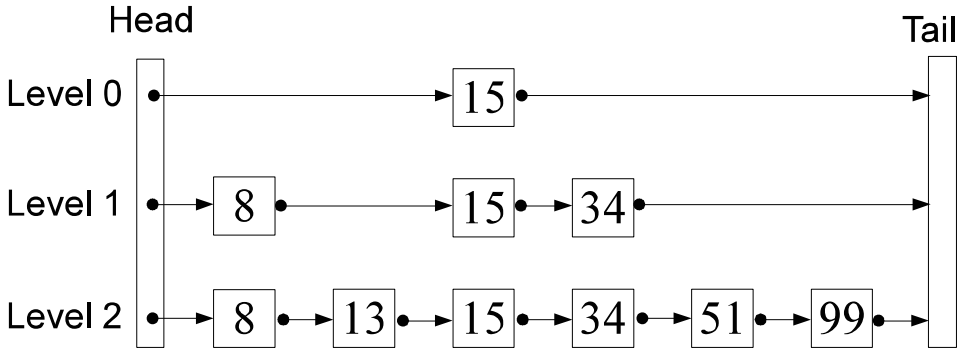


- Design goals
  - A protocol must provide strong protection against **compromise attacks** in keeping with **high performance**
  - There is **tradeoff** between security and performance
- Basic ideas
  - Tree-based is fast, but not secure
  - A random **shift** at each level
  - **dependency** among levels



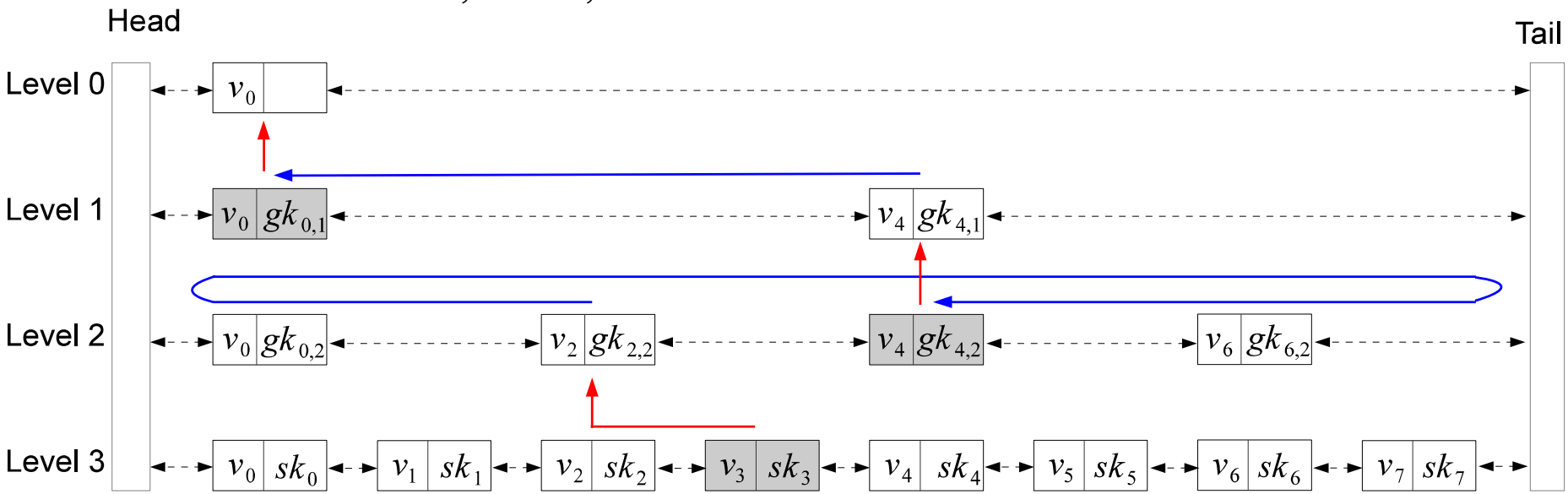
- We proposed a **skip lists-based** protocol
  - **Randomized Skip Lists-Based Authentication (RSLA)**
  - It is as **fast** as the tree-based, and **more secure**

- 4 components
  - Key initialization
  - Authentication
  - Key update
  - System maintenance
    - Tags can join to/leave from the system



- A skip list is generated **deterministically**
- Tags are assigned to the nodes in the bottom list
- Keys on the path from the bottom to the top list are assigned to a tag
- e.g., Tag 3 has **a set of keys** and **random numbers** for shifting

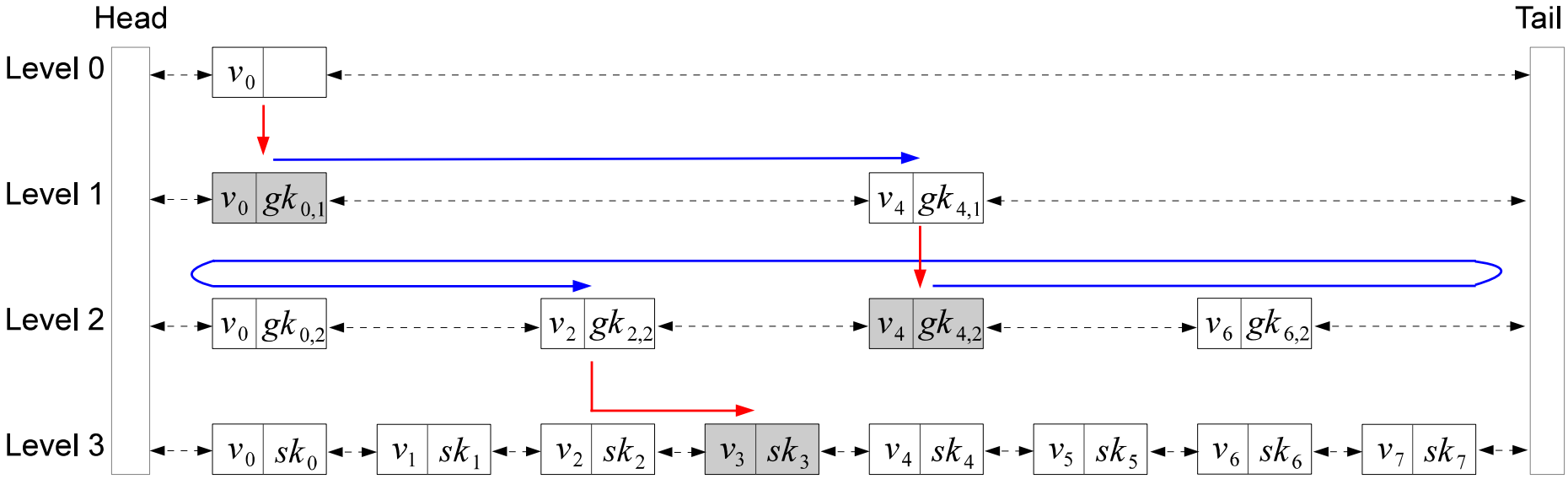
$$gk_{0,1}, gk_{4,2}, sk_3 \quad R = \{3,1\}$$



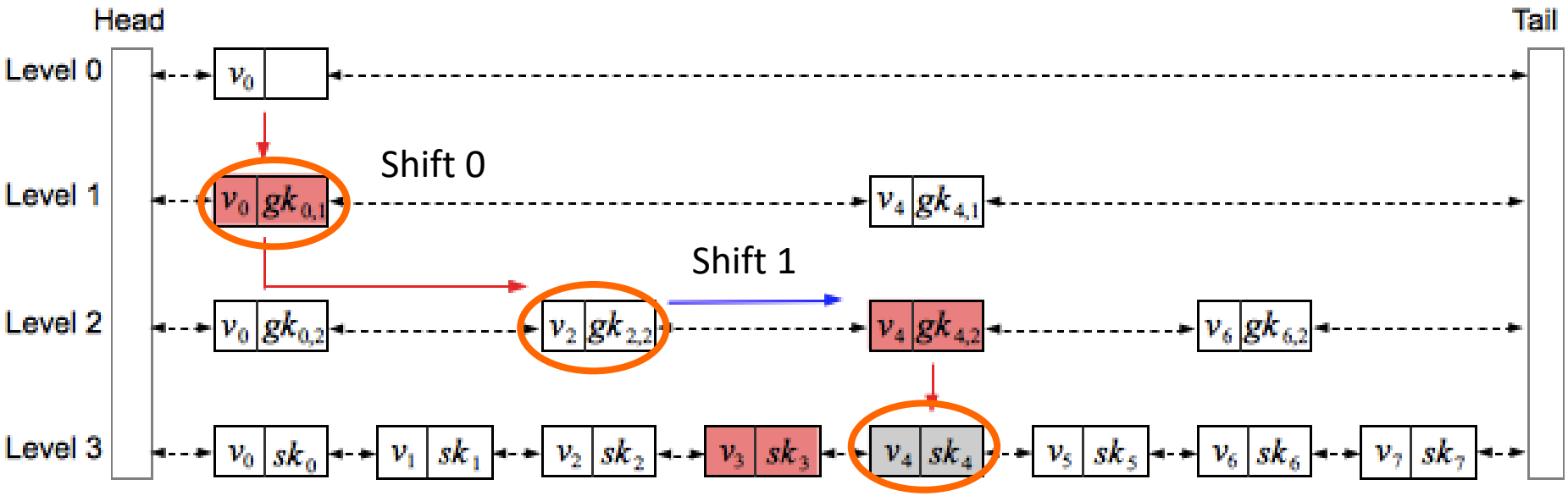
- For each level i, $b_i = H(gk_{i,ji}, b_{i-1} \| n_r \| n_t), E(gk_i, R_i)$
- A tag's reply consists of $n_t, b = \{b_1, b_2, \ldots, b_{\log N}\}$
- e.g., Tag 3 replies with $b_1 = H(gk_{0,1}, f \| n_r \| n_t), E(gk_{0,1}, 1)$

$$b_2 = H(gk_{4,2}, b_1 \| n_r \| n_t), E(gk_{4,2}, 3)$$

$$b_3 = H(sk_3, b_2 \| n_r \| n_t), E(sk_3, n_r \| n_t)\ n_t$$



- Assume Tag 3 is compromised $gk_{0,1}, gk_{4,2}, sk_3$
- Another tag belongs to an anonymous set with size (N – 1) unless it has the all group keys in common
  - Dependency among levels and random shifting
- e.g., Tag 4 has $gk_{0,1}, gk_{2,2}, sk_4 \quad R = \{0,1\}$
  - Tag 4 belongs to an anonymous set size (N – 1)



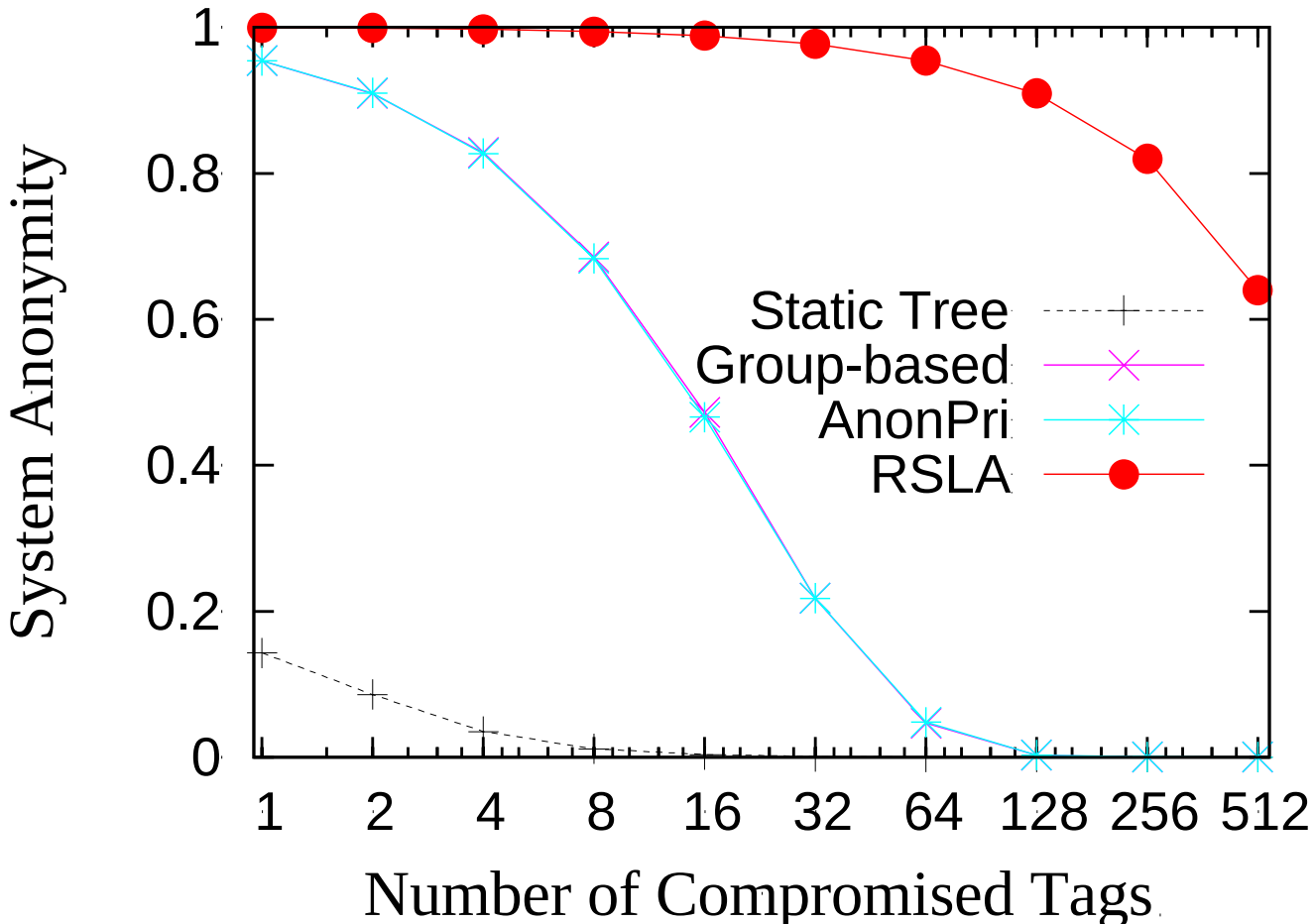
- Key update
  - The system updates the entire skip lists
  - Each node has a new key and the old key
  - Keys at tags are updated when they are interrogated
- System maintenance
  - A new tag joins to the system
    - A tag is assigned to a leaf
    - When the skip lists is full, a new set of skip lists is created
  - A tag leaves from the system
    - The corresponding leaf node is deleted

- The proposed skip lists-based is **fast** and **secure**

	Unstructured	Tree-based	Group-based	Skip Lists-Based
Running time	$O(N)$	$O(\log N)$	$O(N/t)$	$O(\log N)$
Key cost	$O(N)$	$O(N)$	$O(N+N/t)$	$O(N)$
Security	Good	Very poor	Poor	Good

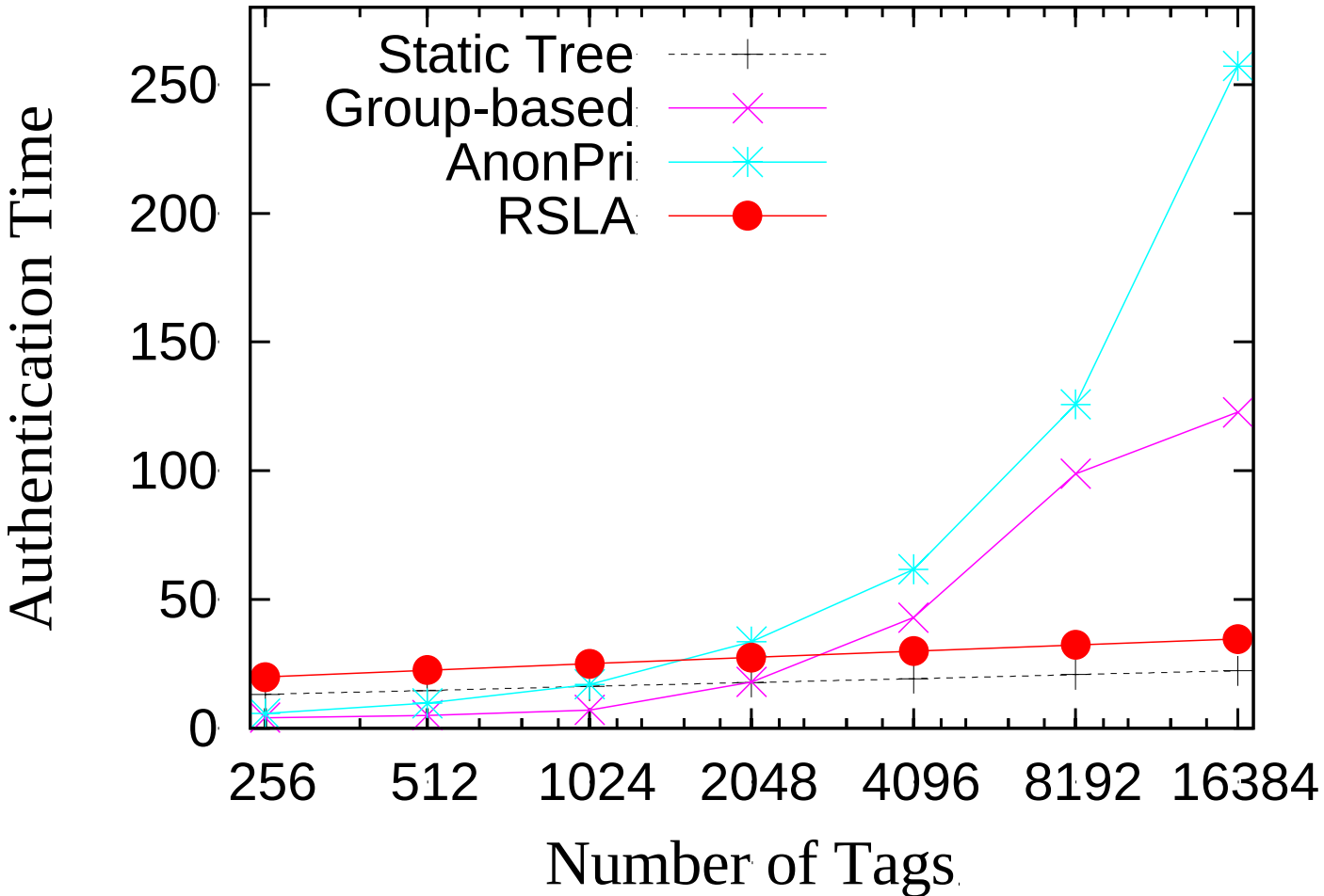
N is the number of tags in the system
t is the number of groups

- RSLA v.s. existing solutions (Tree-based, Group-based, and AnonPri (group-based)
- Anonymity of the system against compromised attacks



Simulation Results (Cont.)

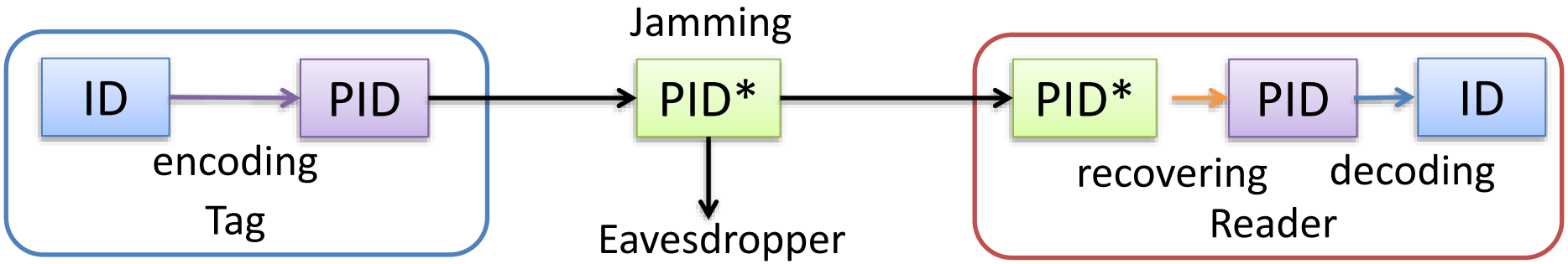
- Time required for an RF reader to authenticate tags in the system



1. Background of RFID Systems
    1. Introduction to RFID Technologies
    2. Security and Privacy Issues in RFID Systems
2. Encryption-Based Private Authentication
3. Non-Encryption-Based Private Authentication
4. Conclusion

Non-Encryption-Based

- In some applications, shared secrets are **not possible**
  - e.g., transport payment, smart cards
- Tags cannot perform public/private key operations
- One way to protect tags' reply is use of **Jamming**
- Baseline
  - Tag encodes its ID to a pseudo ID (PID)
  - Jamming is conducted during the data transmission
  - A reader recovers corrupted PID, and decodes it



Security Issues in RFID Backward Channel

- Most of the research have focused on forward channel protection
- Only two solutions have been proposed for the backward channel protection
  - Privacy Masking
  - Randomized Bit Encoding

Privacy Masking

- Privacy masking, (Choi and RohI, ICCSA'06)
  - A reader sends mask bits when a tag sends ID
  - The reader can recover with the mask even if some of bits of tag ID collide

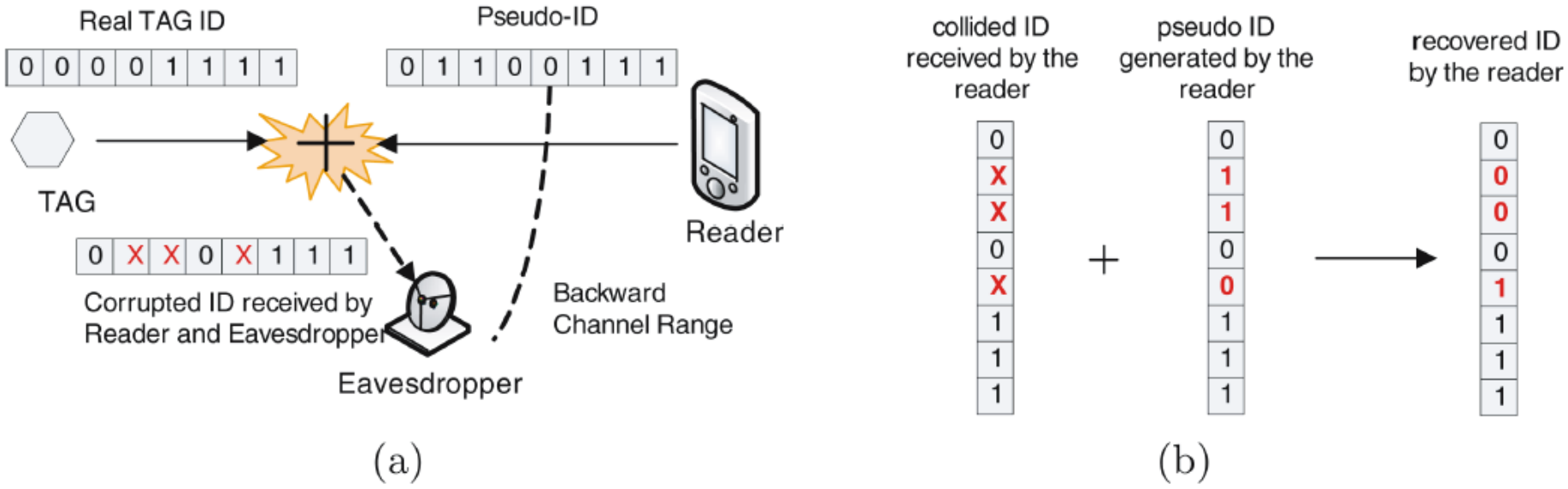


Fig. 2. Example operation of proposed method (a) collision (b) recovery

Issues of Privacy Masking

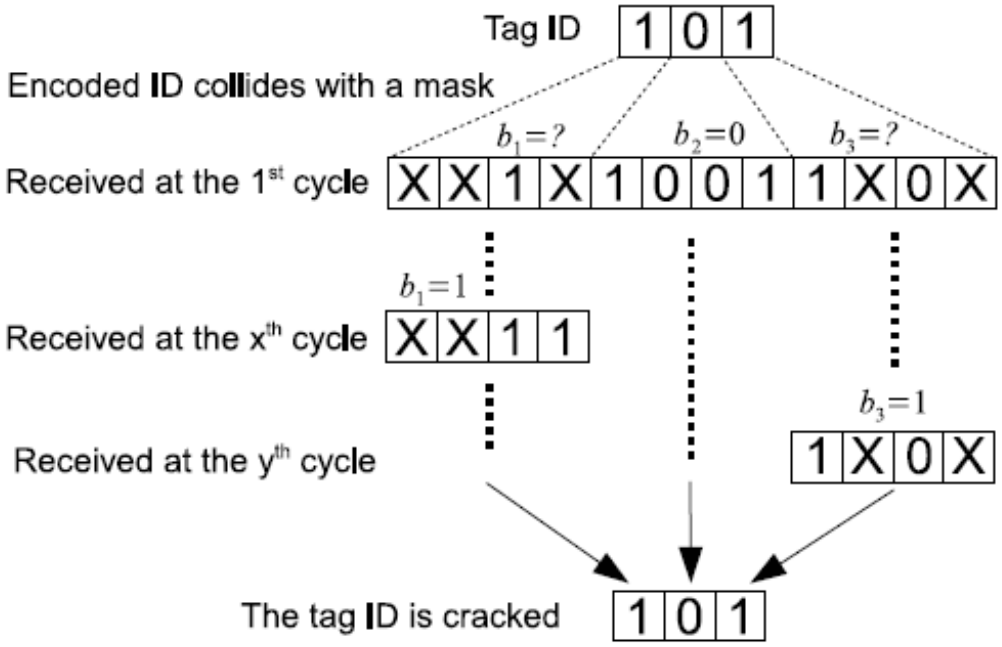
- Each bit has 50% of chance to be recovered
  - A higher level of protection is required
- Attackers can create their own “unprotected” reader!
  - The backward channel protection is completely cracked.

Randomized Bit Encoding (RBE)

- To alleviate the same bit problem, Lim et al (Lim, Li, and Yeo PerCom 08) proposed Randomized Bit Encoding (RBE)
- The idea is that **an encoded ID is transmitted in privacy masking environment**
  - **Each source bit is encoded into a codeword**
  - A tag sends pseudo ID
  - If a source bit is "0", the hamming weight of codeword is even, otherwise odd
    - Example. Source bit (the real tag ID) is "0101"
      - "0" -> "00", "1" -> "01", "0" -> "11", "1" -> "10"
      - The pseudo ID is "00011110"
  - An authorized reader recovers pseudo ID, and then identifies the real ID
    - **We assume a reader needs to know a source bit and the corresponding codeword**
    - Example, ID is "00011110", and mask is "10000110"
      - Received ID is "X00XX110",
      - Reader gets "0101", but an eavesdropper gets "XXX1"

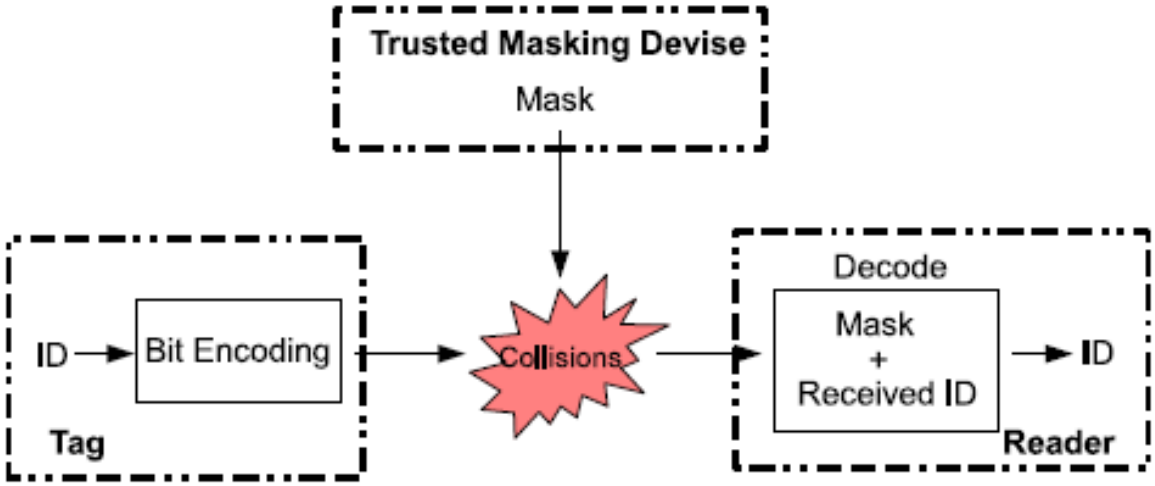
More Issues of RBE

- It is vulnerable to the **correlation attack**
  - Each source bit is independently encoded
  - An eavesdropper may listen to a channel for a long period of time
  - This attack works for both Privacy Masking and RBE!



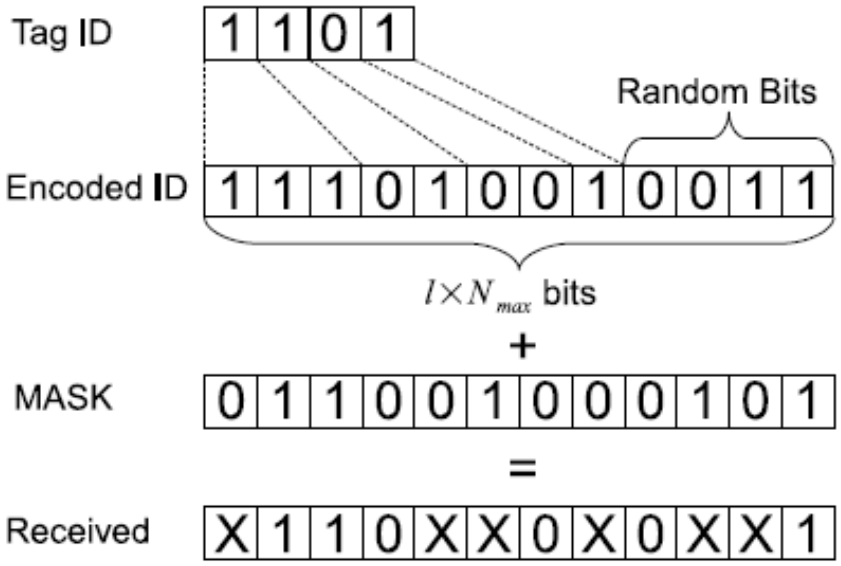
New System Architecture (To Eliminate Unprotected Readers)

- A reader queries tags
- Tags sends its pseudo ID under the masking environment
- At the same time, Trusted Masking Device (TMD) sends mask bits
- A secure channel is established between reader and TMD
- The reader can recover pseudo ID and obtain the real tag IDs



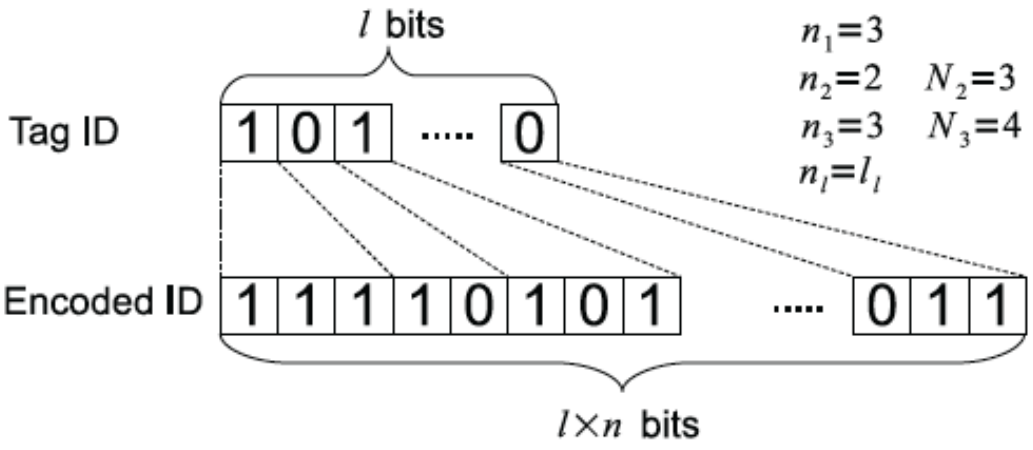
Dynamic Bit Encoding

- The idea is that **the codeword length is changed dynamically**
  - The first codeword length is N_{max}
  - The codeword length of the i-th bit is F(key), where F() is a hash function w/ value <= N_{max}
  - Random bits are inserted at the end to make the pseudo ID length l x N_{max}
- To identify i-th bit, an attacker needs (i-1)-th codeword
- Example,
  - N_{max} = 3, l = 4
  - F(key) = key mod N_{max} + 1
  - Key is the prev. codeword
    - e.g. F(111) = 2



Optimized DBE

- The **Optimal Dynamic Bit Encoding (ODBE)** is proposed to further improve performance of DBE
  - Length of i-th codeword is F(key) = key mod N_i + 1, where $N_i = n \cdot i - \sum_{k=1}^{i-1} n_k$.
  - The last codeword length is $l_l = n \cdot l - \sum_{k=1}^{l-1} n_k$
- Example
  - N_{max} = 3, i-length ID



Analysis

- **The correct guess probability** is the prob. that an eavesdropper successfully guesses the original ID from received pseudo ID

Lower bound is P = $(1/2)^l$

No encoding

$$P = \left\{\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{2}\right\}^l = \left(\frac{3}{4}\right)^l$$

RBE

$$P(n) = \left\{\frac{1}{2^n} + (1-\frac{1}{2^n})\cdot\frac{1}{2}\right\}^l, \quad (n \geq 1)$$

DBE

$$P(N_{max}) = (1-\frac{1}{2^{N_{max}}})\cdot(\frac{1}{2})^l + \sum_{i=1}^{l}(\frac{1}{2^{N_{max}}})\cdot(\overline{n})^{i-1}\cdot\{1-(\overline{n})^{l-i}\}\cdot(\frac{1}{2})^{l-i}$$

ODBE

$$P(n) = (1-\frac{1}{2^n})\cdot(\frac{1}{2})^l + \sum_{i=1}^{l}(\frac{1}{2^n})\cdot(\frac{1}{2^{\overline{N}}})^{i-1}\cdot\{1-(\frac{1}{2^{\overline{N}}})^{l-i}\}\cdot(\frac{1}{2})^{l-i}$$

The Correct Guess Probability

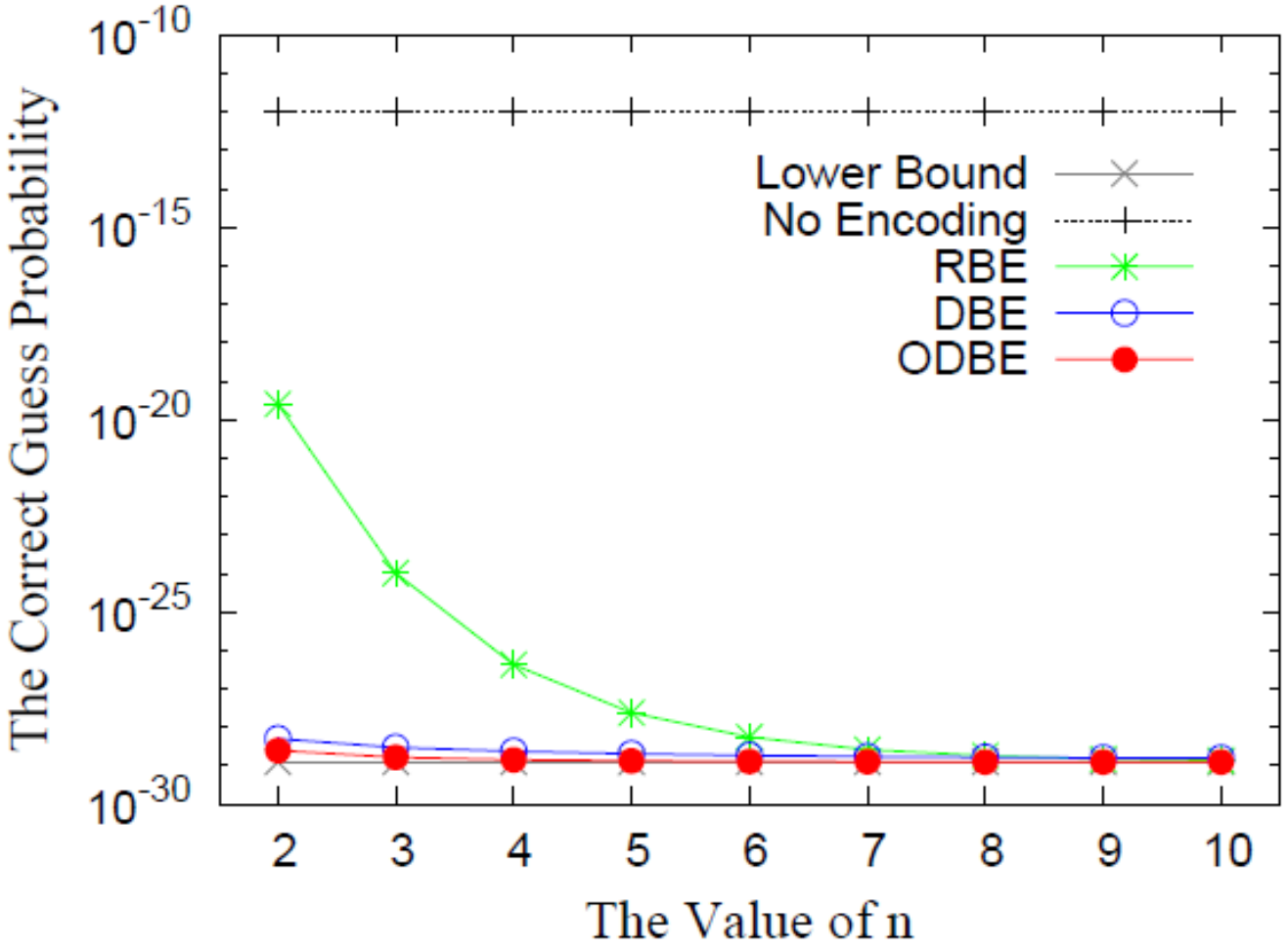


Fig. 6. The correct guess.

Results

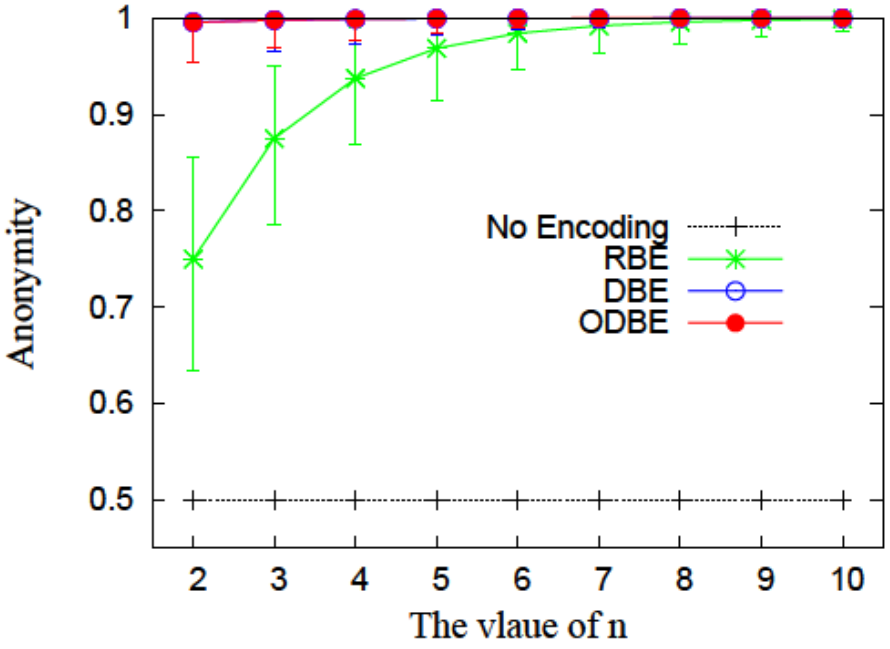


Fig. 7. Anonymity.

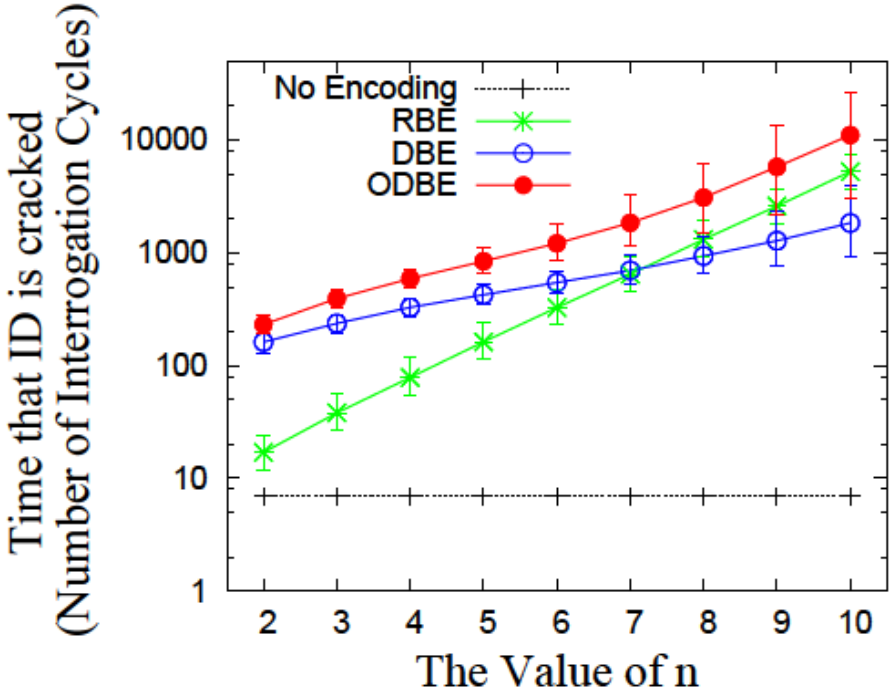
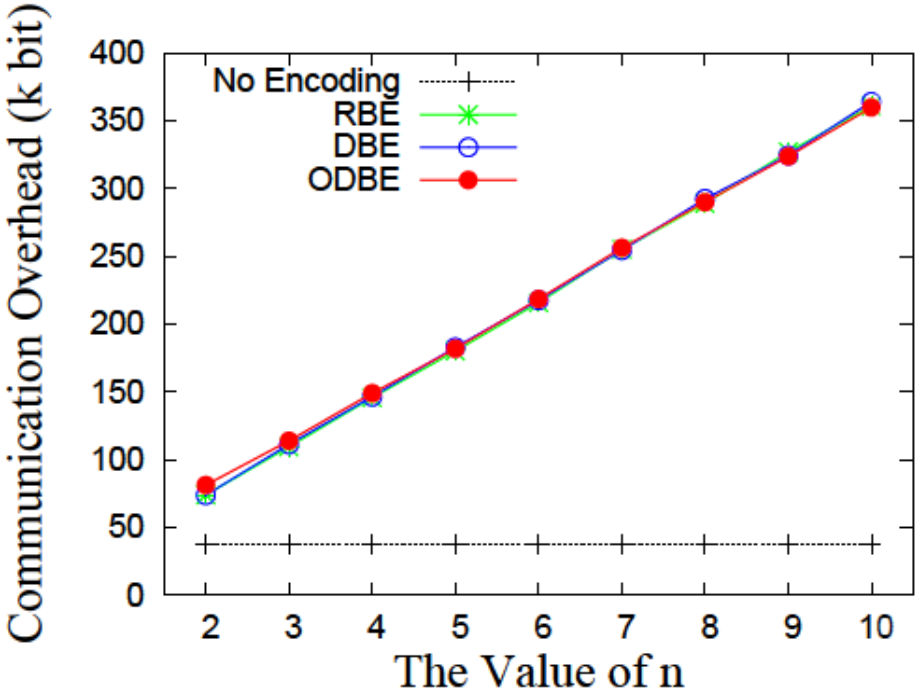


Fig. 8. Time to crack IDs.

Results (Cont.)



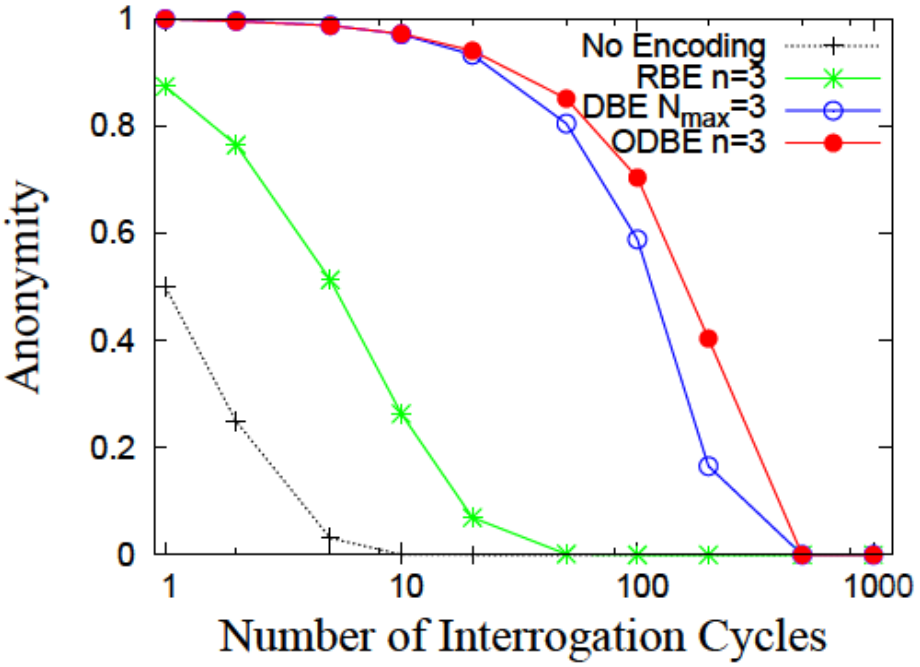
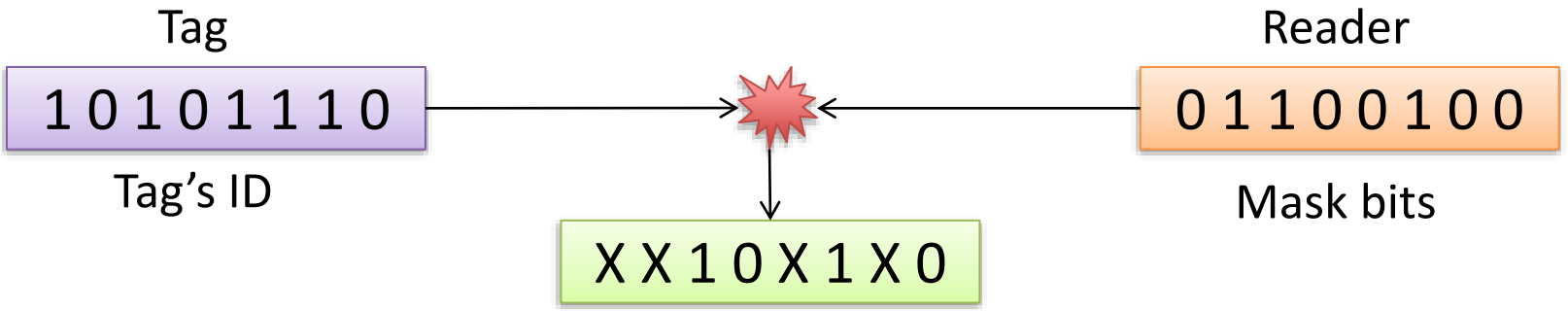


Fig. 9. Communication overhead.

Issue of DBE and ODBE

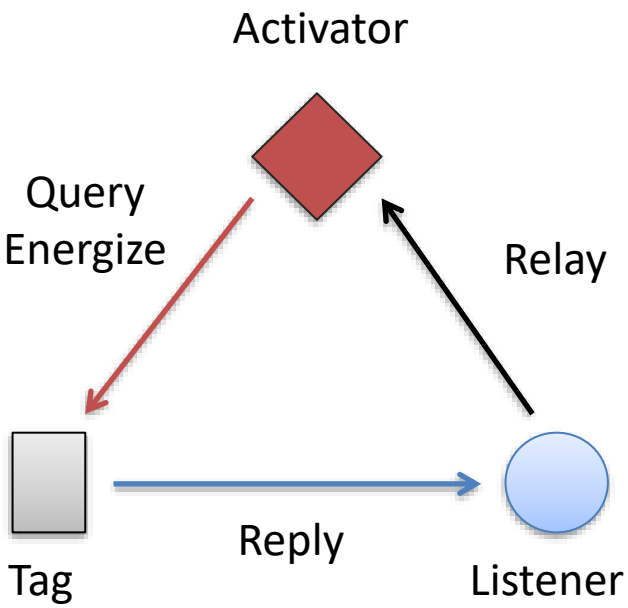
- Unrealistic physical layer assumptions
  - Tag's reply and the mask must be **perfectly** synchronized
  - The channel is assumed to be **additive**
    - i.e., 0 + 0 = 0, 1 + 0 = X, 0 + 1 = X, 1 + 1 = 1
  - Only deal with backward channel

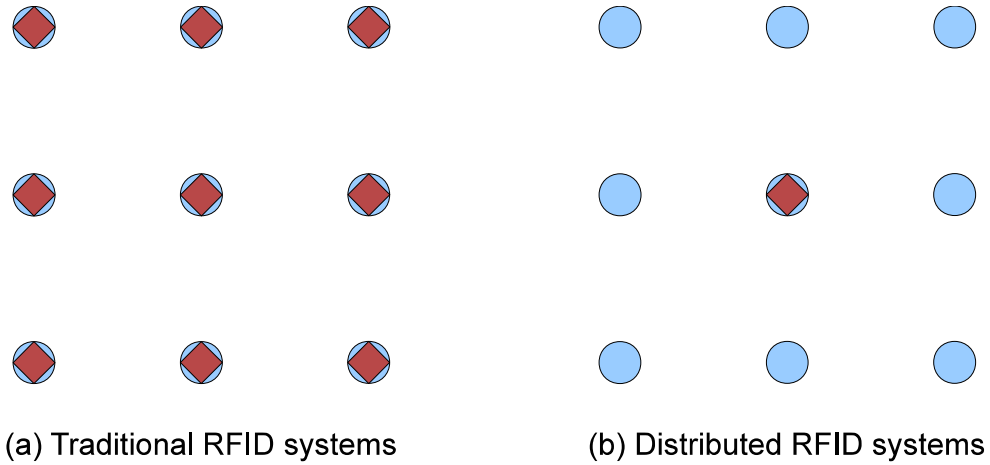


A reader and an eavesdropper receives a corrupted ID

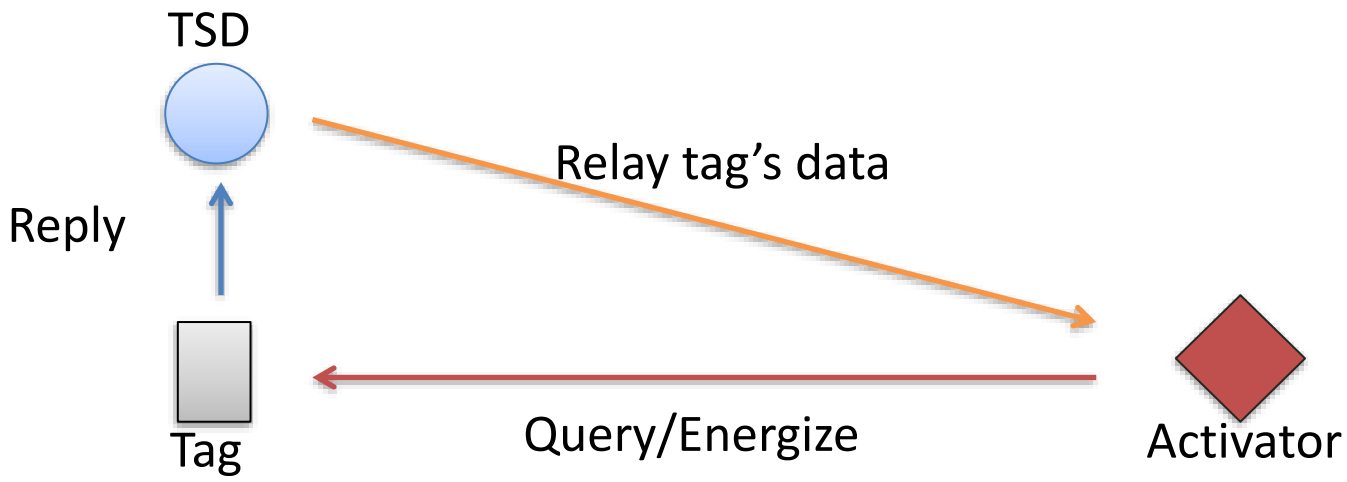
- Existing solutions has 3 components
  - A system architecture, a jamming model and a encoding scheme
- Proposed non-encryption-based authentication
  1. We applied the **distributed RFID architecture**
  2. We redesigned **a jamming model**
  3. We have developed **a new coding scheme** that achieves perfect secrecy [Wire-Tap, 1975]

- A RF reader is divided into two components
  - An **RF activator** and **RF listeners** [Mobicom 10]
  - The forward channel is long, the backward channel is short



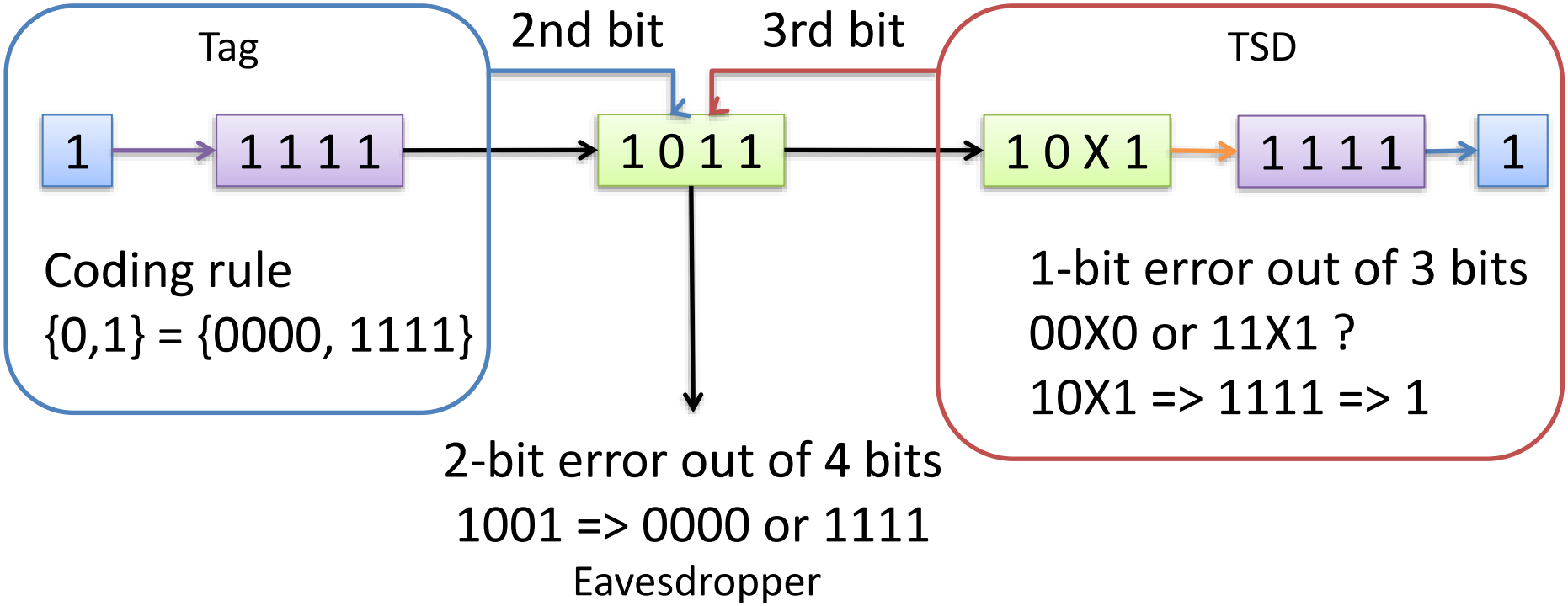


- System architecture
  - An RF activator queries tags
  - An RF tag replies its ID to a **TSD (trusted shield device)**
    - A TSD (RF listener) could be implemented in smart phones, etc.
    - A listener is located at user-side
  - The listener relays data to the activator
    - The traditional communication link

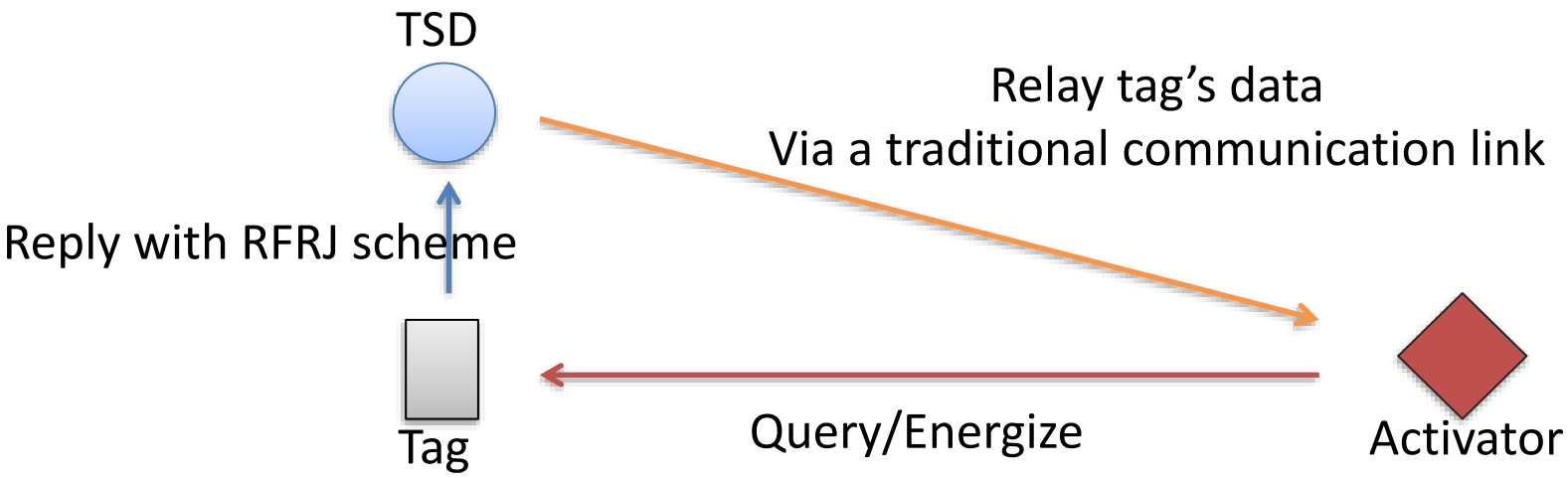


- Jamming Model Assumption
  - **Probabilistic jamming model**
    - A bit is flipped with a given probability when jamming is conducted
  - **Bit level jamming** is assumed
- TSD conducts jamming when a tag replies
  - **Full-duplex** mode is assumed [Mobihoc 12]
    - A node can send signal and receive signal simultaneously

- Proposed the 1-to-4 bit coding
  - A tag randomly **flips** one bit in a codeword
  - A TSD randomly **jams** one bit in a codeword
- The index of flipped or jammed bit is **secret**

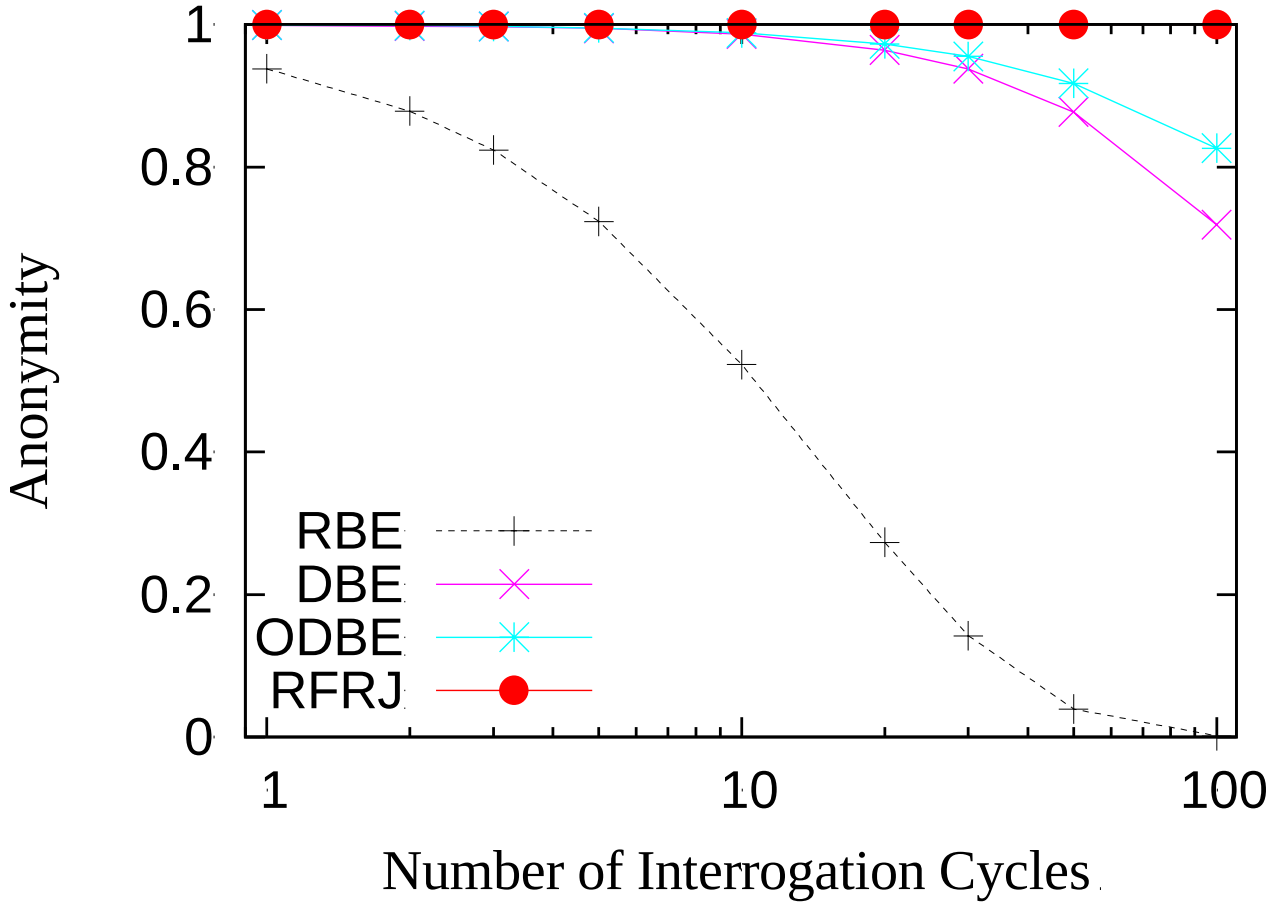


- We proposed **Random Flipping and Random Jamming (RFRJ)** private authentication protocol
  - Distributed architecture
  - A new jamming model
  - A new coding scheme

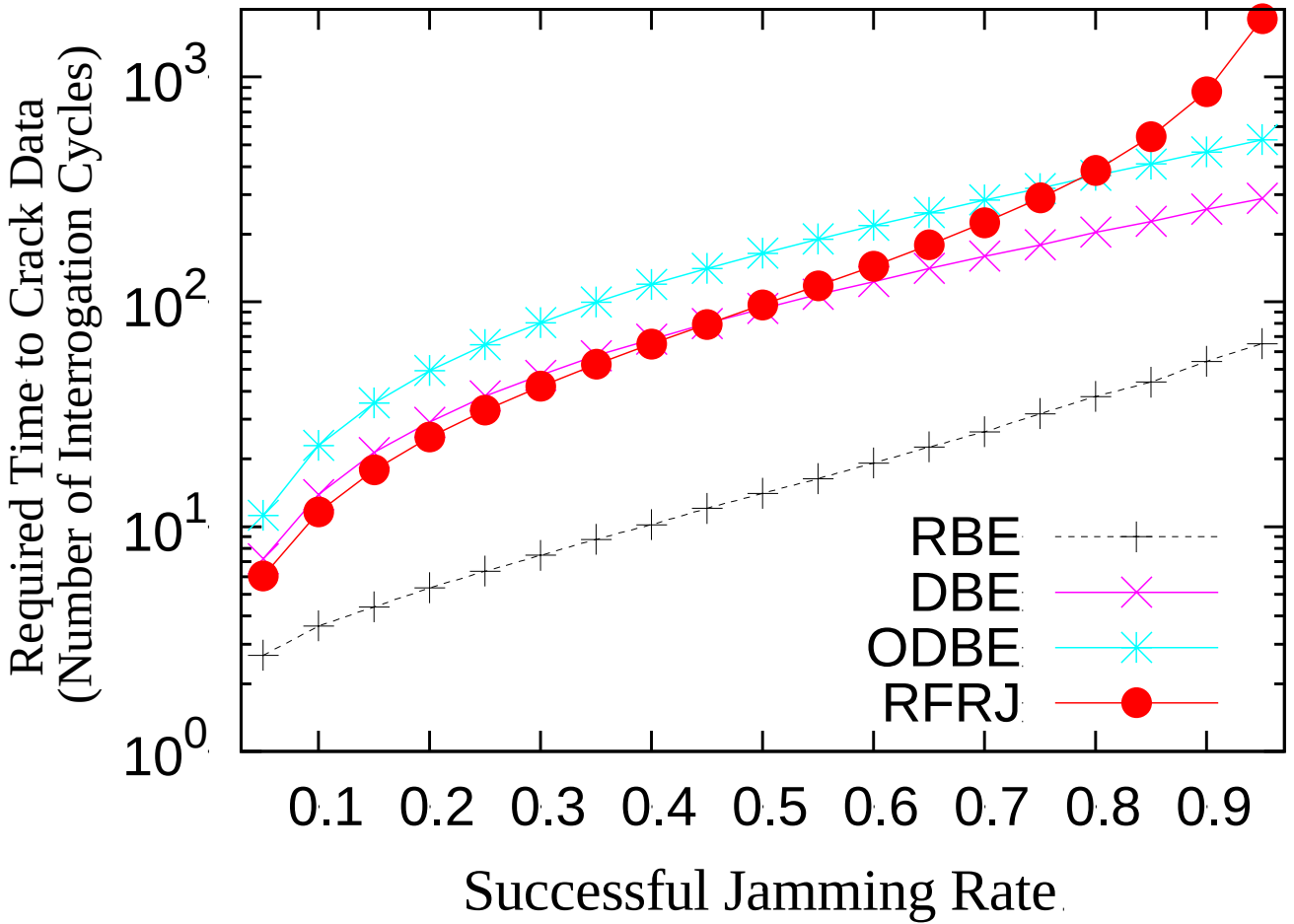


- Tag's ID may partially disclosed to adversaries if jamming fails to flip a bit in a codeword
- **Anonymity**
  - State of not being identified in an **anonymous set**
  - e.g., an eavesdropper receives 101XX
    - Anonymous set is {10100, 10101, 10110, 10111}
    - The original bit-string is identified by 0.25
    - Anonymity = (5 – 3) / 5 = 0.4
- **Perfect secrecy**
  - The system achieves perfect secrecy if the anonymity always equals to 1
  - If the jamming successful rate is 100%, **RFRJ protocol achieves the perfect secrecy**

- RFRJ v.s. existing solutions (RBE, DBE, and ODBE)
- Anonymity of the system
- Jamming successful rate is 100%



- Required time for an attacker crack the original tag's ID



Conclusions

- RFID systems bring productivity gains, but also raise security threats for individuals and organizations
- There are security and privacy issues in large-scale RFID forward and backward channels
- In the tutorial, **security and privacy issues** are addressed
  - **Private tag authentication**
    - Encryption-based and non-encryption-based protocols
    - The proposed schemes achieve high degree of security and performance
  - **Two RFID architectures**
    - Trusted Masking Device and Distributed RFID system