PERFORMANCE METRICS AND PARAMETERS OF SIGNATURE BASED AIR INDEX FOR DATA ACCESS IN BROADCASTING WIRELESS ENVIRONMENTS

Prof. Pragati Goel
Asso. Prof. MCA Dept.,
Sterling Institute of Management Studies,Nerul, Navi Mumbai.
Email: goelpragati78@gmail.com

Broadcast is one of the most suitable forms of information dissemination over wireless networks. It is particularly attractive for re-source limited mobile clients in asymmetric communications. To support faster access to information and conserve battery power of mobile clients, a number of indexing schemes have been proposed in recent years. In this paper, we report on our study of signature indexing scheme and studies on the performance issues (i.e., access latency and energy conservation) of wireless data broadcast. Signature-based indexes are particularly good for sequentially structured media (such as a broadcast channel) and multi-attribute indexing. By naturally encoding all the indexed attributes in a bit-vector (i.e., a signature), signature-based techniques allow clients to efficiently filter out unwanted data items and thus improve the performance. However, the important security issues have not been well addressed. This paper studies signature indexing and reveals various metrics and parameters affecting performance and security of signature-based air index schemes in wireless data broadcast.

Keywords *: Index scheme, Data broadcast, Signature based index*

INTRODUCTION

With rapid advent of wireless technology and growing popularity of smart wireless devices, there is a strong demand on pervasive data services, which facilitate wireless devices and information appliances alike to access much needed information from anywhere, anytime.

Wireless and mobile computing is one of the high growth areas in information technology. Mobile users require information while on the move. This gives great flexibility and convenience for data access. There are increasing numbers of applications that utilize wireless data access. For example, in *Geographical Information Systems (GIS),* mobile clients could ask for geographical information to find a restaurant of their choice in the vicinity or the highest peak in the area of interest. Another example is *wireless stock market data delivery*. Stock information from any stock exchange in the world could be broadcast on wireless channels or sent to mobile users upon requests. There are several research issues related to wireless and mobile computing [5]. In this paper, we focus on efficient data access. There are two fundamental modes for providing information for wireless applications: *Broadcast* and *on-demand*. For broadcast mode, information is broadcast over the wireless channel. Clients ”listen” to the channel and filter the interested information. This mode is also called *push-based* mode. On-demand mode provides information to the mobile clients by the fixed server only upon a request. Such message delivery method is sometimes referred to as *pull-based* mode. Two key requirements for data access in wireless environments are conserving power consumption and minimizing client waiting time.

In a wireless data broadcast environment, any client with appropriate equipment can monitor the broadcast channel and log the data items being broadcast. If the broadcast data items are not encrypted, the broadcast data content is open to the public and any person can access them. Key-based encryption is a natural choice for ensuring secure access of data on air (i.e., only the subscribers who own the valid keys can decrypt the received packets to obtain the data items). Therefore, a search for broadcast data items is answered by receiving all the broadcast data items off the air and decrypting them for further processing (i.e., filtering out unwanted data items). To help alleviating the high cost of receiving, decrypting and filtering broadcast data, auxiliary information may be provided on the broad-cast channel to annotate the broadcast data items. This technique is called air indexing. The basic idea is that, based on index information broadcast along with data items (including indexed attribute values, arrival schedule, length of data items, etc.), mobile clients are able to selectively skip unauthorized or unwanted data items by slipping into doze mode and switch back to active mode only when the data of desire arrives. This technique, substantially reducing workload and energy consumption of mobile clients, is particularly important for encrypted data broadcast. To facilitate efficient access of data on air, index information is preferred to be non-encrypted. Nevertheless, non-encrypted index information may allow unauthorized attackers to infer the data content on broadcast and therefore cause confidentiality loss. In this paper, we examine both performance and security issues in signature air indexing techniques.

Existing air indexing techniques can be roughly classified as tree-based and signature-based indexes [15]. The tree-based indexes, typically based on clustered data organizations, provide a very accurate and complete global view (in the form of index information) of data items being broadcast on air and thus are very energy efficient for clustered data items. Nevertheless, this 'complete' and 'accurate' index in-formation, if not encrypted, causes significant confidentiality loss which is the major security issue we are concerned in this study. On the other hand, signature-based indexes are particularly good for sequentially structured media (such as a broadcast channel) and multi-attribute indexing. By naturally encoding all the indexed attributes in a bit-vector (i.e., a signature), signature-based techniques allow clients to efficiently filter out unwanted data items and thus improve the performance. Since signature-based techniques do not provide the most clear index information, unauthorized attackers cannot be sure of the content of data items and thus reduced the confidentiality loss. In this paper, we investigate the tradeoff between performance and confidentiality of signature-based air indexes by analysis and experiments.The main contributions of our study are

- The tradeoff between performance and security requirements in signature based air indexes are analyzed in terms of false drop and false guess probabilities of the signatures.
- Performance and security of the examined signature schemes are analyzed. Analytical model for the impact of different control parameters is studied to serve as guidance for configuring signatures to meet the performance and security requirements of wireless data broadcast applications.

The rest of this paper is organized as follows. In Section 2, we present the overview of signature technique and their application in the wireless data broad-cast. In section 3 we briefly review some related work to this study. Section 4, tells the metrics for performance and security. Finally, we conclude this paper in Section 5.

RELATED WORK

Air indexing is commonly adopted to conserve battery power in mobile clients. Several tree-based indexing techniques, such as flexible indexing and distributed tree indexing, for broadcast channels were first introduced by Imielinski et al. [6, 7]. Based on the index tree method, work presented in [3, 13] uses unbalanced indexes to improve performance for skewed data access. However, these studies concentrated on one-dimensional indexes for equality-based queries. Lee et al. have addressed general queries with a semantics-based broadcast approach [10]. Tan and Yu have developed a broadcast program that supports range queries [14]. Traditional index techniques, such as hashing [7] and signature file [4], were also applied in air indexing, along with hybrid approach [5]. Besides the design of different indexing structures for different scenarios, index organization algorithms were also studied [8]. However, none of the above techniques addresses any security issue.

There is some related work done in the networking area. For example, [12] focused on secure multicast group key management in the network and [1] proposed broadcast en-cryption schemes that disseminate a secret to only the privileged clients. However, key management and cryptography are not the focus of this paper and we try to address the se-curity issue from data management aspect. Another related work in networking is Bloom Filter [2]. Different from the signature technique, it adopts multiple hashing functions to set the bit strings. As a result, the generation and comparison processes of the bit string become more complicated and time-consuming. It is not as suitable for the wireless broadcast systems as the signature techniques.

OVERVIEW OF THE SIGNATURE TECHNIQUES

A signature is essentially an abstraction of the information stored in a record. It is generated by a specific signature function. By examining a record's signature, one can tell if the record possibly has the matching information. Since the size of a signature is much smaller than that of the data record itself, it is consider-ably more power efficient to examine signatures first instead of simply searching through all data records. Indexing schemes making use of signatures of data records are called signature indexing. In [8], three signature indexing schemes are proposed: simple signature, integrated signature, and multi-level signature. The latter two schemes originate from the simple signature indexing. Since our focus is on comparing different indexing techniques, only the simple signature scheme is covered in this paper.

Access Protocol. The access protocol for simple signature indexing is as follows (assume *K* and *S* are the key and signature of the required record respectively, and *K(i)* and *S(i)* are the key and signature of the *i*th record):mobile client requires data item with key K tune in to broadcast channel keep listening until the first complete signature bucket arrives

(1) read the current signature bucket if S(i) = S(k)

download the data bucket that follows it if K(i) = K

search terminated successfully

```
else
false drop occurs
continue to read the next signature bucket repeat from (1)
else
go to doze mode
tune in again when the next signature bucket comes repeat from (1)
```

Simple Signature Scheme

Signatures are constructed from the information frames and broadcasted together with the information frames. The signatures may be broadcasted as a group before the information frames or interleaved with the corresponding frames. Figures 1 and 2 illustrate these two approaches. For the non-interleaved signature approach, since the client may start monitoring the broadcast channels at any moment, missing the signature segment means the client has to wait until the next broadcast cycle to access the signatures. The period of time from the moment a client tunes in until the first signature is received is called the initial probe time.

The most intuitive approach for interleaving signatures with information frames is to construct a frame signature for each information frame. The signature frame is broadcasted before the corresponding information frame (see Figure 2). When a mobile client wants to retrieve information from the broadcast channel, client specifies a query. A query signature S_Q is generated based on the specified query. Then the client tunes into the channel and uses S_Q to compare with the frame signatures received. When a match is found, the corresponding information frame is received by the client for further checking in order to eliminate false drops. If the frame is not a false drop, it will be retained in the result set. When a received frame signature does not match with the query signature, the client will switch into doze mode until the next signature frame arrives. If most of the frame signatures don't match with the query signature, the client will stay in doze mode for the most part of a broadcast cycle, thus saving a lot of energy. In this scheme, the average initial probe time is half of the average size of an information frame and its signature frame. The access time and tune-in time, however, are dependent on the positions of the initial probe:

position A: If the initial probe falls in the middle of a signature frame (point A in Fig. 2), the client has to stay active for the rest of the signature frame in order to detect the beginning of a new frame. Then, the client switches to doze mode. The initial probe time, access time and tune-in time are as follows.

Initial probe time = length of the partial signature scanned + the length of the first information frame.

Access time = initial probe time + a broadcast cycle.

Tune-in time = length of the partially scanned signature + every signature in a broadcast cycle + false drop and true drop information frames in the cycle.

position B: If the client initially tunes into the middle of an information frame (point B in Fig. 2), it will stay active for the rest of the frame so that it can detect the beginning of the next signature.

Initial probe time = the partial information frame.

Access time = initial probe time + a broadcast cycle.

Tune-in time = initial probe time + every signatures in a broadcast cycle + false drop and true drop frames in the cycle.

Figure 1: Non interleaved Signatures

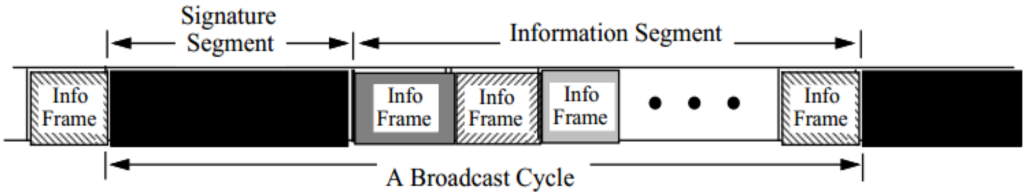
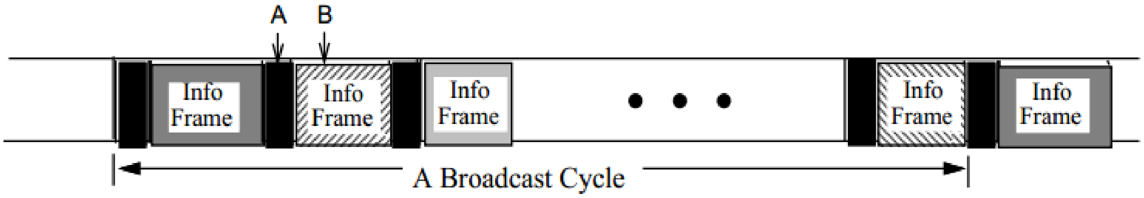


Figure 2: Interleaving Signatures



Information Broadcasting Using the Signature Technique

There are a number of ways to generate signatures. Given a set of data items to be indexed by multiple attributes, the signature S_i of data item i is typically formed by first hashing each indexed attribute in the data item into a bit string and then superimposing (i.e., bitwise-OR, denoted as ∨) all these bit strings into a signature. Note that the size of a signature equals the size of the bit string. An example of signature generation is depicted in Figure 1, in which an attribute is hashed into a 12-bit string. To process a query, a query signature S_Q corresponding to the query Q is generated similarly. The query processing is proceeded by comparing S_Q and the signatures of examined data items using bitwise-AND (denoted as ∧). The signatures match if for every bit set in S_Q , the corresponding bit in the compared data signature S_i is also set. There are two possible outcomes of the comparison:

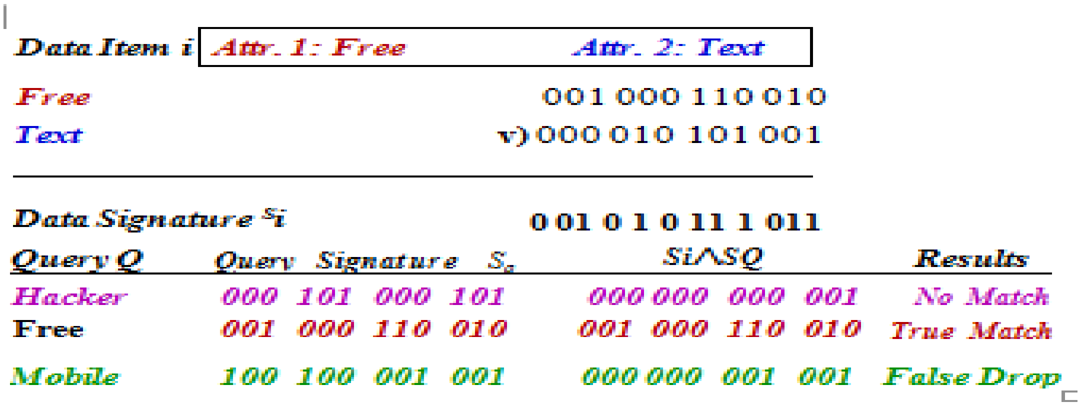
- SQ ∧ Si ≠ SQ : data item i does not match query Q.
- SQ ∧ Si = SQ : a match has two possible implications:

– true match: the data item is really what the query searches for; and

– false drop: data item in fact does not satisfy the search criteria although the signature comparison indicates a match.

As shown in Figure 3, three queries are issued and their corresponding signatures are produced. Based on the result of S_Q ∧ S_i , the examined data item is not qualified for the first query, Q=Hacker, and hence can be discarded. However, it shows a match for both queries Q=Free (true match) and Q=Mobile (false drop). Thus, the data item has to be retrieved for further checking.

Figure 3: Signature Generation and Comparison



Data Item i | Attr. 1: Free Attr. 2: Text

Free 001 000 110 010

Text v) 000 010 101 001

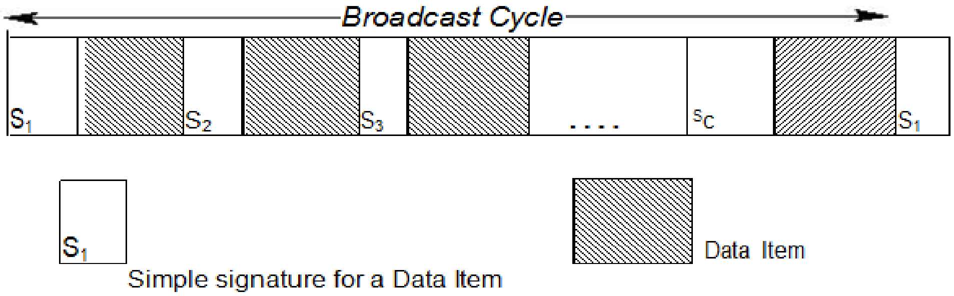
Data Signature S_i 0 01 0 1 0 11 1 011

Query Q	Query Signature S_Q	Si∧SQ	Results
Hacker	000 101 000 101	000 000 000 001	No Match
Free	001 000 110 010	001 000 110 010	True Match
Mobile	100 100 001 001	000 000 001 001	False Drop

Signature techniques have been employed for air indexing [11]. The idea is to put data items into groups and generate a signature for each group. A server broadcasts a signature before its corresponding group of data items. Since the data items are periodically broad-cast in the broadcast channel, a complete broadcast of the data items is called a broadcast cycle.

As shown below in Figure 4 respectively, each group in simple signature scheme contains only one data item while each group in integrated signature scheme contains multiple data items. It is assumed that the hashing function H adopted to generate the signatures is known to all the clients. Therefore, when a client issues a query Q, it first generates the query signature S_Q based on H and then starts the retrieval process. Since the signatures and data items are interleaved, the client listens to the signatures to decide whether to retrieve their corresponding data items. When a signature shows a mismatch, the client, by tuning into doze mode, skips the corresponding data items to save energy. It turns back to active mode again when the next signature is broadcast to continue the search process. We assume that each data item has the same size. Thus, the arrival time of the next signature packet is predictable.

Figure 4: Signature Generation and Comparison



PERFORMANCE AND SECURITY METRICS

Access time and tune-in time have been widely used as performance metrics in the studies of wireless data broad-cast. The former represents the access latency and the latter estimates the energy consumption in mobile clients. Since index information consumes extra bandwidth, balance between the overhead of access time and the gain from tune-in time is extremely important to all the broadcast systems adopting air indexes. A query issued by a client may request multiple items broadcast separately. Thus, a client has to listen to the whole broadcast cycle to avoid a miss of any right answer. Therefore, the access time is only affected by the bandwidth overhead caused by the signature. The tune-in time depends on the filtering ability of signatures. As long as a signature shows a match, whether a true match or a false drop, the data items have to be down-loaded and decrypted for further checking. Therefore, the false drop probability should be reduced to minimize the energy consumption on retrieving and decrypting unwanted data items. On the other hand, the fact that one signature can match different queries provide an uncertainty which prevents attackers (also called hackers) from knowing the indexed at-tribute values of data items. The hackers scan the broadcast channel, download indexes and data items, and try to guess the encrypted content of data items from indexing information. Hence, when a hacker downloads a signature from the broadcast channel, he might start a dictionary attack. He uses all the attributes in his dictionary $D_{\mathbf{H}}$ to generate $|D_{\mathbf{H}}|$ signatures and compares each of them with a down-loaded signature. Assuming that the attacker's dictionary is comprehensive, he will find a set of matches in $D_{\mathbf{H}}$ Among those matches, there are correct guesses and false guesses.

A key challenge for the system administrator is to determine confidentiality loss by answering "how much information has been leaked to attackers?" It is important for system administrators to be able to estimate and minimize the information leakage. Thus, information leaking degree (ILD) is defined as the security metric in this paper. An important job of the the system administrator is to facilitate highly energy efficient data access to only the authorized clients (maybe with a cost of some small access latency delay and bandwidth overhead), while minimizing ILD.

The tradeoff between performance and confidentiality, both of which are important to a broadcast system, is studied in this paper. The formal definitions of all the metrics are summarized as follows:

- Access time (ACC). The period of time from the moment a query is issued to the moment the client finishes receiving all the qualified data items. Assuming a fixed transmission rate, it is measured in byte.
- Tune-in time (TUNE). The time duration that a client has to stay in the active mode to answer a query. We use a normalized tune-in time (i.e., the ratio of the tune-in time to the whole broadcast cycle) here.
- Bandwidth overhead (BO). The bandwidth consumed on broadcasting signatures. Obviously, it is closely related to the signature length and the number of the groups within one broadcast cycle. It is also represented in the unit of byte in this paper.
- Information leaking degree (ILD). The expected number of correct guesses out of all the matched guesses obtained by an attacker. Intuitively, this depends on not only the deployed air index, but also the size and the quality of dictionaries used by attackers.

Control Parameters

The important parameters that affect the performance and security are listed as follows:

•Signature length. The number of bits in one signature (denoted by m).

•Bit setting. The fixed number of bits set to 1 in a bit string (denoted by w_b). Signature generation can be controlled by setting w_b between 1 and m.

•The number of indexed attributes. The number of attributes in a data item that contribute to the signature (denoted by u).

Due to the fact that a data signature is superimposed from the bit strings of the indexed attributes, the value of u impacts the filtering ability of the signature. In general, a signature superimposed from a small number of attributes provides a more accurate representation of the indexed item. Although this parameter is usually application dependent, it can still be adjusted subject to the system needs. In traditional information retrieval applications, the size of the signature, m, is set to a large value and the number of bits set, w_b , is carefully selected to provide a large space of hashed bit strings and minimize hash collisions. However, for secure wireless data broadcast systems, a large signature consumes too much bandwidth and extends both access latency and tune-in time. Furthermore, a larger signature may result in a higher ILD. Consequently, it is extremely important for the administrators to consider all the factors and choose proper control parameters.

ANALYSIS OF SIGNATURE-BASED AIR INDEX TECHNIQUE

As discussed earlier, a secure wireless broadcast system can meet different performance/security requirements by tuning the control parameters. We conduct several experiments by simulation to demonstrate the flexibility of signature-based air index. All the experiments are implemented in C language in a windows system. As shown in Table 2, the application domain D_A has 100 attribute values. We assume the application has 1000 data items to broadcast. Each data item is characterized by 10 indexed attribute values drawn from D_A . On the other hand, the attacker's dictionary D_H contains 1000 attributes which is a superset of D_A (i.e., we made a conservative assumption from the administrator's standpoint). The signature size, m, is set to 64 and 128. By tuning w_b from 1 to m, the system administrator can generate different configurations of signatures.

The experimental results shown here are obtained from the average of 100 queries, each of which is based on an attribute value drawn from D_A .

Table 1: Simulation Settings

$\|D_A\| = 100$	$\|D_H\| = 1000$	$u = 10$
$C = 1000$	$m = 64, 128$	$n = 64$
$s = 4$	$w_b = [1,m]$	$P_s = 0.01$

The analytical value of false drop rate shown here is approximated by :

Given a query Q and corresponding signature S_Q, false drop probability can be experimentally obtained as follows. Among the total C signatures within one broadcast cycle, let C_t be the true matches, C_f be the false drops, and C_m0 be the number of signatures that do not match S_Q, i.e

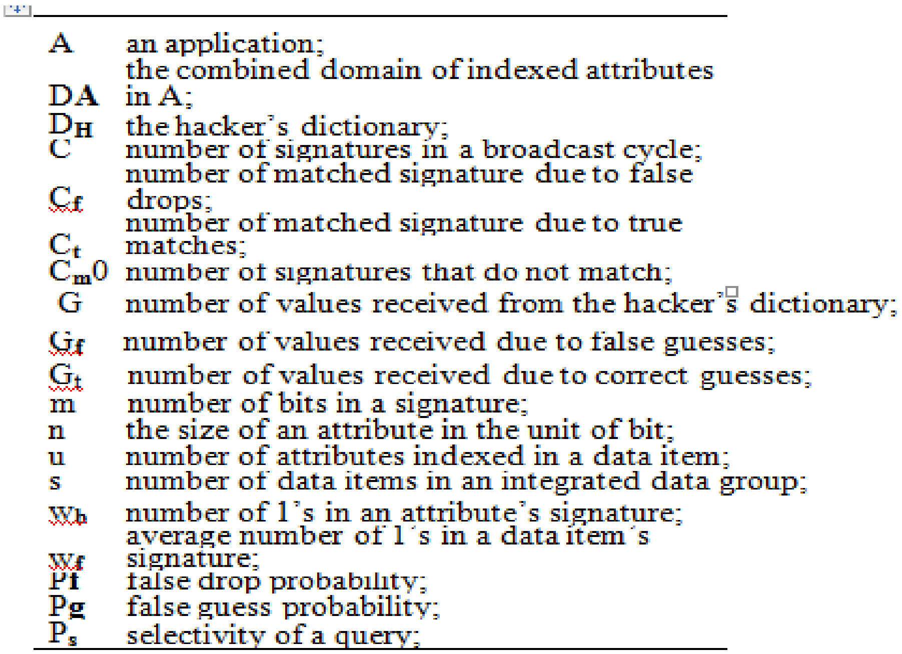
$C = C_t + C_f + C_m0$. The false drop probability P_f is defined as the ratio of C_f to $(C - C_t)$.

$$P_f = \frac{Cf}{C - C_t} \qquad \ldots\ldots\ldots\ldots\ldots(1)$$

The false guess probability is similarly obtained by :

Suppose an attacker receives a signature S_d from the broad-cast channel, he produces a set of signatures S_j where $j \in [1, |D_H|]$ from his dictionary and compares each of them with S_d . Let G be the total number of matched guesses, G_t be the correct guesses and G_f be the false guesses (i.e, $G = G_t + G_f$). The false guess probability P_g (from the attacker's view) can be defined as the probability that a dictionary value matches a signature of an item but actually is different from the item.

$$P_g = \frac{G_f}{|D_H| - G_t} \qquad \ldots\ldots\ldots(2)$$



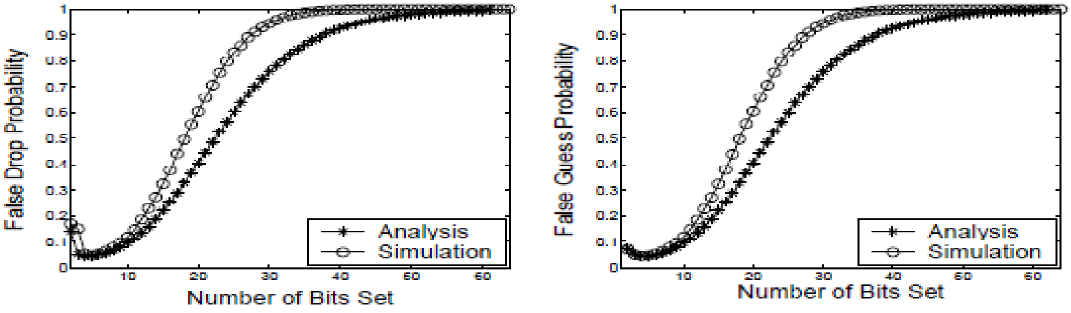
A	an application;
DA	the combined domain of indexed attributes in A;
D_H	the hacker's dictionary;
C	number of signatures in a broadcast cycle;
C_f	number of matched signature due to false drops;
C_t	number of matched signature due to true matches;
C_m0	number of signatures that do not match;
G	number of values received from the hacker's dictionary;
G_f	number of values received due to false guesses;
G_t	number of values received due to correct guesses;
m	number of bits in a signature;
n	the size of an attribute in the unit of bit;
u	number of attributes indexed in a data item;
s	number of data items in an integrated data group;
w_b	number of 1's in an attribute's signature;
w_f	average number of 1´s in a data item´s signature;
Pf	false drop probability;
Pg	false guess probability;
P_s	selectivity of a query;

Table 1: Notations.

RESULT

It's very important for the administrator to be able to efficiently decide the best signature configurations that meet specific performance and security requirements.

Figure 3: Performance/Security Metrics of Simple Signature Scheme (m = 64)



(a) P_f (b) P_g

For example, if an application requires a higher security, the administrator can raise the false guess probability higher by choosing a larger w_b . For an application to facilitate better energy conservation at clients, a lower false drop probability can be obtained by setting a smaller w_b. From the performance perspective, keeping low false drop probability helps clients retrieve the information from a broadcast channel efficiently. Meanwhile, from the security perspective, achieving high false guess probability prevents the hacker from guessing the information easily.

CONCLUSION

Air indexing is an important technique to facilitate energy conservation of mobile clients in wireless broadcast system. However, the crucial security issues on air indexing have not been discussed. This paper is effort to address both performance and security parameters and metrics in wireless data broadcast systems for signature-based air index .It is an ideal technique to meet the performance and security requirements of applications because the tradeoff between performance and security metrics can be properly tuned by system administrators. We define a security metrics called information leaking degree to measure confidentiality loss in air indexes both security and performance metrics in terms of a number of controllable parameters. The analysis provides much insight and guidance about tuning the system..

REFERENCES

[1] How to broadcast a secret, S. Berkovit, In Proc. of Eurocrypt'91, pages 536–541, Brighton, UK, 1991.

[2] Space/time trade-offs in hash coding with allowable errors, B. Bloom, Comm. of ACM, 13(7), July 1970.

[3] Indexed sequential data broadcasting in wireless mobile computing, M. Chen, P. S. Yu, and K. Wu, In Proc. of the 17th International Conference on Distributed Computing Systems, pages 124–131, Baltimore, MD, USA, 1997.

[4] Indexing techniques for wireless data broadcast under data clustering and scheduling, Q. Hu, W. Lee, and D. Lee, In Proc. of the 8th International Conference on Information and Knowledge Management, pages 351–358, Kansas City, USA, 1999.

[5] A hybrid index technique for power efficient data broadcast, Q. Hu, W. Lee, and D. Lee, Distrib. Parallel Databases, 9(2):151–177, 2001.

[6] Energy efficient indexing on air, T. Imielinski, S. Viswanathan, and B. R. Badrinath, In Proc. of the International Conference on Management of Data, pages 25–36, Minneapolis, MI, USA, 1994.

[7] Power efficient filtering of data on air, T. Imielinski, S. Viswanathan, and B. R. Badrinath, In Proc. of the 4th International Conference on Extending Database Technology, pages 245–258,

Cambridge, UK, 1994.

[8] Data on air: Organization and access, T. Imielinski, S. Viswanathan, and B. R. Badrinath, IEEE Trans. Knowledge and Data Engineering, 9(3):353–372, 1997.

[9] A partitioned signature file structure for multiattribute and text retrieval, D. L. Lee and C. Leng, In Proc. of the 6th International Conference on Data Engineering, pages 389–397, Los Angeles, USA, 1990.

[10] A semantic broadcast scheme for a mobile environment based on dynamic chunking, K. Lee, H. V. Leong, and A, Si. In Proc. of the 20th International Conference on Distributed Computing Systems, pages 522–530, Taipei, Taiwan, 2000.

[11] Using signature and caching techniques for information filtering in wireless and mobile environments, W. Lee and D. Lee, Journal of Distributed and Parallel Databases, 4(3):57–67, 1996.

[12] Iolus: a framework for scalable secure multicasting, S. Mittra, In Proc. of the International Conference of the Special Interest Group on Data Communication, pages 277–288, Cannes, France, 1997.

[13] Energy efficient indexing for information dissemination in wireless systems, N.Shivakumar and S.Venkatasubramanian, ACM-Baltzer Journal of NOMAD, pages 433–446, 1995.

[14] Generating broadcast programs that support range queries, K. Tan and J. X. Yu, IEEE. Trans. on Knowledge and Data Eng., 10(4):668–672, 1998.

[15] Index structures for selective dissemination of information under the boolean model, T. W. Yan and H. Garc´ıa-Molina, ACM Trans. Database Syst., 19(2):332–364, 1994.