A Study of Real-time Data Encryption in the Smart Grid Wide Area Measurement System based on Storm

Shaomin Zhang[1,a], Jie Sun[1,b],Baoyi Wang[1,c]

[1]School of Control and Computer Engineering, North China Electric Power University, Baoding,071003, China

[a]email: zhangshaomin@126.com, [b]email:sunjiedodo@163.com, [c]email: wangbaoyiqj@126.com

Keywords：wide area measurement system(WAMS); cloud computing; Storm; data encryption

Abstract. In order to solve the problem of data security in smart grid wide-area measurement system(WAMS),this paper analyzes the characteristics of the existing data encryption solutions, and combines with the characteristics of massive data as well as high requirement of real time in WAMS, then introduces Storm distributed real-time platform to process real-time encryption of large-scale data. We design WAMS data storage system based on cloud platform, and on top of this system design AES algorithm based on Storm. The algorithm is mainly divided into three steps: data collection, fast parallel data encryption and cloud storage of encrypted results. The procedures are implemented in the Storm predefined programming components. The experimental comparisons verify that the proposed algorithm has the characteristics of low latency and high encryption efficiency.

Introduction

Wide area measurement system (WAMS),as a power grid dynamic monitoring technology platform, is an important part of smart grid real-time monitoring, which can realize real-time high-speed data acquisition from synchronous phase angles and each site of the grid[1].

In recent years,data from wide area measurement system tends towards high sampling rate, continuous steady state record and large storage, besides the sampling rate can reach 100 times/s. If we calculate the result according to sampling frequency 100 times/s, then the system will generate 459 TB data per year (459 TB=16 bytes/ /s*100 frames/s *86400 s *365 days). When the number of sub unit PMUs in the WAMS increases, more data will be generated[2],[3],[4]. In this context, the centralized data processing and storage model will cause great pressure on the network, and it is a necessity to use distributed processing and storage. At the same time, because of the demand of the power grid stability and data itself having the need for redundancy backup, the parallel processing model and distributed storage platform of cloud computing are suitable for future smart grid's need of dispersion, reliability, security and mass data[5]. Besides,because power companies are not allowed to use the power system data publicly[6], the security of data is the most important problem of the wide area measurement system's application.

The encryption of data on the cloud platform and encryption algorithm have been studied in Many papers. According to the characteristics of cloud computing, paper[7] redesigns the traditional encryption algorithm, which combines the traditional DES algorithm with RSA algorithm, and improves the security. Paper [8] combines and improves the implementation of the internal operation of the AES algorithm, reducing the delay, but in order to reduce the relevance of the data block, this paper uses ECB packet encryption mode, which is vulnerable to replay attacks. Paper [9] designs and implements a distributed RSA algorithm based on Hadoop platform, which improves the efficiency of encryption. However, in this paper，the plain data set is just divided into ordered plain text fragments, and the plain text fragments are encrypted separately, then the encryption results are combined to generate the cipher text. As a result, this method is also vulnerable to replay attacks. Paper [10] designs an security storage scheme based on Hadoop of smart grid data, which uses AES encryption algorithm to improve the security. But the MapReduce method in Hadoop belongs to the batch processing mode of persistent data. In each process, if the

system waits until the batch reaches a certain size, it will increase the processing delay and still not meet the real-time needs[11], so the Hadoop platform is not suitable for high-speed data stream real-time encryption.

In this paper, we apply the Storm platform to real-time encryption of wide area measurement system data, and design the AES encryption algorithm based on Storm, which real-time encrypts and processes the mass high speed wide area measurement system data.

Related technology

Description of Storm platform

Storm is a distributed real-time computing system which is free open source, high fault tolerance. It has the characteristics of high low latency, scalability, high reliability, multi-language support etc.[12] Storm is commonly used in real-time analysis, online machine learning, continuous computing, distributed remote invocation and ETL etc.

Storm is mainly divided into two parts: Nimbus and Supervisor [12]. The structure of the Storm cluster is as shown in figure 1.

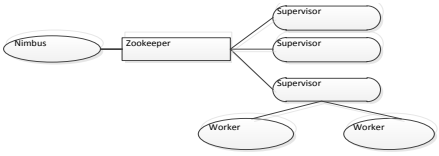


Fig.1 The architecture of Storm cluster

(1) Nimbus is responsible for sending code in the cluster, assigning jobs to machines, and monitoring the status. There is only one globally.

(2) Supervisor is responsible for monitoring the work of the machine that is assigned to it, and starting or closing the work process Worker according to the need.

(3) Zookeeper is an external resource that Storm mainly depends on. Nimbus, Supervisor and Worker store their heartbeats in Zookeeper.

(4) In the Storm cluster, the real-time processing of data stream is packaged as a Topology for publication. As shown in Figure 2, Topology is composed of Spout and Bolt. Spout represents a main data entry of Storm Topology, acting as a role of collector, which connects the data source, and transforms the data into a Tuple, then transmits Tuple as a data stream. Bolt treats one or more data streams as input, and after operating the data, outputs one or more data streams selectively. Bolt can perform the typical functions including filtering Tuple, join and aggregation, computing as well as database reading and writing.

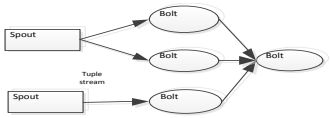


Fig.2 The architecture of Topology

Description of AES algorithm

AES is a symmetric encryption algorithm. Symmetric encryption algorithms encrypt and decrypt using the same key, so the encryption speed is high.

The characteristic of AES algorithm is that the algorithm is safe, efficient and simple. It has excellent resistance to differential and linear analysis. The core of the algorithm is the round transformation, and it can get a strong security with less number of rounds[13]. AES algorithm is symmetric encryption algorithm, of which packet length is 128 bit, and the key length is 128 / 192 / 256 bit. According to the different key length, the number of rounds of data encryption and decryption is also different.

AES takes the input blocks as a 4 x 4 byte matrix (the state matrix), and then impose the matrix a different transformation. AES encryption algorithm mainly consists of three parts[7]:

(1) XOR with initial cyclic sub key

(2) N_{r-1} rounds cyclic encrypting

(3) the last round encrypting

The decryption is the inverse process of the encryption.

Selection of AES encryption mode

Interleaved CBC (ICBC) mode is introduced to realize the parallelization of encryption algorithm[15]. ICBC mode can generate multiple interleaved encrypted data streams rather than one[8].For example, in two way interleaved chaining, the first, third, and every second block thereafter are encrypted in CBC mode; the second, forth, and every second block thereafter are encrypted as another stream, and so on. The processes of encryption and decryption both require two initial vectors(IV). Thus, ICBC mode can overlap multiple encrypted streams in order to realize the parallel of block encryption algorithm. The decryption of ICBC mode is similar to its encryption, which also supports multi-thread and parallelism. AES encryption in two-way ICBC mode is shown in Figure 3.

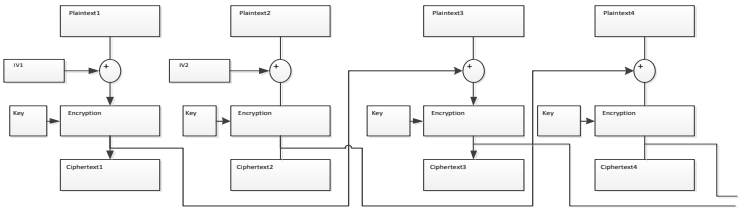


Fig.3 AES encryption in two-way ICBC

It can be seen from Figure 3 that ICBC mode supports the parallel implementation of AES algorithm, which can guarantee encryption speed is similar to that in ECB mode, and its security is no less than that in CBC mode.

Real-time encryption design of WAMS data based on Storm

Firstly, WAMS data storage system based on cloud platform is designed. After WAMS data is transferred into the system, the data is processed into three steps:

(1) The data files are processed by generating indexes according to the station and the time of generation, which is helpful to the quick inquiry of the later use.

(2) Data files are encrypted on Storm platform. To enhance the security， data files from different station use different secret key K_1.

(3) The encrypted files are stored into the HDFS on Hadoop platform after sorting. At the same time, the key information with file indexes is stored in HBase after Hash. The overall design of the system is shown in figure 4.

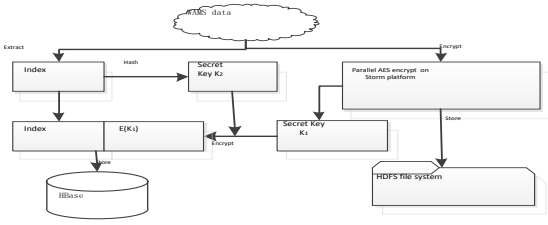


Fig.4 WAMS data storage system based on cloud platform

WAMS data are parallel encrypted using AES in ICBC mode on the Storm platform. The process of parallel encryption is mainly divided into three steps: data access, fast parallel data encryption and storage of the encrypted results. Real time encryption of data is shown in Figure 5.

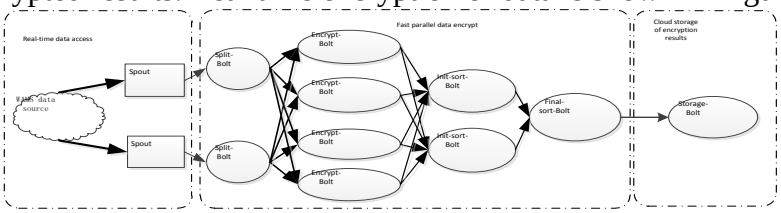


Fig.5 Real-time data encryption process

(1)Data access

In Storm, Spout is data source, Spout reads data from external real-time data source, and then converts the source data into Topology. Spout access WAMS data through IRichSpout interface ,

and Storm send a Tuple by invoking nextTuple().In order to achieve reliable message processing, unique ID is given to each Tuple, and the ID is passed to emit() of SpoutOutputCollector as a parameter.

If this Tuple is not successfully processed by Storm, Spouts will emit a new Tuple; Storm will invoke ack () and confirm the message response when it has detected that the Tuple is successfully processed by the Topology, otherwise invoke fail ().

(2)fast parallel data encryption

All message processing logic of Storm is encapsulated in Bolts. Split-Bolts are responsible for receiving Tuple sent by Spout , and diverse the Tuple. The plaintext from Spouts are split into four word length Block , then in accordance with the rules of ICBC mode the small plain blocks are divided, and finally the divided groups are sent to encrypt-Bolts. Encryption logic is written in execute () of the Bolt. In order to guarantee the sequence of the processed data, the system has assigned each Tuple a unique ID before the execution, when the encryption process is over, the data is sent to the init-sort-Bolts, and is sorted initially according to the IDs; Finally, data is sent to the final-sort-Bolts to be sorted finally and output.

(3)storage of the encrypted results

As Storm platform itself is not responsible for the preservation of encrypted results, so access to storage-Bolt after final-sort-Bolt for data storage. In view of the l data volume, this paper designs that the data will be stored in HDFS on Hadoop platform.

Analysis of examples

The nodes of the cluster are equally configured, each machine has 2G memory, 160G hard disk space, and is installed Hadoop-2.2.0 and Storm-0.8.2, where Hadoop platform has one Master node and three Slave nodes, Storm platform has one Nimbus node and six Supervisor nodes, through a high-speed switch to form the internal network. AES encryption algorithm is used in the experiment, and the key length is 128 bits.

Instructions of WAMS data source, key, and cipher text storage

(1)Data structure of WAMS data source

WAMS acquire data through PMUs installed in stations, and data files' sampling frequency is 100 frames /s. Each data file records the 60-second data before and after disturbance.

(2)Key encryption storage

The file index extracted is hashed by the Hash function to generate a fixed-length key K_2, and then make K_2 encrypt the symmetric key K_1 used on Storm platform, and finally store the file index and encrypted $E(K_1)$ in HBase. In order to ensure security, the Hash function should be kept confidential. Symmetric key information hiding process is shown in figure 6.

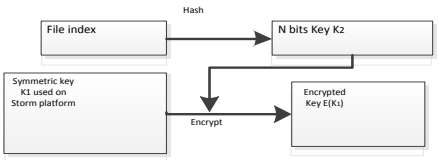


Fig.6 Symmetric key information hiding process

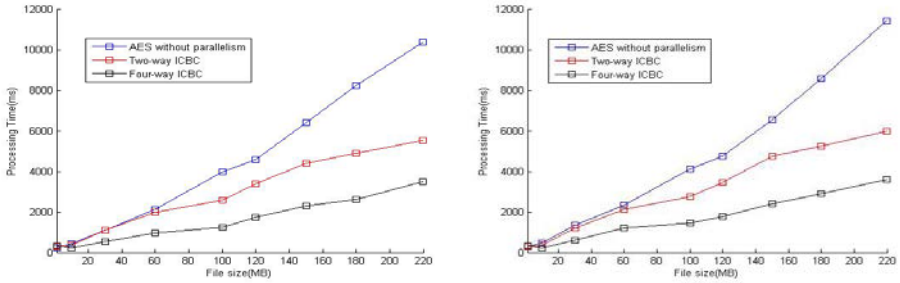
HBase stores the file's indexes and encrypted symmetric key. There are three columns in the table, which are RowKey, Timestamp and Contents. RowKey is the unique identifier of the data row in the table, and is the primary key of the HDFS. Timestamp correspond the time associated with each data operation ,and is automatically generated by the system. Contents currently contains a EncryptedKey tag for storing encrypted symmetric keys. The logical view of the index and key data storage in HBase is shown in table 1.

Table 1 Logical view of index and secret key stored in HBase

RowKey	Timestamp	Column"Contents"
FileIndex		Contents：EncryptedKey="…"

Experiments and analysis of data encryption and decryption on Storm platform

In order to test the performance of the scheme, we experiment different size data files, and the data file size is between 20M and 220M. Figure 7 (a) (b) separately show the time performance of AES algorithm in CBC and two-way ICBC and four-way ICBC on Storm platform.



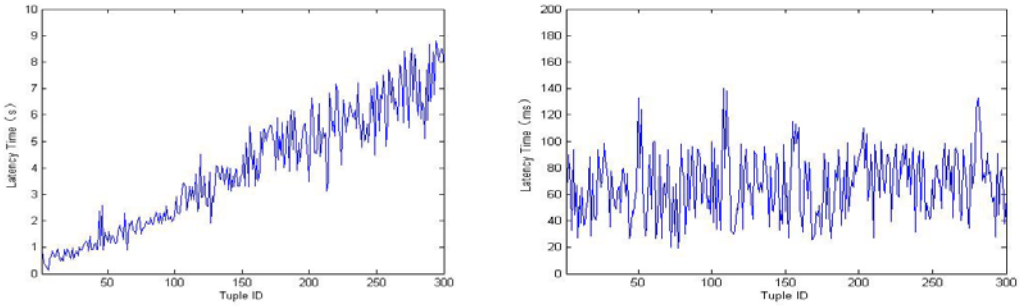
(a)Encryption (b)Decryption

Fig.7 Comparison of AES encryption and decryption processing time in three different modes

Figure 7(a) (b) show that when using AES in ordinary CBC to encrypt or decrypt, with the file size increasing, processing time of encryption and decryption has been a linear growth. However, AES in ICBC mode achieve a significant increase in data encryption and decryption speed. Compared with two-way interleaved ICBC ,encryption or decryption speed of AES in four-way interleaved ICBC has been improved.

Experiments and analysis of data processing delay time on Storm platform

In this experiment, the delay time of data processing = the end moment of WAMS data is processed in final-sort-bolt –the moment of data is emitted. Setting the number of Spout components is two, and the number of Bolts that are used to encrypt or decrypt using AES in parallel is separately 2 and 6, and the data processing delay of two tasks is recorded. Results are shown in Figure 8(a) (b).



(a)The number is 2 (b)The number is 6

Fig.8 The delay when the number of Spout is 2 or 6

Figure 8 (a) (b) show that when the number of tasks is two, due to processing speed is far less than that of generating data , pending data continue to accumulate, causing that the delay of processing increasing; However when setting the number of tasks is 6, most of the delay is between 20ms and 100ms, satisfying WAMS data ‘s delay requirements 100ms~1s ,which belongs to the controllable scope. Analysis suggests that, occasional larger delay is because Spout at one point gained more resources to send more resources, but the multiple tasks can in a short time get rid of these resources, making the delay recover a controllable range.

Conclusion

According to the characteristics of WAMS data, we designs a WAMS data storage system based on cloud platform. On the basis of this system, a fast data encryption scheme based on Storm is proposed, which is based on the Storm framework. The data processing process is divided into real-time data access, fast parallel data encryption, encryption results cloud storage. Relevant experiments demonstrate the effectiveness of the design scheme.

Acknowledgement

In this paper, the research was sponsored by the National Natural Science Foundation of China (Project No. 61300040) and Scientific Research Project of Hebei Province (Project No. Z2012077).

References

[1] Qu Zhaoyang, Zhu Li, Zhang Shilin. Data processing of Hadoop-based wide area measurement system[J]. Automation of Electric Power Systems,2013,04:92-97.

[2] Song Yaqi, Zhou Guoliang, Zhu Yongli.Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37(4):927-935

[3] Song Yaqi, Liu Shuren, Zhu Yongli, Wang Dewen, Li Li. Cloud storage of power equipment state data sampled with high speed[J]. Electric Power Automation Equipment, 2013,10:150-156.

[4] Zhou Guoliang,Soug Yaqi,Wang Guilan, Zhu Yougli. Research of condition monitoring big data storage and clustering[J]. Transactions of China Electrotechnical Society, 2013,S2:337-344.

[5] Cao Junwei, Wan Yuxin, Tu Guoyu, Zhang Shuqing, Xia Aixuan, Liu Xiaofei,Chen Zhen, Lu Chao.Information system architecture for smart grids[J].Chinese Journal of Computers, 2013,01:143-167.

[6] Zhang Dongxia, Miao Xin, Liu Liping, Zhang Yan, Liu Keyan. Research on development strategy for smart grid big data[J].Proceedings of the CSEE, 2015,01:2-12.

[7] Guo Ping,Dan Guangxiang. Mixed encryption algorithm in cloud computing[J]. Journal of Jilin University(Engineering and Technology Edition), 2012,S1:327-331.

[8] Cai Yudong, Shen Haibin, Yan Xiaolang. A High-speed implementation of AES[J]. Microelectronics&Computer, 2004,21(1):83-85.DOI:10.3969/j.issn.1000-7180.2004.01.022.

[9] Zhou Jian. The design and implementation of fast encryption algorithm for distributed RSA based on Hadoop[D].Shaanxi:Shaanxi Normal University,2013.

[10] Zhang Shaomin, Li Xiaoqiang, Wang Baoyi. Design of data security storage in smart grid based on Hadoop[J]. Power Svstem Protection and Control, 2013,14:136-140

[11] Qi Kaiyuan, Zhao Zhuofeng, Fang Jun, Ma Qiang. Real-time processing for high speed data stream over large scale data[J]. Chinese Journal of Computers, 2012,03:477-490.

[12]ANDERSON Q. Storm real-time processing cookbook[M].Birmingham: Packt Publishing, 2013.

[13] Shi Panpan. Parallel optimization and implementation of AES algorithm based on multi-core[D].Zhengzhou:Zhengzhou university,2012.

[14] Shi Jingang, Zheng Yan, Sun Huanliang, Luan Fangjun. Parallel processing of block cipher for massive data in cloud computing[J]. Journal of Frontiers of Computer Science and Technology, 2014,02:161-170.

[15] Desai A, Ankalgi K, Yamanur H, et al. Parallelization of AES algorithm for disk encryption using CBC and ICBC modes[C]// Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on. IEEE, 2013:1 - 7.