ство, и строки, содержащие выбросы, обнаруживаются с помощью нового механизма измерения сходства в двумерном пространстве.

Результат. Сначала, с использованием матричных вычислений, была определена частота употребления слов в строках данных, а затем, с ее помощью, вычислялось сходство семантики и структуры. После перевода строки данных из одномерного в двумерное пространство, с помощью меры сходства, были определены ошибочные значения.

Научная новизна. Проведено исследование по обнаружению строк, содержащих выбросы, для очистки данных. Во-первых, сформулирована модель вычисления сходства с учетом семантического и структурного факторов. Во-вторых, с помощью построения ячейки сходства для проецирования строки данных, осуществлялось измерение расстояния сходства.

Практическая значимость. Метод может быть использован для очистки строк с аномалиями в информации о клиентах на любом предприятии, чтобы гарантировать качество данных в информации о клиентах, а также снизить затраты на обслуживание данных. Проведено исчерпывающее количество моделирующих экспериментов с целью доказать целесообразность и рациональность этого метода. Результаты показали, что этот метод позволяет улучшить точность обнаружения строк с выбросами.

Ключевые слова: *качество данных, очистка данных, обнаружение выбросов, вычисление матрицы, семантическое сходство, структурное сходство, ячейка сходства, расстояние сходства*

Рекомендовано до публікації докт. техн наук В. В. Гнатушенком Дата надходження рукопису 18.04.15.

Changwang Liu[1], Chao Yin[2], Yihua Lan[1]

1 – Nanyang Normal University, Nanyang, China
2 –Jiujiang University, Jiujiang, China

IMPROVED BINARYANTI-COLLISION ALGORITHM FOR RFID

Чанван Лю[1], Чао Інь[2], Іхуа Лань[1]

1 – Наньянський педагогічний університет, м. Наньян, КНР
2 – Цзюцзянський університет, м. Цзюцзян, КНР

ПОКРАЩЕНИЙ БІНАРНИЙ АНТИКОЛЛІЗІЙНИЙ АЛГОРИТМ ДЛЯ РАДІОЧАСТОТНОЇ ІДЕНТИФІКАЦІЇ

Purpose. Internet of Things (IoT) represents the future direction of the development of computer and communication technology, which is considered to be the third wave of development in the field of information industry after the computer. IoT is the implementation of a network of goods real-time information system based on Frequency Identification (RFID) and Electronic Product Code Radio (EPC). In the process, all kinds of existing technology will face many new opportunities and challenges, especially RFID.

Methodology. There are two kinds of problems in the research of the problem of label collision at home and abroad. One is binary anti-collision algorithm based on the tree, another isanti-collision algorithm based on time slot ALOHA. But ALOHA algorithm is rapidly deteriorated so that it is not suitable for large-scale application in the IoT.

Findings. In the binary tree anti-collision algorithm, the mature algorithms are the binary tree anti-collision algorithm based on pruning branches (pruning branches algorithm) and similar binary anti-collision algorithm (similar algorithm).

Originality. We have developed a new anti-collision algorithm called improved anti-collision algorithm (IAC), which is able to reduce the number of data in each time slot, the number of times and searches.

Practical value. Test results show that the IAC algorithm can improve the performance comparing to traditional pruning branches algorithm and similar algorithm. At the same time, IAC algorithm can reduce the search time very much.

Keywords: *RFID, binary tree anti-collision algorithm, ALOHA algorithm*

Introduction. IoT is the network, which connects the Internet with any goods in order to realize intelligent identification, location, tracking, monitoring and management. It uses radio frequency identification sensors, infrared sensors, global positioning systems, laser scanners and other information gathering equipment for exchange and communication [1, 2].

RFID is a non-contact automatic identification technology [3], which is based on radio frequency signal (inductive or electromagnetic) transmission characteristics to achieve automatic identification of objects or goods. RFID technology has the advantages of strong anti-interference ability, a large amount of information, a non-visual range of reading and writing and long life comparing with other automatic identification technologies, such as barcode technology, optical recognition and biometric technology, which includes the iris, face, voice and fingerprint [4]. It is widely used in logistics, supply chain, animal and vehicle identification, access control system, library management, automatic charge and production, etc.

Multiple RFID tags response to readers known as a multi-access technology. The development of the multiple access technology has many ways to resolve conflicts [5,6]. It includes the following four species: Frequency Division Multiple Access (FDMA), Space Division Multiple Access (SDMA), Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA). Because the design of RFID tag requires low power consumption and low complication in memory and computing capability, the anti-collision algorithm generally used in many systems is based on TDMA.

There are two kinds of anti-collision algorithms in the TDMA method. One is based on ALOHA, another is based on the tree structure [7]. The former uses a random selection of tags to send information, such as ALOHA algorithm. The latter uses a reader to check the information returned by each tag. The most common method is the binary search algorithm.

With the development of the IoT, the existing algorithms became unable to meet the requirements of the IoT. For example, the scale of the data information is huge, which often causes the lack of consideration in the previous protocol design.

The existing binary tree-based anti-collision algorithm and the binary anti-collision algorithm remain unable to meet the requirements of large-scale tags and mass data processing. In this paper, we have proposed IAC algorithm, which can improve the performance of the algorithm from two aspects. Firstly, it can reduce the number of data in each time slot. Secondly, it also can reduce the number of times and the number of searches. The simulation results show that the improved algorithm proposed in this paper has a great improvement in time efficiency and the number of times. Especially in the number of tags, the performance optimization is more obvious, which is suitable for the recognition of a large number of tags.

The contributions of this paper are described as follows:

1. Reduce the data for each transmitted, making use of judging the content and time of the conflict is small, but also makes the reception of data more efficient.

2. After distinguishing a completed ID tag, the algorithm will start the search from the nest of the collision position instead of the beginning, which can reduce the sending numbers and improve efficiency.

3. Experimental results show that IAC algorithm has improved a lot in the performance of the system.

Related works. ***RFID technology.*** RFID technology is a new automatic identification technology directly inherited from the concept of the radar. In 1948, Harry Stockman published the paper “The Use of Reflective Power Communication”, which lays the theoretical foundation of RFID. In the past half a century, the RFID technology has improved very much.

After 2001, as the RFID standardization problem attracted more attention, the RFID product category became more abundant. Active tags, passive tags, and semi-passive tags have been developed; at the same time, the cost of tags has been reduced. The expansion of the industry stimulates the development of RFID technology theory.

Anti-collision algorithm. When there are a number of tags to be identified in the reader’s scope, there are two different basic communication methods: data transmission from the reader to the tag and the data transmission from the tag to the reader. The first communication mode is also known as the radio broadcast. The second communication mode means that the data with many tags is transmitted to the reader in the function [8].

When the reader receives different responses from different tags, the reader cannot identify any tags with the different signals. The phenomenon is called tag anti-collision. The algorithm for solving the problem of tag collision is called tag anti-collision algorithm.

In all times, the problem of multi-access in radio communication existed. As we already mentioned, the anti-collision algorithm is divided into four methods: FDMA, SDMA, CDMA, and TDMA. The TDMA is commonly used because of the limitations of technology and costs, especially the label production costs.

There are two kinds of problems in the research of the label collision. One is binary anti-collision algorithm based on the tree, another is anti-collision algorithm based on time slot ALOHA [9]. The binary anti-collision algorithm based on the tree is to filter out the different serial numbers, and the signal exchange between the reader and the tag is often carried out by the binary anti-collision algorithm. It is based on a unique serial number to identify tags. In order to select one of a set of tags, the reader sends a request command. The data of the tag serial number is transmitted to the reader. If there is a collision, the next step is to narrow the search.

ALOHA anti-collision algorithm based on time slot reduces the probability of collision as much as possible by making the tag choose the different time to reply to the information [10]. With the increase in the number of tags, the performance of the ALOHA algorithm falls dramatically. It is not suitable for large-scale application in the IoT.

The IAC algorithm. ***The Design of the Algorithm.*** Because ALOHA algorithm is not suitable for large-scale application in the IoT, we focused on the binary anti-collision algorithm based on the tree.

In the binary tree anti-collision algorithm, the mature algorithms are pruning branches algorithm and similar algorithm. There is redundant information in the pruning branches algorithm, which reduces the transmission efficiency. While similar algorithm reduces the time of packet search and greatly improves the efficiency of data transmission, but it increases the transmission times of the command.

The analysis of the two kinds of anti-collision algorithms showed that there are two aspects to improve to obtain the algorithm with better performance. Firstly, we should reduce the number of bytes transmitted in each time slot. Secondly,we should reduce the number of times of the search command.

Comparing to pruning branches algorithm, IAC reduces the transmitted data each time, which reduces the content and time of judging the conflict. It also makes the reception of data more efficient. The system sends 1-bit data each time, even if there is a collision, it can be automatically identified based on the binary value. Especially when the collision occurred at last position, it can identify the sequence number of two tags. The improved binary anti-collision algorithm is a complete identification tag by ID, no longer start the search, but from the beginning of a collision to reduce the number of inquiries, send commands, and improve efficiency. Comparing to similar binary anti-collision algorithm, IAC searches the next collision position instead of beginning position, which can reduce the number of inquiries and improve the efficiency.

Algorithm Description. In order to identify the exact location of the bits in the data collision of the reader, we use the Manchester encoding. The general binary search algorithm is used to obtain the tag information, which makes the encoding way freer. But it increases the burden of software and hardware. This is why the IAC algorithm uses Manchester encoding to confirm the exact location of the crash site.

The command in the algorithm is defined as follows:

1. Register definition

There is a collision counter in the tag, which records the position of the parent node of the collision. If the tag sequence number has eight bits, the counter at least has three bits.

2. Tag states

There are three tag states:

State NEW means it enters the reader's scope at first. State ACTIVATE means it has been activated. State IDENTIFIED means it has been identified.

3. Search process

a. Request order: Request (n, m).

b. If the collision counter is equal to m, the tag is required to send n bit data.

c. Judge whether the data returned the tag is a conflict or not.

d. If there is a conflict, add the collision counter 1.

e. Under next request order Request ($n+1$, m) or Request ($n+1$, $m+1$).

Algorithm Implement. Assuming that the EPC code is 8, there are four tags in the reader function area. The four tags are shown as follows:

T1 = 01001000
T2 = 01001100
T3 = 10110101
T4 = 11011010

The principle of the algorithm is shown in Fig. 1.

1. All the new entrants to the reader's scope of action are in the NEW state. First, the reader sends the ACTIVE command to activate the tag, so that all is activated by the label in the NEW state. At the same time, the collision counter is equal to zero.

2. The reader sends command Request (1, 0), which lets all tags with collision counter that are equal to "0" send the first data. If the data has a collision, the tag will be added one.

3. The reader sends command Request (2, l), which lets all tags with collision counter that are equal to "1" send the second data. If the data has a collision, the tag will be added one.

4. The reader sends command Request (3, 2), which lets all tags with collision counter that are equal to "2" send the third data. If the data has a collision, the reader continues to send command Request (4, 2) until the reader sends command Request (8, 2) without a collision. At this time, the tag ID is identified by T4.

5. The state of the tag is changed from ACTIVATE to IDENTIFIED, which means the tag has been identified and no longer to be read.

6. The reader turns back to the last conflict point and sends command Request (3, l). It makes all tags with collision counter that are equal to "1" send data. And so on, until the T3 is identified from the NEW state changed to IDENTIFIED state.

7. The reader turns back to the conflict point again, adjust the Request command to send Request (2, 0). It makes all tags with collision counter that are equal to "0" send the second data. And so on, until the T1 and T2 are identified from the NEW state changed to IDENTIFIED state.

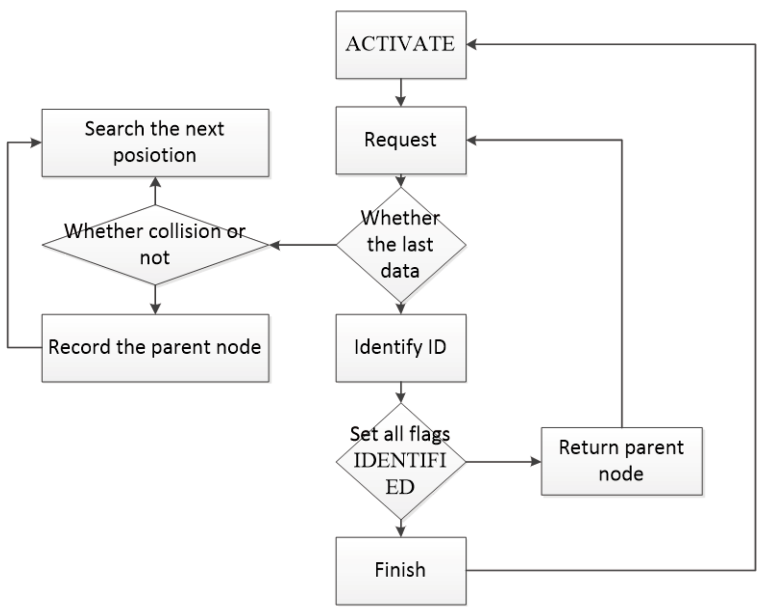

Fig. 1. Algorithm Description

8. All tags in the state NEW are identified, that is, all tags are set to state IDENTIFIED. The reader sends command ACTIVATE to determine whether there is a new tag to enter the reader's scope. If not, the whole process of identification is over.

Experimental results. ***Algorithm Implement.*** First, we give some definition as Table 1. We suppose that the algorithm time T and search number is a linear relationship.

Table 1

Definition

Name	Description
N	Tags in the reader's range
T	The total time required for the identification of N tags
12.5 us/bit	Time for sending tag ID

The total recognition time for a complete pruning branch algorithm can be defined as, us

$$T = (2N - 1) * M * 12.5.$$

The total time for N Tags of similar binary search algorithm can be evaluated by

$$T = N * M * t.$$

For IAC algorithm, we can evaluate the total time required to identify the N tags

$$T = \sum_{i=1}^{M-1} 2^i * t = (2^M - 1)t.$$

It is assumed that 256 eight bit tags need to be identified. First, we calculate the transmission time using pruning branches algorithm, ms

$$T = (2 * 256 - 1) * 8 * 12.5\text{us} = 51.1.$$

Secondly, we use the similar binary algorithm, ms

$$T = 256 * 8 * 12.5\text{us} = 25.6.$$

The improved algorithm needs the data transmission time, ms

$$T = (2^8 - 1) * 12.5\text{us} = 3.1875.$$

The recognition time of IAC algorithm is about 1/16 of the time of the pruning branch algorithm while it is about 1/8 of the time of the similar binary algorithm. The result shows that IAC algorithm has greatly reduced the recognition time.

Tag identification time. The bigger the tag numbers, the more optimized the identification time of IAC algorithm is. IAC algorithm has obvious performance advantage than the others. The identification of the three algorithms in the same conditions is shown in Fig. 2.

Search times. Comparing to the similar binary algorithm, IAC algorithm reduces the numbers of the command. If there is a set of 256 eight-bit tags need to be identified, the number of search times are 2048 for thesimilar binary algorithm. While that of IAC algorithm are 255, which greatly improves the efficiency of the query. The bigger the number of tags is, the more obvious the efficiency increase is. The result is shown in Fig. 3.

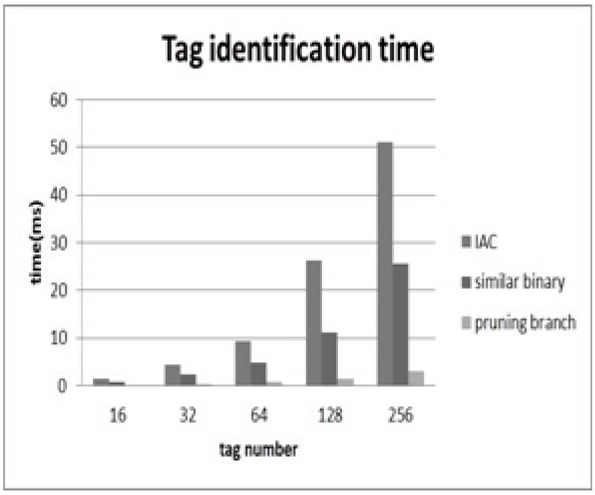

Fig. 2. The identification of the three algorithms in the same conditions

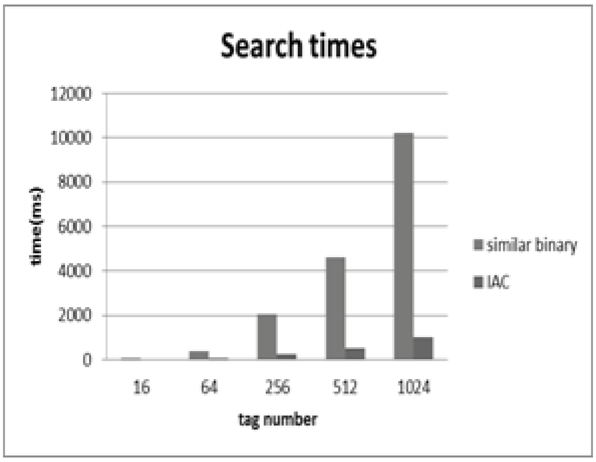

Fig. 3. The search times of the three algorithms under the same conditions

Conclusions. With the development of the IoT, the existing algorithms have been unable to meet the requirements of the RFID. We have analyzed the pruning branches algorithm and similar algorithm and developed a new anti-collision algorithm called IAC algorithm. It allows reducing the number of data in each time slot and the number of times and searches. The experimental results show that the algorithm is able to improve the performance of the system.

Acknowledgements. This work was supported by Scientific Research Project of Jiuji ang University (No. 2015LGYB28).

References / Список літератури

1. Hong, K., Lillethun, D., Ramachandran, U., Ottenwälder, B. and Koldehofe, B., 2013. Mobile fog: a programming model for large-scale applications on the internet of things.*Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing*, pp. 15–20.

2. Vicaire, P.A., Hoque, E., Xie, Z. and Stankovic, J.A., 2012, Bundle: a group-based programming abstraction for cyber-physical systems.*IEEE Transactions on Industrial Informatics*,vol. 8, no. 2, pp. 379–392.

3. Vujovic, V., and Maksimovic, M., 2014, Raspberry Pi as a wireless sensor node: performance and constraints. *Proceedings of 37th International Convention on Information and Communication Technology, Electronic and Microelectronics*, pp. 1247–1252.

4. Su, P.H., Shih, C.S., Hsu, J.Y.J., Lin, K.J. and Wang, Y.C., 2014. Decentralized fault tolerance mechanism for intelligent IoT/M2M middleware. *IEEE World Forum on Internet of Things*, pp. 45–50.

5. Laprie, J.C., 2008. From dependability to resilience. *Proceedings of 38th IEEE / IFIP International Conference on Dependable and Secure Network*, pp. G8–G9.

6. Chetan, S., Ranganathan, A. and Campbell, R., 2005. Towards fault-tolerant pervasive computing. *IEEE Technology & Society Magazine*, vol. 24, no. 1, pp. 38–44.

7. Cherrier, S., Ghamri-Doudane, Y.M., Lohier, S. and Roussel, G., 2014. Fault-recovery and coherence in internet of things choreographies. *IEEE World Forum on Internet of Things*, pp. 532–537.

8. Xu, J., Romanovsky, A. and Randell, B., 2000. Concurrent exception handling and resolution in distributed object systems. *IEEE Transactions on Parallel & Distributed Systems*, vol. 11, no. 10, pp. 1019–1032.

9. Kulkarni, D. and Tripathi, A., 2010. A framework for programming robust context-aware applications. *IEEE Transactions on Software Engineering*, vol. 36, no. 2, pp. 184–197.

10. Tariq, M.U., Florence, J. and Wolf, M., 2014. Design specification of cyber-physical systems: towards a domain-specific modeling language based on simulink, eclipse modeling framework and giotto. *Proceedings of the 7th International Workshop on Model-based Architecting*, pp. 6–15.

Мета. Інтернет речей (IoT) являє собою майбутній напрям розвитку комп›ютерної та комунікаційної технології, що вважається третьою хвилею розвитку в області інформаційної індустрії після появи комп’ютера. IoT є втіленням інформаційної системи мережі фізичних об’єктів, що функціонує в режимі реального часу на основі радіочастотної ідентифікації (RFID) і електронних кодів продуктів (EPC). Надалі всі види існуючої технології будуть стикатися з багатьма новими можливостями та викликами, особливо RFID.

Методика. У вітчизняних і зарубіжних дослідженнях проблеми колізії міток існує два напрями: двійковий антиколізійний алгоритм обходу дерева, і антиколізійний алгоритм ALOHA на основі тимчасових інтервалів. Однак, алгоритм ALOHA швидко застаріває, і тому не підходить для широкомасштабного застосування в IoT.

Результат. Серед бінарних антиколізійних алгоритмів обходу дерева, найбільш розвиненими є алгоритм видалення гілки дерева (алгоритм обрізки) і алгоритм подібності.

Наукова новизна. Розроблено покращений антиколізійний алгоритм (IAC), що дозволяє не тільки зменшити кількість даних у кожному часовому інтервалі, але й зменшити кількість спроб пошуку.

Практична значимість. Результати тестів показали, що розроблений покращений антиколізійний алгоритм дозволяє підвищити продуктивність в порівнянні зі стандартними алгоритмами обрізки й подібності. Більш того, покращений антиколізійний алгоритм дозволяє істотно скоротити тривалість пошуку.

Ключові слова: *RFID, двійковий антиколізійний алгоритм обходу дерева, алгоритм ALOHA*

Цель. Интернет вещей (IoT) представляет собой будущее направление развития компьютерной и коммуникационной технологии, которое считается третьей волной развития в области информационной индустрии после появления компьютера. IoT является воплощением информационной системы сети физических объектов, функционирующей в режиме реального времени на основе радиочастотной идентификации (RFID) и электронных кодов продуктов (EPC). В дальнейшем все виды существующей технологии будут сталкиваться со многими новыми возможностями и вызовами, особенно RFID.

Методика. В отечественных и зарубежных исследованиях проблемы коллизии меток существует два направления: двоичный антиколлизионный алгоритм обхода дерева, и антиколлизионный алгоритм ALOHA на основе временных интервалов. Однако, алгоритм ALOHA быстро устаревает, и поэтому не подходит для широкомасштабного применения в IoT.

Результат. Среди двоичных антиколлизионных алгоритмов обхода дерева, наиболее развиты алгоритм удаления ветви дерева (алгоритм обрезки) и алгоритм сходства.

Научная новизна. Разработан улучшенный антиколлизионный алгоритм (IAC), который позволяет не только уменьшить количество данных в каждом временном интервале, но и уменьшить количество попыток поиска.

Практическая значимость. Результаты тестов показали, что разработанный улучшенный антиколлизионный алгоритм позволяет повысить производительность в сравнении со стандартными алгоритмами обрезки и сходства. Более того, улучшенный антиколлизионный алгоритм позволяет существенно сократить продолжительность поиска.

Ключевые слова: *RFID, двоичный антиколлизионный алгоритм обхода дерева, алгоритм ALOHA*

Рекомендовано до публікації докт. техн. наук В.В.Гнатушенком. Дата надходження рукопису 07.05.15.