International Journal of Security and Its Applications
Vol. 9, No. 5 (2015), pp. 325-334
http://dx.doi.org/10.14257/ijsia.2015.9.5.32

# A Monitoring Algorithm Based on IP Address Statistics in DCS Isolation Device

Yongguang MA and Zhiyu YE

*North China Electric Power University, Baoding, China*
*Mr_ma@163.com, 11860542@qq.com*

***Abstract***

*Distributed Control System (DCS) requires communication with external system, which probably causes the invasion of the viruses and malicious programs, finally resulting in the fault of units. One-way physical isolation device is efficient to solve the current threat but not aware of the potential danger. In this paper, a monitoring algorithm is introduced to predict and detect the abnormal communication based of IP address statistics, thus increasing reliability and flexibility of the communication.*

***Keywords:*** *DCS, IP address statistics, isolation device, monitoring algorithm*

## 1. Introduction

With the wide use of Distributed Control System (DCS) in thermal plants, most of the real time data has to be transmitted to the external system for processing, storage and analysis. However, the external system, which mainly consists of Supervisory Information System (SIS) and Management Information System (MIS), is connected with the Internet, threatening the internal operating system with viruses and malicious software [1, 2]. So the data transmission between them should be monitored closely, to analyze every potential danger and assure the safety of DCS.

The physical isolation device between DCS and SIS can guarantee the security, but the detailed information about corresponding cannot be gained. It appears to be not so flexible to manage with no monitoring measure [3]. The potential danger can never be predicted. When it goes wrong, the whole system is in danger. Meanwhile, the upper monitor of data sending information is a efficient way to foresee and relieve the danger.

How to monitor such large quantities of Ethernet data frames with various information and lengths? Anti-virus firewall is not enough, because it is basically a kind of "passive defense", which can possibly prevent the known viruses and conventional attack only with upgrading software and hardware continuously [3, 4]. If an upper monitor algorithm is based on statistics of corresponding IP addresses, which makes it common and active to defend the communication launched by abnormal IP addresses.

## 2. Design of Monitor Algorithm

### 2.1. Design Requirements Analysis

When DCS is communicating to SIS, the Destination Address (DA) and Source Address (SA) should be told to the upper monitor. High statistical probabilities of IP ports are trusted and low ones are mistrusted. If there are additional communication ports, the algorithm firstly mistrusts it and gives the alarm, but if the new ones are reused again and again then, the algorithm should memorize this shift and gradually start to trust them by self-learning progress [5]. However, for those IP ports that are now rarely used but often used in the past, the algorithm should also give alarms. So the monitoring algorithm

ISSN: 1738-9984 IJSIA

should give every single decision after a strict process of learning, memorizing and trusting [6, 7].

### 2.2. Algorithm Constraints

In order to calculate all the statistics of the effective IP Addresses the requirements mentioned above can be concluded into the following two constraints.

**Constraint 1**. IP Address Trust and Elimination Rule

While DCS is sending data to the SIS, there would certainly be IP ports that are frequently used and infrequently used. This algorithm aims to set the degree of statistical confidence by the use frequency of each IP port. Then the statistical confidence is divided into three levels, trust, doubt and mistrust, which is the basis for the algorithm to give the conclusion.

After setting parameter 'N', the system trustworthy IP port number, the program will intercept IP addresses from the acquired Ethernet data frame, and IP addresses will be sorted by the use frequency. Then the algorithm will judge whether is proper to let the data frame to pass the isolation device [8]. Taking N=4 as an example, the IP addresses are arrayed in the following table in descending order of statistical confidence.

**Table 1. IP Addresses and Corresponding Conclusion (N=4)**

| IP Addresses (HEX) | Times of occurrence | Rank | Description | Trust level | Corresponding Measure |
|---|---|---|---|---|---|
| 4F8020010640 | 8 | 1 | Top N | Totally Trust | Allowed |
| 4FA0200104B0 | 7 | 2 | | | |
| 4FA020010578 | 5 | 3 | | | |
| 4FC0200103E8 | 3 | 4 | | | |
| 4FE01F010190 | 2 | 5 | Top 2N | Doubt | Allowed but Warning |
| 4FE020010320 | 2 | 6 | | | |
| 50001F0100C8 | 1 | 7 | | | |
| 500020020000 | 1 | 8 | | | |
| 502024000000 | 1 | 9 | -- | Mistrust | Alarm and Waiting for Check |
| 08C01E010578 | 0 | 10 | | | |
| 08E01E0104B0 | 0 | 11 | | | |
| 08E01E010708 | 0 | 12 | | | |
| 09001E010640 | 0 | 13 | | | |
| 09001E0107D0 | 0 | 14 | | | |
| 09401E0203E8 | 0 | 15 | | | |
| 09001E0107D0 | 0 | 16 | | | |

IP addresses of Top N are considered to be a trusted address, and the system let the frame with these IP addresses pass and does not take any other actions. IP addresses of ranking N+1 to 2N are doubted, and the system releases them with a warning though. IP addresses out of Top 2N are assumed to pose a threat to the system, which will be placed in a waiting queue to confirm before admission.

 

**Constraint 2**. Trustworthy Time Limit Rule

Since DCS to SIS data transmission appears the following: an IP port was used frequently, but due to a number of factors, it is now replaced by other communicating IP ports [9]. If only in accordance with the Guideline 1, the IP port which has been idle for a long time will still be trusted. Hence, the program needs a parameter, "M", to limit trustworthy period that can discard all the useless statistics, as is shown in Figure 2.

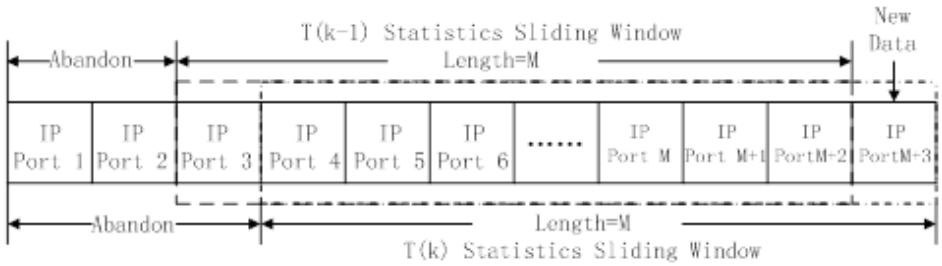


**Figure 1. Sliding Window of Contributing Statistics**

The sliding window, M data units long, makes IP statistics valid only inside and ignores the outmoded data [10, 11]. Therefore, all IP port addresses will be recorded and arrayed strictly according to the times of occurrence, so there will possibly be mistrusted ones promoted to be trusted as well as the elimination for trusted ones due to the dynamic statistics. In this way, the algorithm can determine the threat level of the data frame according to the Constraints 1 and 2.

### 2.3. Modular Algorithm Design

**Module 1**. Initialization Module

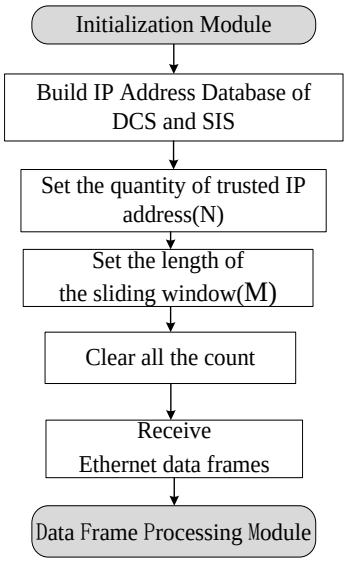


**Figure 2. Initialization Module**

In initialization module, firstly all the possible IP addresses of DCS and SIS systems should be included, while the number of trusted IP addresses(N) and sliding window of length m (valid statistical data for this) need to be set. And then the times of every IP address occurrence is set to 0, ready to receive Ethernet data frames.

**Module 2**. Data Frame Processing Module

In the Ethernet frame format, the first 7 numbers of the frame is the preamble, the 8th number is the starting delimiter, the 9th -12th is destination address DA, and 13th -16th is the source address SA, thus extract of DA and SA is very easy to operate. Then the algorithm is interested in the corresponding between DCS and SIS, so this module will get the DA that is from DCS/SIS to SIS/DCS. If so, it goes on to the next module; if not, it is neglected.

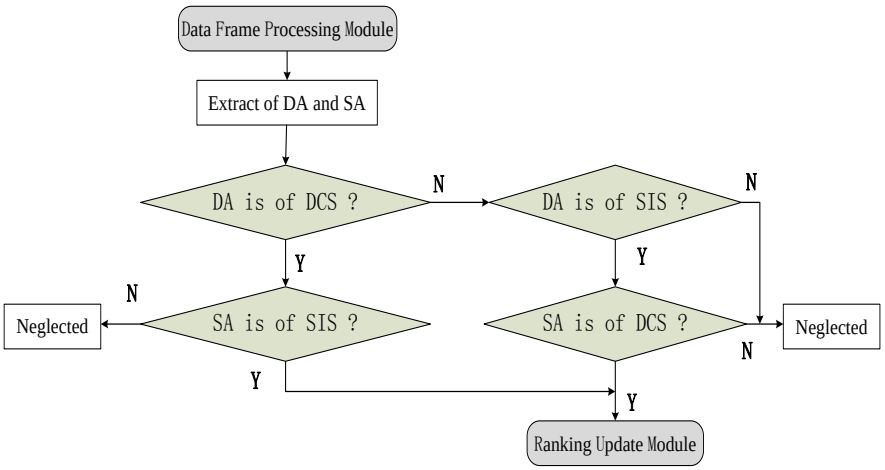


**Figure 3. Data Frame Processing Module**

**Module 3**. Ranking Update Module

This module combines the Constraint 1 and 2. Within the sliding window length M at the beginning of statistics, the sliding window does not move. Once the statistical frequency exceeds M, the window starts to move when new IP address is counted. To ensure timeliness and effectiveness of the statistics, the statistical data out of sliding window will be discarded. On the other hand, the occurrence number of IP addresses statistical tables in each new increase in the number of IP will plus one, and then the data is re-ranked.

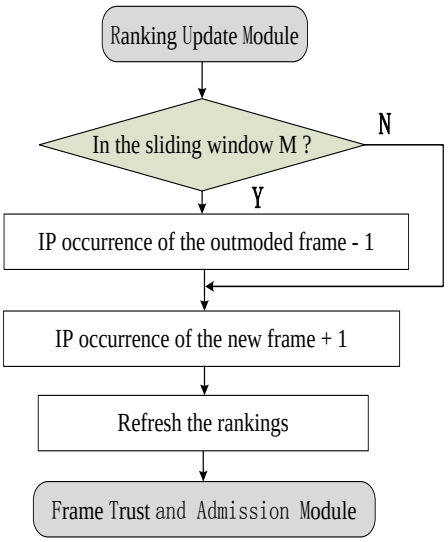


**Figure 4. Ranking Update Module**

**Module 4**. Frame Trust and Admission Module

In this module, according to the occurrence number of data frames of its DA and SA, the upper PC will determine the trust level of the frames, which leads to different measure as shown in Table 1.

1) IP address sorted within the first Nth is considered to be a trusted address, and the system does not take any action;

2) IP address sorted from $(N+1)^{th}$ to $(2N)^{th}$ is thought to be a doubtful address, the system will give a warning, but it will be released ;

3) IP address outside $(2N)^{th}$ is considered to be the new IP address , it may pose a threat to the system, and before being released, it needs to be confirmed.

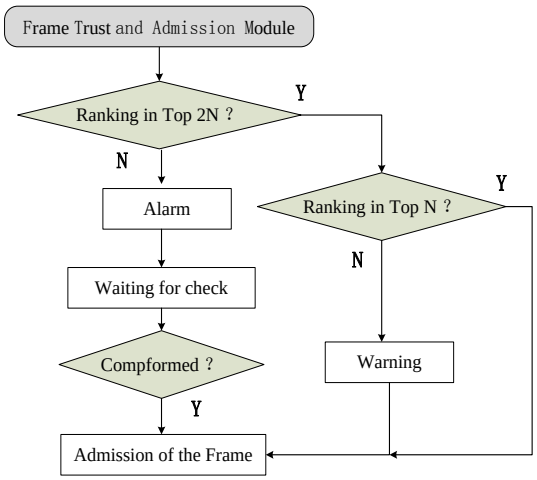


**Figure 5. Frame Trust and Admission Module**

## 3. Algorithm Test

Firstly, IP Address Database is necessary in algorithm test to be the template of statistics. Then we need a set of representative data frames to test all the functions and bugs in it. With expected result based on theoretical analysis, the test result can be contrasted accordingly and finally we can decide the applicability of this algorithm.

### 3.1. IP Address Database

**Table 2. All IP Address (Possible DA) in Corresponding between SIS and DCS**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 502024000000 | 5 | 4FE020010258 | 9 | 4FA020010578 | 13 | 08C01E010578 |
| 2 | 500020020000 | 6 | 4FE020010320 | 10 | 4F8020010640 | 14 | 09001E010640 |
| 3 | 50001F0100C8 | 7 | 4FC0200103E8 | 11 | 09401E0203E8 | 15 | 08E01E010708 |
| 4 | 4FE01F010190 | 8 | 4FA0200104B0 | 12 | 08E01E0104B0 | 16 | 09001E0107D0 |

Taking a small power plant as a case, we get the IP address of all devices in DCS and SIS to the table above, so that the communication between the DCS and SIS must be completed between two of them. Then all the DA can be directly compared with the ones in the table, after which the situation of communication is gained clearly.

### 3.2. The Set of Test Data Frame

We should select a set of data to test the algorithm performance in detecting the threat and trusting the safe ones. So the set should contain different parts, where the DA varies in a given law, so that the test performance can be compared to the expected result.

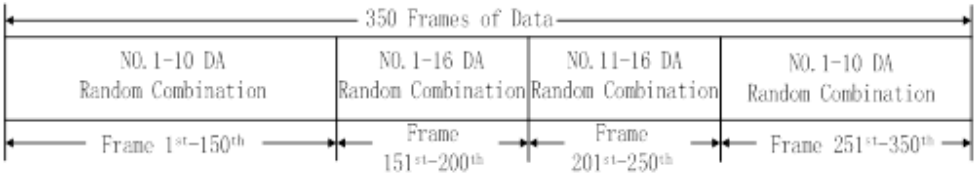


**Figure 6. Composition of Test Frame Set**

First 150 groups of frames are with 10 different DA, the following 50 groups with all 16 possible DA in the IP database, the next 50 groups with the other 6 DA and finally the last 100 groups with the original 10 DA. It would be 350 frames altogether and the capability of this algorithm can be shown in an obvious way.

### 3.3. Expected Performance

The expected result should be divided into 4 parts: Steady trusting performance in first 150 groups, very frequent threats in the following 50 groups(especially at the beginning), fair usual threats in the next 50 groups(rather less than before) and back to steady with seldom threat alarms.

## 4. Test Result Analysis

Set N=4 (the number of trusted IP addresses), then the number of doubted IP address (2N) is 8, and set the length of the sliding window (Trustworthy Time Limit) M=24. Get 350 groups of data frame, sent from DCS to SIS, to test the algorithm.

### 4.1. Overall Analysis for Data Frame Group

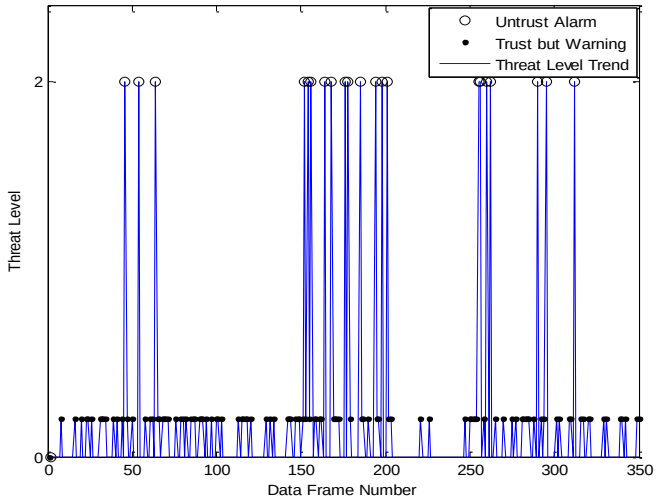


**Figure 7. Ethernet Frames Monitoring Algorithm Test Results**

As Figure 7 shows, the data frames are classified to be trusted, doubted and mistrusted as different threat level. And every detailed analysis is as following:

1) Among the first 150 groups of data, only at the 46th, 54th and 64th give alarms (mistrusted address appears). The sliding window length is 24, and before statistics window has not been sliding before that, so no alarm occurs until then. On the other hand, the first 150 sets of data are composed of random pick from the 1st -10th IP ports, and trust IP ports number N is 4, so the probability to get confidence is relatively high, and the test result in this part is as expected.

2) The frequent alarms are given between the 150th -180th frames, for the 151st -200th groups groups of data are random picks from the 1st-16th IP addresses, a combination of 16 random numbers. While N=4, which is much smaller than 16, the probability to trust is relatively low.

3) The 201st -250th frames, of which the DA are picked from the 11th-16th IP ports, are more likely to get trusted, so the alarms are rare during this period. Only when the corresponding IP port is just switched and the algorithm is still learning, there would be the alarm.

4) In the 250th -350th group of data frames, DA turn into the 1st -10th again, so the performance is similar to the first 150 sets of data, with a higher probability to trust. When the IP address is suddenly switched, the algorithm should recount the new IP addresses, so there are many alarms at the beginning of this part.

### 4.2. Individual IP Port Analysis

Then we focus on each IP Port, taking one of the IP addresses, the 10th Destination Address, 4F8020010640 (HEX), for instance, to analyze. In this way, the relationship between the trust level and times of occurrence is even clearer, as shown in Figure 8.

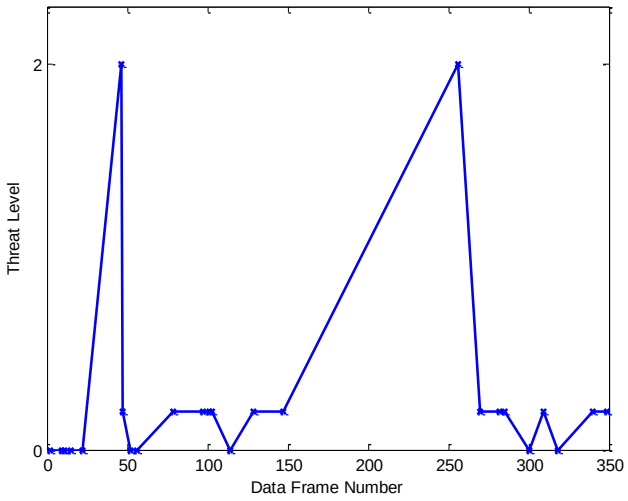


**Figure 8. Monitoring Test Results for Data Frame with the 10th Destination Address**

Originally, in the first 22 groups, the frames with the 10th Destination Address occurs so frequently that algorithm trusted this IP. While the 10th DA has not appeared through the 23rd- 46th frames, it is after the limited trust time (M=24) when the 46th data frame arrives. Because the sliding window has already abandoned the first 22 groups of statistics, the algorithm no longer trusts the IP address.

But the 10th DA comes up again in the 47th frame, the algorithm becomes to doubt instead of mistrusting it. Then it shows up again in the 52nd and 56th frames, algorithm begins to trust. That is because the statistics at the 46th and 47th frames are still valid in the sliding window and the algorithm gradually learned. The algorithm test results can be analyzed in the same way as above and the same conclusion can be gained.

### 4.3. Parametric Analysis of Rationality

Comparing the test data, the algorithm performance is bad in the 150$^{th}$ -180$^{th}$ frames, where alarms are frequent, which is because the types of IP address in use have changed from the original 10 into 16, and the selection of parameter N and M in this group is not appropriate any more. In that case, the rationality of the parameters N and M is connected with the quantity of IP addresses in use [12].

After debugging and analyzing for several times, N is decided by the quantity of SIS corresponding ports in use and the empirical formula for parameters N and M is given as following.

$$5N \leq M \leq 6N \tag{1}$$

## 5. Conclusion

Eventually, we can conclude the monitoring algorithm into a chart. The three plots, ranking the first, the N and 2N, divide the space into three passband. The light-colored passband above is the completely trust region, where the system does not take any action. And the dark passband in the middle is the doubted area, where the system releases data frame but gives warning. The passband at the bottom is mistrusted by the system, where the data frames will be intercepted and put into waiting for verification. So the standard of trusting and mistrusting is floating, relying on the statistical condition. Whenever the data frame comes, the algorithm can figure out a scheme to deal with it.

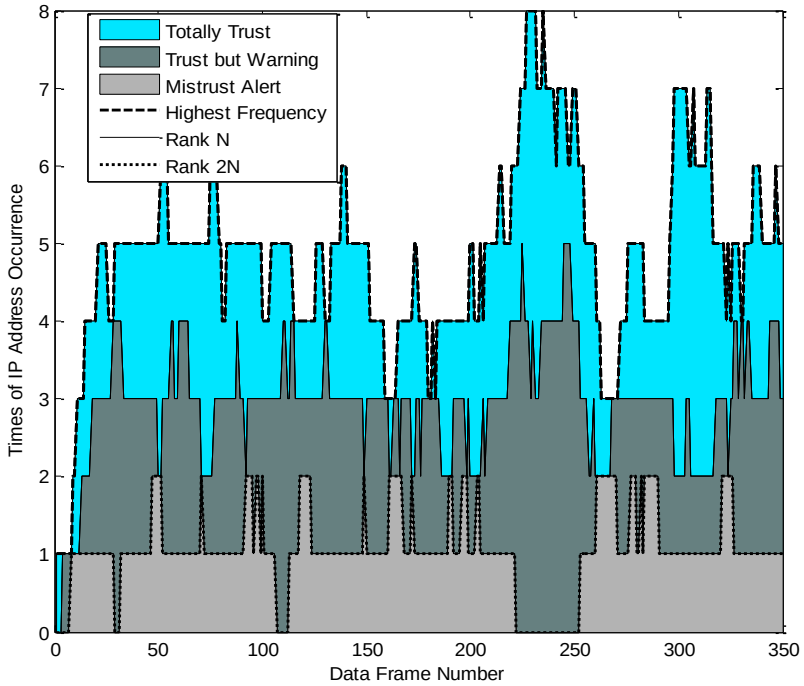


**Figure 9. Monitoring Algorithm Conclusion Based on the Statistical Frequency**

Therefore, this monitoring algorithm has a capability of self-organization and self-learning. Meanwhile, the effectiveness of the statistics is strictly limited and finally the threat warning function is achieved accurately. And this algorithm can be applied into different types of corresponding areas and make it much safer and more reliable.

## References

[1] E. Chong, "Smart Gateway Systems for Internet Security for Broadband Communication Networks", International Conference on ASIC Proceedings Book 2 of 2, Chinese Institute of Electronics (CIE), **(2003)** April.
[2] Z. Hanwei, J. Bo and X. Chunfu, "DCS and MIS Security Isolation Technology", Hebei Electric Power Technology, vol. 03, **(2004)**, pp. 44-46.
[3] Y. Zhou and Y. Wang, "Intrinsically Safe DCS Data Isolation System and Its Application in Industry", Safety and Environment, vol. 07, **(2011)**, pp. 20-23.
[4] B. Sun, J. Feng and T. Lin, "Key Laboratory of Microelectronic Devices and Circuits", Institute of Microelectronics, Peking University, 100871, P. R. China. A New Configuration Scheme for Delay Test in Non-simple LUT FPGA Designs [A], **(2008)** April.
[5] Z. X. Fu and Q. D. Wen, "Three-step semiquantum secure direct communication protocol", Science China, **(2014)** September, pp. 1696-1702.
[6] "Efficient multi-user detector based on box-constrained deregularization and its FPGA design", Journal of Systems Engineering and Electronics, **(2012)** February, pp. 179-187.
[7] "Improved eavesdropping detection strategy based on four-particle cluster state in quantum direct communication protocol", Chinese Science Bulletin, vol. 34, **(2012)**, pp. 4434-4441.
[8] Z. Li, S. Zhang, J. Lang and H. Shao, "The Application and Research of the Liquid Level Control Technology Used in Mineral Flotation Process which Based on the Modbus Communication Protocol", **(2013)** April.
[9] G. Chen, Y. Du, P. Qin, J. Du and N. Li, "Petri Net Based Research of Home Automation Communication Protocol", **(2013)** November.
[10] M. Qiang, Z. Jianguo and Liubingxu, "Implementation of Embedded Ethernet Based on Hardware Protocol Stack in Substation Automation System", Transactions of Tianjin University, vol. 02, **(2008)**, pp. 153-156.
[11] "Improved eavesdropping detection strategy based on four-particle cluster state in quantum direct communication protocol", Chinese Science Bulletin, vol. 34, **(2012)**, pp. 4434-4441.
[12] M. Jianguo, S. Yubo, X. Ling and M. Qiang, "Design of UCL Hardware Filtering System Based on FPGA", Technical Committee on Control Theory, Chinese Association of Automation, **(2010)** April.

## Authors

**Yongguang MA**, 1964, male, professor, master instructor, mainly engaged in the research of power production process simulation technology and automation technology.