

# Evaluating the merits of various RFID tagging protocols with respect to the protocol characteristics.

Chris Turner IEng MIIE  
6 June 2003

*There are many different air interface protocols used by RFID systems. The inventors of the various systems make claims as to their performance. But what is performance and how do you measure it? This paper attempts to clarify the individual characteristics of protocols, provide a means to evaluate them and arrive at a figure of merit.*

## 1 INTRODUCTION

When evaluating or assessing air interface protocols for RFID there is a range of characteristics that need to be considered.

These characteristics are:

- Efficiency as a data carrier
- Speed of reading
- Reliability
- Dynamic Tag Populations
- Variation in the size of the tag population
- Counting
- Group Select
- Bandwidth Requirements
- Complexity (which will affect cost)

From the above a Protocol type figure of merit can be derived in order to attempt a ranking of the different protocols.

### 1.1 EFFICIENCY AS A DATA CARRIER

The efficiency of a tag as a data carrier is the relationship between read time and the amount of data to be read from the tag. The read time consists of two main components;

- The time taken to sort out contention between tags (clash time) and;
- The time required for the tag to transmit its data (message time)

And may be expressed as:

$$\textit{Total Read Time} = \textit{Clash Time} + \textit{Data Transmit Time (message time)}$$

The issue is quite complex and any comparison between protocols on this aspect will require a careful analysis.

In order for an interrogator to collect data from a single tag, a dialogue (or connection) must be established with that tag that will not be disturbed by other tag transmissions. The Binary Search protocols and the simple roll call achieve this goal by their very nature. Aloha protocols achieve this goal by providing slots and via muting. Both systems acknowledge tags read and switch them off.

## 1.2 SPEED OF READING

The speed of reading a tag population is dependent on the total read time per tag and the efficiency of the contention management method. Various contention management systems have been developed which variously use Time Division Multiplexing, Frequency Division Multiplexing or combinations of these. Because of bandwidth limitations imposed by radio regulations, or simply because of tag complexity issues, most tag protocols use a single communications frequency and therefore Time Division Multiplexing is generally used.

The arbitration algorithms fall into two broad categories:

- Polling protocols based on binary or word search which include
  - Modified Binary search
  - Modified Binary search with conflict estimation
  - Clipped Binary search with conflict estimation
  - Tree splitting (binary or word)
  
- Non-polling protocols which include
  - Simple Aloha
  - Slotted Aloha
  - Random hold-off and retry Aloha and slotted Aloha
  - Carrier Sense Multiple Access (listen before talk)

Polling protocols are generally referred to as deterministic and non-polling protocols as probabilistic. However in order to improve the speed of reading in random populations of tags, polling protocols frequently incorporate deliberate randomisation making them to some extent probabilistic. Likewise some non-polling protocols incorporate deterministic elements such as group selection to improve throughput.

## 1.3 BINARY SEARCH PROTOCOLS

Binary Search involves searching through a list of numbers (tag identities). If the identities of the tags present are known to the interrogator, then a simple roll call is a very efficient search method. However if as is usually the case, the interrogator does not know how many tags are present or their identities, then a simple roll call is extremely inefficient. The optimised binary search protocols subdivide the tag population into smaller groups so improving search speed by shortening the search path. Once a tag has been read it is acknowledged and switched off thus removing it from the field and from the search tree.

Cidon and Sidi <sup>1</sup> describe the efficiency of a binary search system as the ratio:

$$M_k/L_k$$

Where:

$k$  is the multiplicity of the conflict, the number of nodes transmitting simultaneously.

$M_k$  is the average number of packets successfully transmitted during the Batch Resolution Interval (**BRI**): the interval between the start and end of the algorithm measured in slots where a slot is the duration of a transmission.

$L_k$  is the average length of a **BRI** given that it starts with an initial conflict of multiplicity  $k$ .

Table 1 shows example efficiency according to Cidon and Sidi for various binary search methods.

**Table 1. Search efficiency for various binary search methods**

Algorithm	Efficiency (limiting value as $k \rightarrow \infty$ )	Efficiency for $k=100$	Efficiency for $k=10$
Basic Binary Search	0.346		
Modified Binary Search	0.381	0.383	0.396
Modified Binary Tree (with conflict estimation phase)	0.468	0.418	0.396
Clipped Binary Tree (with conflict estimation phase)	0.487	0.449	0.407

#### 1.4 NON-POLLING PROTOCOLS

The efficiency of non-polling protocols, has been described by Stallings<sup>2</sup>, for Aloha and CSMA in packet radio access systems. RFID systems are packet based systems where the tag message transmission is analogous to a packet. Stallings defines the throughput (efficiency)  $S$  as the number of packets generated per packet time or the fraction of channel capacity that is used.

Three protocols are discussed:

- ALOHA (probabilistic),
- Slotted ALOHA (probabilistic with a deterministic element) and
- CSMA (Carrier Sense Multiple Access – only transmit packet if the channel is clear). CSMA is achieved in RFID systems by preventing more than one tag from transmitting in a time slot.

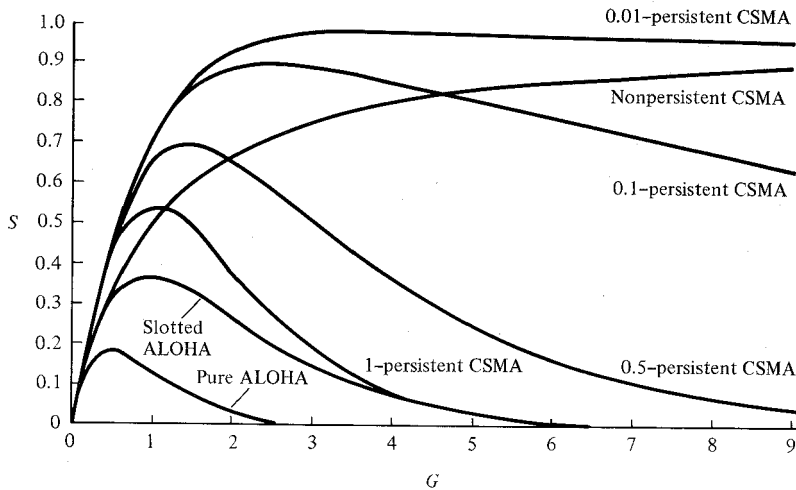
Stallings uses the term ‘persistent’ to refer to the probability that a station will transmit immediately if the channel is free. In a hold-off and re-try Aloha system this is analogous to the maximum wait time (hold-off time) in multiples of slots. Therefore a 0.1 persistent CSMA has a probability of 1 in 10 of transmitting immediately the channel is free. This roughly corresponds to a maximum wait time of 10 slots (where a slot is the time duration of a tag transmission). Once a tag has been read it is acknowledged and effectively removed from the field.

The throughput  $S$  of the system is a function of the load  $G$ , which is the total rate of data presented to the network for transmission. In an RFID system, the load is the ratio of the number of tags present in the interrogator field of view to number of slots available. (For example a load  $G$  of 0.5 would occur for a population of 32 tags where the maximum number of slots is 64).

Figure 1 shows the relationship between  $S$  and  $G$  for various values of persistence.

#### Muting

Muting can be used to improve persistent systems by causing all but the first tag to transmit, to hold off until the channel is clear (the first tag completes its message), thus preventing an overlap or clash from occurring.



**Figure 1. Values of  $S$  and  $G$  vs persistence**

Table 2 shows the optimal efficiency  $S$  and the Optimal load  $G$  for various protocols.

**Table 2. Optimal efficiency for optimal load**

Protocol	Optimal Efficiency (Throughput $S$ )	Optimum value of $G$ (Load)
ALOHA	0.18	0.5
Slotted ALOHA	0.37	1
CSMA (1 - persistent)	0.52	1.2
CSMA (0.5 - persistent)	0.68	1.5
CSMA (0.1 - persistent)	0.88	2.5
CSMA (0.01 - persistent)	0.96	3.5

Note: The values of  $S$  are the maximum possible values of  $S$ .

## 2 EVALUATING READING EFFICIENCY

### 2.1 INVENTORY ACTIVITIES

The four main activities that take place during an inventory or reading process are

- Clash Time,
- Data Transmit Time,
- Data Acknowledgement Time and
- Command Time.

#### 2.1.1 Clash Time

This is the time where tags try get an open period in which to transmit. It is important to note that the tags do not have to transmit all their information, they just have to attract the attention of the interrogator in order to get allocated a chance to transmit their information.

Clash time is wasted time, in that while the tags are clashing, no useful information has been received by the interrogator.

- In a binary search protocol, the tag message length will determine the duration of the clash time.
- In a hold-of and retry protocol, the number of tags, the frequency of transmission and the message length will determine the duration of the clash time. The hold-off time may be varied dynamically to allow for varying numbers of tags in the field.

### 2.1.1 Data Transmit Time (Message Time)

This is when the information is delivered and the duration is determined by the amount of information and the rate at which it is delivered.

- If another tag transmits in this time both transmissions are corrupted and all the time has been wasted. Therefore all other tag transmissions **MUST** be silenced for optimum efficiency

### 2.1.2 Acknowledgement Time

Once the tag has transmitted its message the interrogator will acknowledge the tag to prevent it from transmitting again. Thus if a tag transmission is not acknowledged by the interrogator the tag will transmit again. This has the obvious benefit that as time passes fewer and fewer tags now contend for a speaking slot and the Clash Time decreases.

### 2.1.3 Command Time

This is the time used by the interrogator to instruct the tag how to respond. This may include wakeup commands, arbitration commands (excluding acknowledgement commands) or group selection commands.

## 2.2 MEASURE OF EFFICIENCY

While the overall efficiency of the communications is dependent on the rate of the data transmission the relative efficiency of protocols depends on the minimisation of the Clash Time. The Data Transmit Time and the Data Acknowledgement are constant overheads in the sense that each tag has to transmit its information at least once and each transmission has to be acknowledged as least once. The number of commands transmitted will depend on the collision arbitration method used and so Command Time will vary accordingly.

There are therefore four main efficiencies:

- Clash Efficiency,
- Transmit Efficiency,
- Acknowledgement Efficiency and
- Command Efficiency.

There are two other issues that affect the efficiency of reading which should be accounted for.

These are:

- The number of passes required to read user data from tags in the population (ie is it necessary for tags to be first identified during an inventory process and then the user data collected from the tags?)
- The ratio of the total interrogator transmit time to total tag response time to collect user data from all the tags present in the field.

### 2.2.1 Clash Efficiency:

Any increase in Clash Time as the number of tags in the population increases, should be as close to linear as possible. The Clash Time should not be dependent on the Data Transmit Time (Message Time). The Clash Time should only depend on the number of tags present and the probability of two or more tags clashing. Only one tag should be acknowledged at a time. The communication between tag and interrogator should be efficient - the ideal is that all tag transmissions that do not clash will be heard. In order to achieve this the interrogator should listen continuously.

### 2.2.2 Transmit Efficiency:

- A transmission must not be corrupted, ie two tags should not transmit data simultaneously.
- A corrupted transmission should be terminated as soon as possible.
- A tag should transmit its data only once.
- The data must be transmitted as fast as possible (as many bits per second as possible).
- Unproductive time between tag transmissions must be minimised

### 2.2.3 Acknowledgement Efficiency:

- Only correct transmissions must be acknowledged.
- Acknowledgement protocol must be as fast and as simple as possible.
- The tag must successfully receive the acknowledgement.

### 2.2.4 Command Efficiency

Command Efficiency depends on the amount of time that the interrogator requires to arbitrate the population of tags and to collect their data. For example if the interrogator needs to transmit a relatively long command for each tag response, then the total read time will be extended. Likewise if the interrogator needs to use one command to identify the tag and another command to collect the data then the total read time will be further extended. However if the interrogator uses very short commands for each tag response, then the Total Read Time will be optimised.

If the tag has multiple data fields and the interrogator only needs to collect a portion of the data then the total command time vs the total message time needs to be considered in deciding whether to use selective addressing commands and collect only the required data or to use one command to collect all the data and for the interrogator to discard the unwanted data.

## 2.3 TOTAL READ TIME

The time taken to read a population of tags may be expressed as:

$$\text{Total Read Time} = \text{Clash Time} + \text{Data Transmit Time} + \text{Data Acknowledgement Time} + \text{Command Time}$$

The most important protocol dependent aspect of the Read Time is the Clash Time. Marais<sup>3</sup> has defined a variable which he calls 'Clash Efficiency Criterium' (CEC). In order to provide a measure independent of actual data rate Marais proposes to measure the time periods in the equation in terms of data bits, which can then be later converted to Time. Clash Time can be expressed in terms of Slots because a tag message will have a duration of at least one slot.

The Clash Time can then be expressed as:

$$\textit{(The number of slots taken to sort out the clashes) * (The duration of a slot)}$$

The number of slots needed to sort out the clashes is a statistical parameter that depends on the number of tags and the search algorithm in the case of binary search and the maximum wait time in the case of non-polling algorithms.

The duration of a slot should be expressed in bits to account for different communication data rates in the exchange of information between the interrogator and the tag. The efficiency of the information exchange should be as high as possible, ie. as few as possible bits.

Therefore the definition of Clash Efficiency Criterion may be expressed as:

$$\begin{aligned} \textit{CEC} &= \textit{(Total Clash Time in bits) / No of tags} \\ &= \textit{(The number of slots taken to sort out the clashes) *} \\ &\quad \textit{(The duration of a slot in bits) / No of tags} \end{aligned}$$

In the case of the simple random selection of a wait time for non-polling algorithms, the number of slots required to resolve the clashes is primarily a function of the number of tags and the number of available slots.

In the case of binary search algorithms the number of slots required to resolve clashes is a function of the number of tags on each branch and the distribution of tags over the branches.

The most important protocol dependent aspects of the Read Time are the Clash Time and the Command Time, therefore these should be minimised.

## **2.4 OTHER CONSIDERATIONS**

There are an number of other factors to be considered when evaluating an air interface.

### **2.4.1 Counting**

Where tags are to be used to replace bar codes, it may be necessary for items to which tags are attached to be counted.

Do tags need a unique identity for the arbitration system to work and for items to be counted?

Uniquely numbered tags can simulate counting by concatenating a unique serial number with a product code A 64 bit unique number should be sufficient, but this will still add 64 bits of overhead to the tag number which will decrease the reading speed of the system compared to a system which is able to count tags containing identical user data.

Tags may be anonymous yet only need to be distinguishable for successful arbitration.

### **2.4.2 Group Select**

Does the air interface have a means to select groups of tags by Application or by owner of data structure?

One way this may be achieved is by means of group select commands or by means of a wakeup command from the interrogator which uses a group select filter, such as the AFI/ASF scheme specified in ISO TR 15961.

### 2.4.3 Variation in the size of the tag population

Is the protocol sufficiently robust to cope with a wide range of tag populations size?

A typical example would be a gate through which people pass. For example, if 10 people normally pass through the gate, what will happen if there is an emergency and 30 people try to cram through the gate at once? Will the interrogator cope?

### 2.4.4 Dynamic Tag Populations

#### Late arrivals:

- What happens if a tag comes enters the interrogation field halfway through a reading cycle and misses a wakeup call or start signal?
- What happens if a tag is in the sleep mode when it arrives?
  - Will the late arrival be read?
  - Will a tag miss its sequence when arriving late?
  - Will the late arrival cause the system to hang-up?

*All binary systems that construct the search tree dynamically or do a preliminary survey to determine which branches of the tree are populated, have a problem with late arrivals. For example, the system has done a survey and found 4 tags with a value of 1 in the least significant bit position. The interrogator then steps to the next highest bit position. A fifth tag also with a value of 1 in the LSB position arrives in the field. It will be missed but will not hang up the system.*

*Aloha protocols cope well with Late arrivals. Provided a late arriving tag is in the interrogation field for long enough it will be read without any need for the interrogator to backtrack.*

#### Early departures:

- What happens if a tag leaves halfway through the reading cycle?
  - Will it be read
  - Will it cause the system to hang-up?

*Dynamic binary search and survey systems: The system has done a survey and found 4 tags with 1 as the LSB and moves on to the NSB. One of the 4 tags departs early. It will not be read. The system is expecting 4 and only gets 3. Could cause a hang up.*

*An early departure may also not be read by the Aloha protocols. The faster ones are more likely to read it.*

## 2.4.5 Bandwidth Requirements

Bandwidth needs to be considered for both the interrogator transmission and the receiver.

### Receiver considerations

Passive backscatter tags communicate with the interrogator by modulating the incident energy using some form of impedance modulation. The resultant backscatter signal is usually below the required spurious emission limits. In some jurisdictions this means that the tag emissions are not regulated or restricted, but in other jurisdictions the information bearing part of the emission may not fall outside of the allocated bandwidth whatever its level.

In many regulatory jurisdictions including Europe, the permissible interrogator receiver bandwidth is regulated. This in turn limits the modulation bandwidth of tag signals which can be received.

### Transmitter considerations

Transmitter occupied bandwidth and duty cycle are regulated to a greater or lesser degree by radio regulations depending on the jurisdiction. In addition to the regulatory requirements, consideration should be given to the effect of the transmissions on co-located systems, both of the same type including RFID and of other users of the frequency, which may or may not be RFID.

In the case of passive backscatter tags, the interrogator transmitter serves two purposes;

- It provides power for the tag
- It provides a carrier onto which the tags data can be modulated and returned to the interrogator receiver.

The transmitter carrier therefore needs to remain on for the duration of an interrogation cycle, but the interrogator only needs to modulate when it communicates with tags to conduct an inventory or read tag data.

Channel occupancy may be described in terms of time/power/bandwidth. Interrogator receivers are expected to receive and decode the weak signals backscattered from tags. A typical tag signal will arrive back at an interrogator receiver at a signal level of  $-60$  to  $-70$ dBm. An interrogator radiating a signal of 33dBm (2 Watts) at a distance of 10 metres from another interrogator receiver, will have a path loss in the order of 40dB, which means that the signal arriving at the interrogator receiver will be at a level of  $-13$ dBm or 50 to 60dB stronger than a tag signal. It will therefore completely block reception of a tag.

The occupied bandwidth of the transmitter depends on:

- Modulation data rate (transmission speed)
- Data volume (per burst)
- Transmission repetitions

Spectrum compatibility characteristics are therefore determined by

- Instantaneous transmitter bandwidth
- Duration of the modulation (% of the time that the interrogator talks)
- Radiated power of the transmission
- Directivity of the transmitter radiated field (antenna radiation pattern)
- To a lesser extent - duration of transmitter carrier on time.

It is generally accepted that short bursts of higher bandwidth will be better tolerated by other users of the spectrum (cause less of a nuisance) than longer bursts of lower bandwidth.

Data rate does not affect the communication protocol because all parts of the protocol can be scaled. What is important is the throughput of data, which is affected by data rate and the efficiency of the protocol.

### 2.4.6 Complexity

What complexity is required in a tag to implement the protocol?

In particular what is the effect on:

- Gate count?
- Analogue circuits (clocks etc)?
- Chip area

## 3 PROTOCOL FIGURE OF MERIT

The relative efficiencies of various protocols may be rated according to their “bit hunger”; that is, the number of interrogator transmit bits required per successful read of a tag’s data for a given tag data size.

### 3.1 PUTTING IT TOGETHER

The contributors to reading efficiency during an inventory and data collection process are

- Tag Message Time
- Clash Time
- Command Time comprising
  - Arbitration command time
  - Data collection command time
  - Data Acknowledgement Time
  - Wakeup Command Time

Assume that all systems have a fixed data field of 128 bits. Each system to be evaluated will have a different tag data transmit rate. The total time taken to transmit the data will be given by:

$$\text{Tag Message Time} = \text{number of tags} * (128 + \text{message overhead}) / \text{data rate in bits per second.}$$

The clash time is derived from the efficiency of the collision arbitration protocol. Refer to the Efficiency  $k$  in table 1 for Binary Search and  $S$  in Table 2 for Non-polling algorithms. An efficiency of 0.5 means that 2 messages will be transmitted by each tag for 1 successfully received. (An efficiency of  $0.333 = 1$  in 3 is successful). Clash time is given by:

$$\text{Clash Time} = \text{number of tags} * (\text{Tag Message Time} / \text{Efficiency}) - \text{Tag Message Time}$$

Arbitration command time is the time used by the interrogator to arbitrate the population of tags present in the field. An arbitration command will be sent once for each clash and once for each message. Therefore arbitration command time is given as:

$$\text{Arbitration Command Time} = \text{number of tags} * (\text{number of command bits} / \text{efficiency}) / \text{data rate}$$

Data Collection Command Time is the time used by the interrogator to collect data from the tag. Some systems require the interrogator to first take an inventory of tags (It is possible that more than one command may be required to collect data from each tag), others collect the tag data as part of the arbitration process. In the latter case there will be no Data Collection Command Time. In the former case the time is given by:

$$\text{Data Collection Command Time} = \text{number of tags} * \text{number of commands} * \text{number of command bits} / \text{data rate}$$

Data Acknowledgement Time is the time taken to acknowledge the successful receipt of the tag data and an instruction to tell the tag to be quiet for the rest of the read cycle. In general there will be one acknowledge for each tag in the field (message received). Data Acknowledge time is given by:

$$\text{Data Acknowledge Time} = \text{number of tags} * \text{number of command bits} / \text{data rate}$$

Some systems require an initial wakeup or start of arbitration command. If used this command time must be added to the total cycle time. This time is given by:

$$\text{Wakeup command time} = \text{number of bits} / \text{data rate}.$$

### 3.1 CALCULATING THE READ FIGURE OF MERIT.

The figure of merit is the sum of all the contributing times divided by the number of tags in the field. Annex A provides an example of the figures of merit for various systems evaluated according to this method.

### 3.2 OTHER INFLUENCING FACTORS

Because it is unlikely that RFID systems will be operated in isolation, but more than likely be operated within range of other RFID systems, both channel occupancy and read efficiency affect the overall performance and throughput of an RFID system. If re-tries increase due to co-channel interference, then the actual throughput will decrease.

The increase in retries is similar to an increase in clash time and so channel noise and interference may be accounted for by adjusting the Efficiency in the equations.

### Bibliography

- Cidon, I. and Sidi, M. Conflict Multiplicity Estimation and Batch Resolution Algorithms. IEEE Trans. on Information Theory, Vol. 34, No. 1, p101-110, (1988).  
Stallings, W. "Data and Computer Communications", Macmillan (1994.)  
Marais, Mario, "Analysis of the efficiency of anti-contention protocols". CSIR, 21 February 1997.