Adaptive Access Control Model of Vehicular Network Big Data Based on XACML and Security Risk

Peng-Shou Xie[1,2], Hong-Jin Fan[1], Tao Feng[1], Yan Yan[1], Guo-qiang Ma[1], and Xue-Ming Han[1]

(Corresponding author: Hong-jin Fan)

School of Computer and Communications, Lanzhou University of Technology[1]

Research Center of Engineering and Technology for Manufacturing Informatization of Gansu Province[2]

No. 287, Lan-gong-ping Road, Lanzhou,Gansu 730050, China

(Email: fan_hjin@163.com)

(Received Oct. 17, 2018; Revised and Accepted Mar. 23, 2019; First Online June 22, 2019)

Abstract

With the constant expansion of vehiclescale and the continuous development of Internet of Vehicles, the network environment of the data resource of Internet of Vehicles is becoming more and more complex. Traditional access control models have been difficult to meet the requirements of various access control conditions and dynamic adaptive adjustment of access control strategy. Aimed at the problem of adaptive access control model of vehicular network big data environment, XACML powerful ability of expressing access strategy is used in the paper, and we conduct the risk quantification based on 10 counts of risk factors, risk threshold and risk quota mechanism are also used for risk management. Experimental verification indicated that the risk adaptive access control model is effective, the research results will have great significance for promoting the application research of Internet of Vehicles and its safety technology and improving people's quality of life.

Keywords: Access Control; Big Data Security; Internet of Vehicles; Risk Adaptive; XACML

1 Introduction

The Internet of Vehicles is a large interactive network that contains information about vehicular location, speed, route, etc. Based on mobile communication and the information science technology, the Internet of Vehicles uses wireless communication technology, automotive sensors technology, global-positioning technology and automobile data recorder technology to complete the data collection of vehicular information and the surrounding environment, data transmission and processing, etc, in order to achieve effective intelligent monitoring, planning and management of vehicles, people, roads and locations [6]. It can be seen from the generation of big data in the Internet of vehicles,vehicular network big data has the characteristics of 4V, that is, Volume, high Velocity, Variety and high Value.In addition, it also has the following characteristics: Spatial and temporal scales span , large dynamic variability, high randomness, locality and finite life cycle. These characteristics of big data in the Internet of vehicles require us to provide more convenient services, such as data sharing and efficient computing to improve the processing efficiency of access control. In addition, when users enjoy the service, if do not provide reliable protection to these data with a large number of ownership characteristics, it will bring huge losses.

Access control technology according to the pre-defined access control policy ensures that resources can only be operated legally by legitimate visitors thus preventing unauthorized access to information. With the emergence of new computing environments such as cloud computing, Internet of things, some characteristics of those have brought great challenges to the application of access control technology, which makes the traditional access control model for closed environments such as Discretionary Access Control (DAC), Mandatory Access Control (MAC) and role-based Access Control (RBAC) difficult to apply directly to the new computing environment [10]. Subsequently, many related research work began to emerge, most of which focused on how to extend the access control of traditional models and how to introduce risks in the extended model. Hui Zhen *et al.* [9] proposed a risk-based access control model for medical big data, which can adaptively adjust doctors' access ability and protect patients' privacy. Chen Aiguo *et al.* [4] proposed a dynamic risk-based access control model, which emphasizes the risk measurement as an auxiliary decision indicator. The model uses the sliding window calculation method based on data stream, and the comprehensive final decision is affected by the policy, risk measurement and dynamic threshold. Xu Jing *et al.* [16] proposed a dynamic access control model, which introduced both the times of threat behavior and risk threshold into the trust model, the dynamic authorization was achieved by mapping trust level and permission. Almehmadi Abdulaziz *et al.* [1] proposed intent-based access control model, which uses the intent and intent motivation level to compute the access risk and greatly reduce the damages caused by internal threats. Chattopadhyay, Arup Kumar *et al.* [5] proposed a scheme uses simple Boolean based encryption and decryption of the data files which is low in computational cost, it reduced the risk of highly sensitive data from internal or external attacks.

Amghar,Sara *et al.* [14] proposed a new hybrid model, which uses KP-ABE and authentication system scheme to enhance the security and privacy of shared big data in the cloud.This model realized flexible and fine-grained access control for storing big data. Kibiwott, Kittur Philemon *et al.* [11] proposed a Cloudlet-Based eHealth Big Data System with Outsourced Decryption. It overcomes so many problems, such as confidentiality of data outsourced to the cloud, integrity of stored data, wide area network latency delays, and the resource constraints of the mobile devices. Lee, Ki Young *et al.* [12] proposed spatio-temporal XACML which could accept not only geospatial information but also temporal information and it compensated for the lack of Geo-XACML. Arunkumar [3] demonstrated the ability of the current OASIS standard to control access to XACML policies,described some confusing methods, and made specific suggestions on which elements should be involved in the process of access control. By combining XACML framework with the attribute based on encryption mechanism, Yang Yafeng [17] designed and realized a kind of attribute-based security enhanced cloud storage access control system applicable to cloud storage environment. Hou Shuchen [8] proposed a security access model for strengthening web services-based business system based on XACML system. Some progress has been made in the solution to security risk access control, but there is still a problem that is insufficient adaptive adjustment capability. Importantly, there are relatively few researches on risk adaptive access control methods specific to the vehicular network big data environment.

Considering the shortcomings of the above researches, we propose a security risk adaptive access control model. By the model, data security can be better protected. Based on the full use of XACML's powerful access policy expression capabilities, the introduction of quantitative risk control functions extends the XACML architecture, enabling dynamic adjustment of access policies based on visitor access, greatly improving vehicle network access control flexibility and applicability in complex network environments. The rest of this article is organized as follows. Section 2 introduces the XACML extension framework and the basic structure of the policy set. Section 3 describes the quantification process for the big data security risk of the Internet of Vehicles. Section 4 describes the decision and execution process of the strategy. The effectiveness of the model was tested by simulation experiments in Section 5. Section 6 concludes the solutions.

2 XACML Extended Framework and Policy Set Infrastructure

2.1 XACML Extended Framework

Figure 1 shows the XACML extended framework: the left side represents the XACML module, and the right side is the newly added module.

The functions of each module are as follows:

- Policy Administration Point (PAP): Create and maintain policies,policy sets and use files for storage.
- Policy Decision Point (PDP): Determine whether access requests are allowed, evaluate available policies, and provide authorization decisions.
- Policy Enforcement Point (PEP): Receive and send messages, interact with external applications according to results and obligations.
- Policy Information Point (PIP): Provides attributes information about subject, resources and environment.
- Context Handler (CH): Convert the access request to the XACML format and send it to the PDP.
- Subject (S): A visitor that performs an action on a resource.
- Resource (R): The data, services, and system components that the system provides to the visitor.
- Environment (E): A set of attributes that are related to authorization decisions and that are not related to specific property, resources or actions.
- Risk Engine (RE): PDP is invoked to handle risk-based access control. It mainly analyzes or solves these resource related risk policies; RE gets the attributes and request information from PDP, and these parameters would be substituted into a specific algorithm to calculate the risk value about the whole access request.
- Risk Quantification Function (RQF): Execute the risk measurement, they play a role inside the risk engine and make the use of risk policies more convenient.
- Risk Policies (RP): Define how each risk-based access control policy evaluates each resource. Using XACML's strong access strategy expression capabilities, we can does not change the original policy structure, just by setting the parameters of the rules in the strategy, the strategy for authorizing by risk value can be implemented. And risk management department can adjust risk strategy as needed. It can give

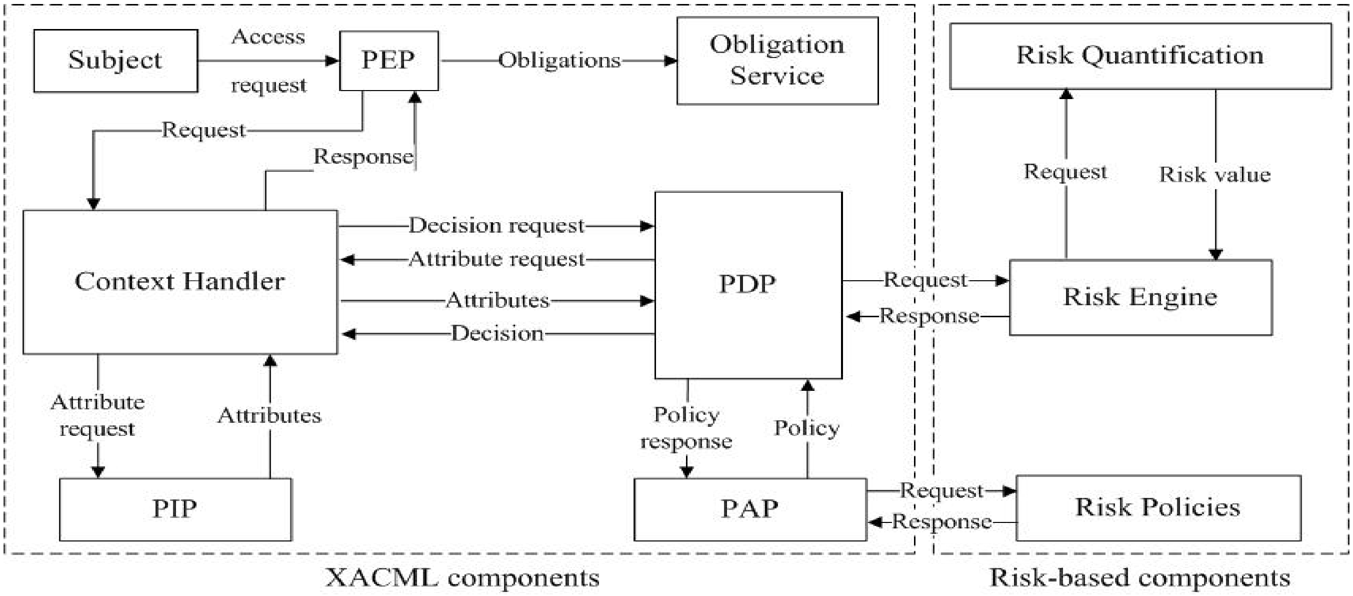

Figure 1: XACML extension framework

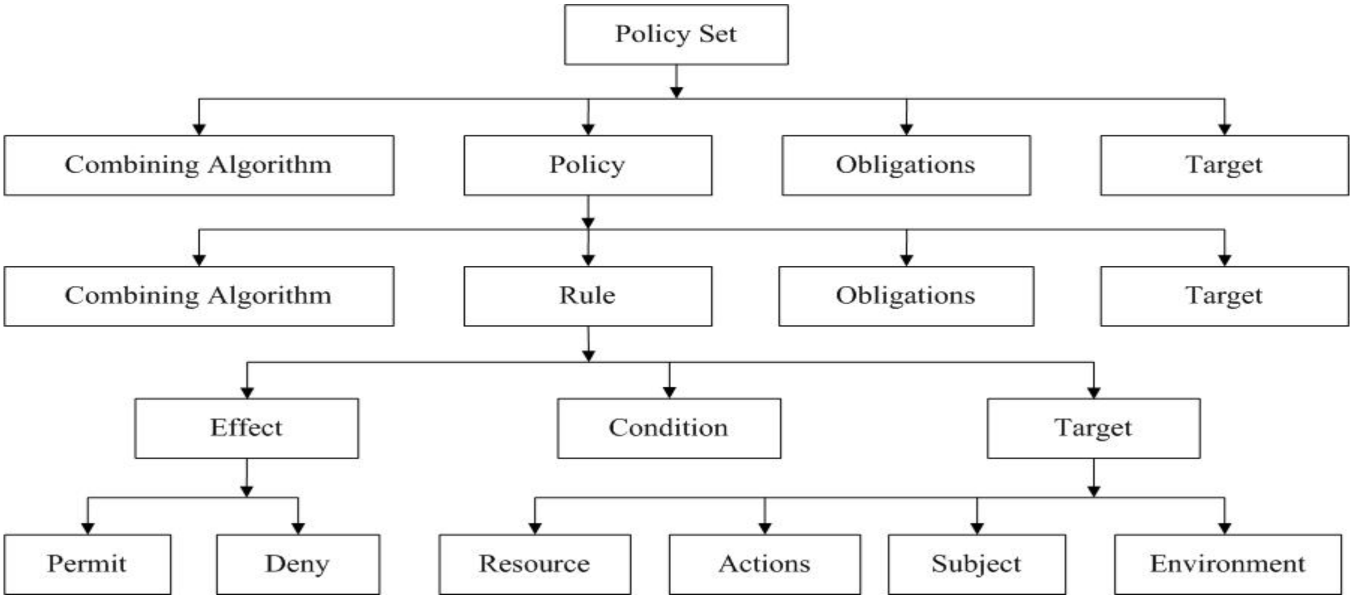

Figure 2: XACML basic structure of strategy set

Table 1: Merge algorithm

Algorithm	**Description**
Reject priority algorithm	If any assessment returns a refusal, the result must be a refusal, even if other assessments have returned permission.
Apply the algorithm first	Rules are evaluated in the order in which they are listed.
Unique application algorithm	For all policies in the policy set, if there is no applicable policy, the result is not applicable.If multiple strategies are applied, the result is indeterminate. If only one policy is applied, the result is the result of evaluating the policy.
License priority algorithm	If any assessment returns a license, the result must be a license, even if other reviews have returned a denial.

the judgment result by comprehensively judging the risk value of the access request and the allowable risk value defined by the policy.

The process framework of XACML after adding the risk point: When the subject issues an accessing request, authorization request will be send to PEP. Then PEP sends it to CH which will standardize the description of the attributes of visitors. In the meanwhile, create an XACML request and send it to PDP to decide whether the accessing authorization can be allowable. The practical policies are stored in PAP. PDP does not use all the policies in each accessing request. Instead, it searches applicative policies to evaluate the request and return the authorization decision back. For risk decisions, PDP will examine whether the resources used this kind of assessment method or not based on the instructions of relevant risk policies. If such a strategy does not exist, the result will be the traditional accessing decisions, conversely, PDP will send the request to RE to check the basic risk policies in the first place. If the basic policy is permitted, RE will Quantitative risk, and the result of risk measurement will be aggregated into a single value before return to PDP. PDP will decide whether to allow requests and send those to PEP and finally fulfil corresponding obligations, which are based on XACML policy and risk policy and merge algorithm.

2.2 The Basic Structure of Strategy Set

As shown in Figure 2, rule is the smallest unit of evaluating access requests which are consisted of three parts: Ondition, target and effectiveness. Logical judgement of accessing request is realized by conditional implementation. The decision result of rule determination is obtained by matching the subject, resource, action, environment in the target and the corresponding attributes in the accessing request. The upper layer of the rule is strategy, which is composed of target, merge algorithm, responsibility set and rule set. Responsibility indicates the tasks to be completed at the stage of strategy implementation. Strategy set is the most top-level structure of policy. XACML implements hierarchical policy management mechanism by this kind of nested structure. The merge algorithm is used to define the merging logic of results decided by multiple rules. According to the combination logic, the results of all rules are merged to get the final decision.

2.3 Merge Algorithm

A policy set may contain multiple policies and multiple rules. Different decision rules may result in conflicts. In order to obtain a unified decision result, a suitable merge algorithm is needed to resolve the conflict. The standard merge logic used by XACML has Deny-overrides, First-applicable, Only-one-applicable, Permit-overrides. Specific description as shown in Table 1.

3 Quantification of Vehicular Network Big Data Security Risks

Quantifying risk is to estimate each visit behavior and classify it into risk levels. Each risk level represents an access decision and action behavior. The top risk level and an access decision reject the contact, meaning the risk is high. Called that the boundary is a hard boundary; The lowest risk level contact with an access decision "allow" that means the risk is low, called that the boundary is a soft boundary; Between the hard and soft boundaries, and also an access decision associate with multiple operational actions. Traditional risk access control, which is static access control, just is allow or deny. But the risk adaptive access control described in this paper is a dynamic, multi-decision access control, that is allow-deny [13].

For risk points, the stage of risk assessment needs to input some factors (denotes subject, denotes object, denotes action, denotes context) to determine whether the accessing request was granted or rejected. This output function is based on a risk threshold and the mechanism of risk quotas to be managed. Specifically, our model determines the risk associated with access requests (visitor trust level and requested object security level and so on.) and then judging such requests according to the risk threshold of situational conditions. And if the quantified risk is below the risk threshold, the access request will be allowed, otherwise it will be denied. Another parallel condition of can Access is that the risk of quantification is less than risk threshold [2]. As shown in Formula (1).

$$canAccess(s,o,a,c) = \begin{cases} 1 \text{ if } risk < riskThreshold \\ \qquad \text{and } riskQuotas > 0 \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

Risk value(s, o, a, c)denotes that the risk comes from when subjects perform operations on objects according to context. Result 1 indicates the right to be granted, while the result of 0 is denied.

3.1 Risk Quantification Base On C, I, A

Table 2 shows the influences on data from different types of accessing behavior, this model means that accessing behaviors do the risk quantification with Confidentiality(C), Integrity (I) and Availability (A). When behaviors include risk attributes, it is designated as 1 otherwise as 0.

CiaRisk can be calculated by Formula (2) and Formula (3):

$$\begin{aligned} ciaRisk &= C * P_b + I * P_b + A * P_b && (2) \\ P_b &= \frac{N_b}{N_{all}}. && (3) \end{aligned}$$

Among them, P_b is the probability of occurrence of behavior, N_b is the number of the behavior occurs, N_{all} is the total number of occurrences of all behaviors, and the

Table 2: Risk value from Santos *et al.* [7]

Behavior	Data attribute	C	I	A
Create	sensitive/insensitive	0	1	1
View	sensitive	1	0	0
View	insensitive	0	0	1
Modify	sensitive/insensitive	0	1	1
Delete	sensitive/insensitive	0	1	1

probability P_b of each behavior can be calculated by using the statistical history of Formula (3). If the probability of the visitor modifying the data is 0.6, then ciaRisk = (0 * 0.6) + (1 * 0.6) + (1 * 0.6) = 1.2.

3.2 Risk Quantification Base on 6 Risk Factors of Internet of Vehicles

According to the results of researchers such as Santos *et al.*, Table 3 presents 6 risk factors under the Internet of Vehicles environment which are the index of risk quantification evaluation. The first group (Charact.of Visitor) shows the relevant resource information of visitors. The second group (Characteristics of Information and Requirements) shows the relevant risk of resource itself. It enumerates 2 groups total 6 risk factors and their weights. The total weight of each group is 0.5(1/2), the weight of each factor in each group is 0.5 / n. N is the number of factors in this group.

ContexRisk can be calculated by Formula (4):

$$contextRisk = \sum_{n=1}^{6} f_n * r_n \tag{4}$$

$$Risk_{Role} = \begin{cases} 1 & \text{R} \in \text{SuperAdmin} \\ 5 & \text{R} \in \text{Admin} \\ 10 & \text{R} \in \text{User} \\ 15 & \text{R} \in \text{Otherwise} \end{cases} \tag{5}$$

Among them, f_n is the weight of the risk factor, r_n is the risk value of the risk factor. The risk value of each risk factor is defined in advance, as Formula (5) defines the risk value for the role factor.

3.3 Risk Quantification Based on C, I, A, H and 6 Risk Factors

Ten risk factors are used for risk quantification in the paper, including 6 contextual factors, C,I,A risk factors and H historical records,that is:

- Safety features of behavior: security impact of confidentiality, integrity and availability behavior on resources.
- Contextual factor: visitor features, information features.
- History record: historical risk associated with the visitor.

The ultimate risk is:

$$\begin{aligned} aggregatedRisk = w_1 * ciaRisk + w_2 * contextRisk \\ + w_3 * hisRisk \end{aligned} \tag{6}$$

The H is the past risk value(hisRisk), which can be obtained by reading the past risk value from the database. If the visitor is first visit, the visitor's hisRisk is 0, w_1, w_2, w_3 are the weights of each metric category.

3.4 Vehicular Network Big Data Risk Threshold and Risk Quota Mechanism

The risk quota indicates how much the system is tolerant of the risk posed by each visitor. For access control, the system periodically assigns each visitor a certain number of risk quotas. Each visitor's visit behavior poses a certain risk and consumes the same amount of risk quota. If the visitor's risk quota is greater than zero, they can continue to access; Otherwise their access request will be denied until a new risk quota is obtained. The allocation of quotas is regular. The risk quota allocated each time should satisfy the normal visitors and will not be exhausted before the next allocation, that is, the request of normal visitors can be successfully passed.

For the formal description, the following symbols will be used.

V: A collection of visitors;

D: A collection of access data;

R: A collection of access records;

T: A collection of the same type of data.

This model periodically analyzes data visitor access records and calculates risk values. In the analysis of the history of the data visitor V_i, the same data access section visited by each visitor is integrated and recorded as D(V_i, D_j), where D_j is the data access section of the visitor, and $D_j \in$ D . The label of one of the types of data is represented by T_k, and $T_k \in$ T, the number of data accesses of the data block D_j and the data type T_k is represented by $F_{Vi}(D_j\ , T_k)$, and T_a represents all data types in the data block D_j. Through this number we can calculate the probability of data visitors accessing T_k data.

Using the calculation formula of information entropy [15], the amount of information obtained by the visitor V_i in the data section D_j.

$$P_{Vi}(T_k|D_j) = \frac{F_{Vi}(D_j, T_k)}{\sum_{T_a \epsilon T} F_{Vi}(D_j, T_a)} \tag{7}$$

$$H_{Vi} = -\sum_{k=1}^{T} P_{Vi}(T_k|D_j) \ln P_{Vi}(T_k|D_j). \tag{8}$$

Similarly, according to the historical access record, the average amount of information of all visitors V_{all} who ac-

Table 3: 6 risk factors of Vehicular Network Big Data

Risk factor	**Weight**
1. Characteristics of Visitor	
1.1 Role	$n_1 = 0.12$
1.2 Access Level	$n_2 = 0.12$
1.3 Previous Violations	$n_3 = 0.12$
1.4 Risk Quotas	$n_4 = 0.12$
2. Characteristics of Information and Requirements	
2.1 Sensitive level	$n_5 = 0.25$
2.2 Permission Level	$n_6 = 0.25$

cessed the data section D_j can be obtained.

$$\overline{H}(D_j) = \frac{H_{all}(D_j)}{C(V_{all})} \tag{9}$$

Among them, H_{all} (D_j) represents the total amount of information of V_{all}, and C (V_{all}) represents the total number of visitors. By comparing the information amount of the visitor Vi and V_{all}, the difference $Risk_{Vi}$ accessing the same data section D_j can be obtained, and then all the access section differences of the visitor V_i can be summed to obtain the risk threshold.

$$Risk_{Vi} = max\left\{0, H_{Vi}(D_j) - \overline{H}(D_j)\right\} \tag{10}$$

$$Risk_{Threshold} = \sum_{T_a \epsilon T} Risk_{Vi}(T_a). \tag{11}$$

A_m is the kth risk quota allocation phase, $Q_{Vd}(A_m)$ is the access quota used by visitor d at this stage, V(A_m) is the total number of visitors in stage A_m, and Formula (12) is the average. In the m + 1 risk quota allocation phase, the quota to be allocated is determined by the average of the quota consumption of the previous m stages. It is considered that the average of the quota consumption of the first m stages is a sample of a normal distribution, and then the mean and the variance s of the distribution can be obtained. The quota to be allocated is in the range of [?- ns, ?+ ns], where n is selected according to the system. Then set the probability ?= [0, 1] as the risk tolerance threshold of the risk adaptive access control system. If the probability of the visitor exhausting the quota in the next stage is less than the visitor can be assigned a new quota.

$$E(A_m) = \frac{Q_{Vd}(A_m)}{V(A_m)} \tag{12}$$

4 Policy Determination and Execution

4.1 Policy Determination

The access request process introduces risk quantification mechanism and policy decision function, the code is Algorithm 1:

Algorithm 1 Introduce Risk and Policy Decision

```
$< RuleRuleId = "" Effect = "Permit" >$
 $< Target > \cdots < /Target >$
<Condition FunctionId = http://research.sun.com/Projects/xacml/names/function\#Risk-quantification?
  <Apply FunctionId =rn:oasis:names:tc:xacml: 1.0: function:integer-one-and-only?
    <EnvironmentAttributeDEsignator
        DataType = http:www.w3.org/2001/XMLSchema\#integer?
  AttributeId =rn:oasis:names:tc:xacml:1.0:environment:riskThreshold?>
  </Apply>
   <AttributeValue = http:www.w3.org/2001/XMLSchema\#integer?
  </AttributeValue>
   </Condition>
</Rule>
```

This rule will be added to every strategy that requires risk determination. Its role is to quantify the access request by calling the method of risk assessment. When its condition is satisfied, the decision effect of the access in this rule is allowed. The Apply function gets the current system riskThreshold provided by the risk strategy.

Algorithm 2 Rule Quantification

```
Input: request
Output: ruleDecision
Begin
requestAttributes = PIP. requestAttributes
riskQuotas = PIP.riskQuotas
requestRisk = RG. quantify
riskThreshold = RP. riskThreshold
if ($requestRisk < riskThreshold$ and $riskQuotas >$ 0) then
    ruleDecision = permit
    return ruleDecision.
end if
    Return noEffect
End
```

Table 4: Dataset metadata

	Heading	**Type of data**	**Data Format**
1	medallion	string	Text string format
2	hack license	string	Text string format
3	IDvendor id	string	Text string format
4	rate code	string	Text string format
5	store and forward flag	string	Text string format
6	pickup datetime	string	Time format YYYY/MM/dd
7	dropoff datetime	string	Time format YYYY/MM/dd
8	passenger count	int	Normal integer format
9	trip time in seconds	int	Normal integer format
10	trip distance	float	Normal floating point format
11	longitude coordinates for the pickup location	float	Normal floating point format
12	latitude coordinates for the pickup location	float	Normal floating point format
13	longitude coordinates for the dropoff location	float	Normal floating point format
14	latitude coordinates for the dropoff location	float	Normal floating point format

The specific method of determination is as follows: First, the parameters related to the attributes provided by PIP are passed to RE to calculate the risk value of the access request. Then determine whether the risk value is less than the riskThreshold and the risk quota is greater than zero. Finally decide whether the access request is allowed. Each risk determination rule will be judged by reference to the risk threshold.In other words, the risk threshold manages the acceptable risk level of the entire system. If the administrator wants information to flow more smoothly, that is, the system can accept a larger risk value, you can increase the value; If the system administrator wants to be more careful about the flow of information, you can turn this value down.

Algorithm 3 Algorithm Decision

```
Input: request
Output: ruleDecision
Begin
policySet = PAP.match + RP.match
policy[] = policySet.match
for i = 1 to policy. quantity do
    // rule[] = policy [i] . match
    for  j = 1 to rule . quantity do
        // rule Decision [j] = rule[j].rulequantify.combine
        policy Result [i] = policy [i] . policyquantify.combine
        result = policySet. policyResult[i].combine
        return result
    end for
end for
End
```

The code is Algorithm 2.

The judgment result of each rule is merged by the merge logic preset by the policy.Finally, the judgment result of the strategy is obtained, and the corresponding obligation is added according to the judgment result.If there is a policy set at the top level, merge the decision results of each strategy with the merge logic preset by the policy set. Get the final judgment result and total obligation.

Table 5: Data visitor access log table

Heading	**Type of data**
name	string
age	int
gender	string
accountID	int
departmentID	longint
position	string
permission	string
risk quota	int
hisRisk	int
previous Violations	string
actionTime	string
action	string
path	string

The pseudo-code for the entire quantization process is Algorithm 3.

After the judgment phase is completed, the result information containing the judgment result and all the obligation are returned to the PEP. The PEP enters the next policy execution phase based on the content of the result information.

4.2 Policy Execution

Figure 3 shows the access control decision process.Similarly, the left side represents the XACML component and the right side represents the risk module. First, the principal issues an access request. Then the external

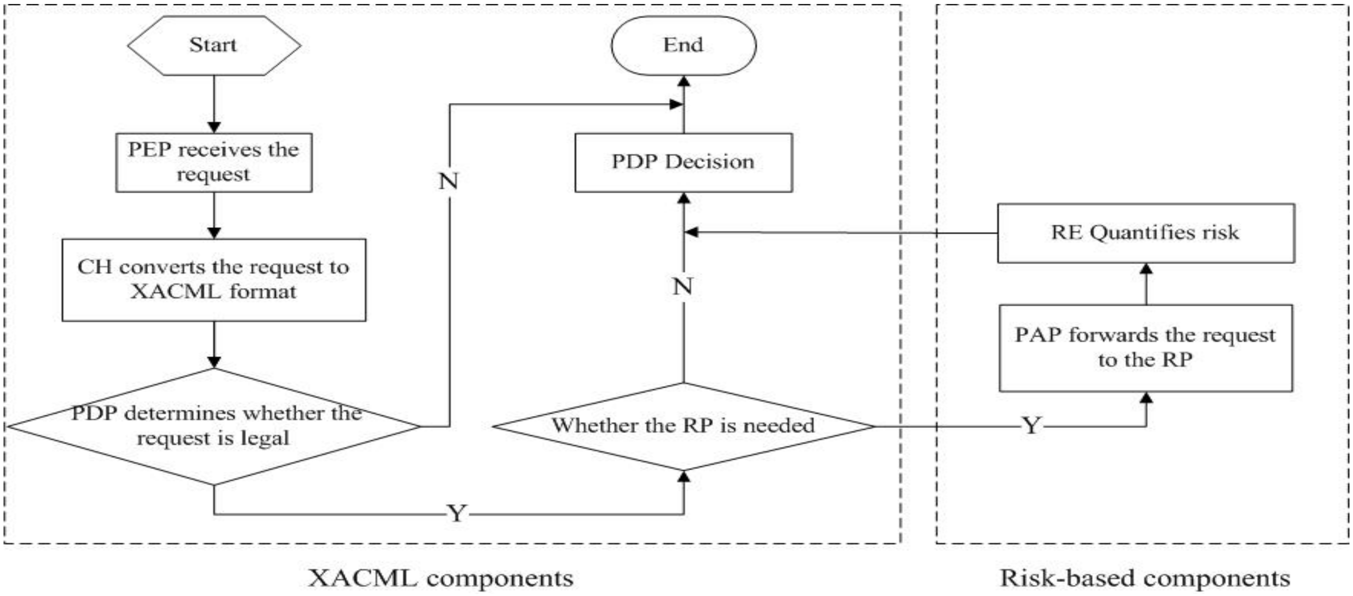

Figure 3: XACML extension framework

application passes the access request to the PEP.Finally, the PEP interacts with the external application. PEP sends an access request to CH. CH converts the access request format to XACML format and sends it to the PDP. The PDP is used to determine whether the access request is legal. The policy or policy set provided by the PAP is required in the decision, and the attribute information provided by the PIP is required. If the access request is illegal, it ends directly; If it is legal, it determines whether the access requires a risk policy. If a risk strategy is not required, the PDP evaluates directly; If required, the PAP makes a request to the risk policy. First, RE quantifies the risk and sends the result to the PDP. Then the PDP makes the decision and sends the result to the PEP. Finally, the PEP performs the relevant obligations.

5 Simulation

5.1 Experiment Setup

In the experiment, the taxi driving position record is used in this model to verify the privacy protection of the big data of the Internet of Vehicles. The data here comes from the real taxi detailed driving position data, including medallion, hack license , vendor id, rate code, store and for ward flag, pick up datetime, drop off datetime, passenger count, trip time in seconds, trip distance, latitude and longitude coordinates for the pickup location, latitude and longitude coordinates for the dropoff location and so on, the specific information is shown in Table 4 . We simulate access requests from two types of visitors, including each visitor's role, access rights, historical violations, risk quotas, et.al for each access record. The specific information is shown in Table 5. The visitor holds the access requirement to access the data, and finally, the risk value is calculated by the visitor's access record through the risk access control model.

5.2 Experimental Result

In the experiment, simulated the access history of 600 visitors as the experimental data, the information included in the history is shown in Table 5, it is about abnormal visitors, and the rests are normal visitors.Calculating the risk value for each visitor and sorting by risk.To test the effectiveness of the method, two indicators were examined.Accuracy rate represents the proportion of abnormal visitors among the top K visitors with the highest risk. Recall rate is the proportion of abnormal visitors in the top K visitors at all abnormal visitors. In each component module based on the XACML access control mechanism, the program is implemented in the Java language based on the Eclipse development platform. Important third-party development kits are based on SunXACML and the University of Murcia (UMU). The API of SunXACML implements the parsing and decision calculation of xacml.UMU uses the Java language to develop a UML-XACML-Editor V1.3.2 policy editor that supports the XACML 2.0 specification, which can be used to edit its own policy documents.

1) Experimental results under different visits. The experiment is mainly used to test the effect ofthe model on the number of different accesses requests. According to Formulas (2)-(11) risk value and risk threshold calculation method, the risk value and risk threshold of the visitor access data are calculated separately. As shown in Table 6, ESAV indicates that

Table 6: Results of 600 visitors

Visits	index	Quantity	Risk value	Risk threshold	Accuracy	Arain2017 Recall rate
5	ESAV	60		3.39	49/60	49/60
	AIAV	49	[4.27, 6.29]			
	AINV	551	[0.90, 3.28]			
10	ESAV	60		3.41	51/60	51/60
	AIAV	51	[4.09, 6.23]			
	AINV	549	[0.86, 3.26]			
15	ESAV	60		3.43	53/60	53/60
	AIAV	53	[3.77, 6.14]			
	AINV	547	[0.82, 3.23]			
20	ESAV	60		3.49	54/60	54/60
	AIAV	54	[3.68, 6.12]			
	AINV	546	[0.79, 3.10]			

Table 7: Risk Threshold calculation process when the number of visits is 5

	Sensitive data access times	Insensitive data access times	Total visits	entropy	amount of information	Risk threshold
D_1	189	267	456	0.00219298	0.67844549	3.393949
D_2	343	215	558	0.00179212	0.666601401	
D_3	467	511	978	0.0010225	0.692134799	
D_4	267	337	604	0.00165563	0.686416351	
D_5	165	230	395	0.00253165	0.679545906	

the number of abnormal visitors is set by the laboratory, AIAV indicates that the number of abnormal visitors identified by the algorithm, AINV indicates that the number of normal visitor identified by the algorithm.The experiment counts the identification of abnormal visitors under different access times. For example, when the number of visits is 5, the risk threshold is 3.39, and the risk value of abnormal visitors is between [4.27, 6.29]. The normal visitor's risk value is between [0.90, 3.28], the accuracy rate and the recall rate also reach 82% (49/60), and the accuracy and recall rate increase with the number of visits. It shows that the model in this paper can clearly distinguish two types of visitors, that is, the model is effective.

In the case of 5 visits, among the 5 data blocks, the sensitive data of the data block D_1 is accessed 189 times, the insensitive data is 267 times; The sensitive data of the data block D_2 is accessed 343 times, and the insensitive data is accessed 215 times; The sensitive data of the data block D_3 is accessed 476 times, and the insensitive data is accessed 511 times; The sensitive data of the data block D_4 is accessed 267 times, and the insensitive data is accessed 337 times; The sensitive data of the data block D_5 is accessed 165 times, and the insensitive data is accessed 230 times; According to Formulas (5)-(9) risk thresholds can be obtained when the number of visits is 5. The specific information is shown in Table 7.

In addition, this experiment also carried out extended statistics, which respectively counted the identification of abnormal visitors in the top 10, top 20, top 30, top 40 and top 50 highest risks. In Table 8 of the risk ranking results, the proportion of abnormal visitors in Top 10 is 100% (10/10), and in Top 50, our accuracy rate is also above 88% (44/50); In the case of recall rate, when the number of access log records of the system is 20 and K is 50, the recall rate is also above 78% (47/60), and the accuracy and recall rate both increase with the number of visits increase, which is because more visits can be more thorough understanding of the behavior and impact of visitors, and the calculated risk value is more accurate.

2) Experimental results under different abnormal visitor proportion. In this experiment, the number of visitors were still 600, mainly testing the identification of abnormal visitors at 5% (30 people), 10% (60 people), 15% (90 people) and 20% (120 people). And set the number of visits is 15 for per visitor, the test results are shown in Table 9. As can be seen from the table, the risk value of abnormal visitors is significantly higher than that of normal visitors. In this experiment, only the number of abnormal visitors is compared, so the accuracy and recall rate is the same in the same proportion. Moreover, as the propor-

Table 8: Accuracy and recall rate under different access times

Measure	Visits	K(Top K visitors with the highest risk value)				
		10	20	30	40	50
Accuracy	5	10/10	19/20	28/30	36/40	44/50
	10	10/10	20/20	29/30	37/40	45/50
	15	10/10	20/20	29/30	38/40	46/50
	20	10/10	20/20	29/30	39/40	47/50
Recall rate	5	10/60	19/60	28/60	36/60	44/60
	10	10/60	20/60	29/60	37/60	45/60
	15	10/60	20/60	29/60	38/60	46/60
	20	10/60	20/60	30/60	39/60	47/60

Table 9: Experimental results for different abnormal visitor ratios

Measure	The proportion of abnormal visitors to all visitors			
	5%(30)	10%(60)	15%(90)	20%(120)
Normal visitor risk value	[0.86,3.31]	[0.82,3.23]	[0.79,3.21]	[0.75,3.18]
Abnormal visitor risk value	[3.83,6.15]	[3.77,6.14]	[3.76,6.09]	[3.72,6.01]
Accuracy	25/30	53/60	81/90	110/120
Recall rate	25/30	53/60	81/90	110/120

tion of abnormal visitors increases, the accuracy and recall rate also increases from 83% (25/30) to 92% (110/120), and the overall performance of the model increases. Experiments show that this model is valid for different proportion of abnormal visitors.

6 Conclusion

Security risks adaptive access control for vehicular network big data is the theme of the paper, it combines the characteristics of XACML's powerful access policy expression capabilities to introduce risk extension XACML framework. It mainly introduces the process of determining and executing the risk quantification process and strategy. Finally, the effectiveness of the model is verified by simulation experiments. In a distributed environment, different enterprises or departments may have different requirements for authorization management, and they use different access control methods.The compatibility of multiple access control technologies must be considered during the development of dynamic authorization decision center.Later, we also need to test the time that takes for the visitor's request from the browser to the fully loaded and the delay in the number of risk metrics for the entire decision.

7 Acknowledgement

This research is supported by the National Natural Science Foundations of China under Grants No.61862040 and No.61762059. The authors gratefully acknowledge the anonymous reviewers for their helpful comments and suggestions.

References

[1] A. Abdulaziz and El-Khatib Khalil, "On the possibility of insider threat prevention using intent-based access control (ibac)," *IEEE Systems Journal*, vol. 11, no. 2, pp. 373–384, 2017.

[2] M. Abomhara, G. M. Køien, V. A. Oleshchuk, and M. Hamid, "Towards risk-aware access control framework for healthcare information sharing," in *The 4th International Conference on Information Systems Security and Privacy (ICISSP'18)*, pp. 312–321, 2018.

[3] S. Arunkumar, M. Srivatsa, B. Soyluoglu, M Sensoy, and F. Cerutti, "Privacy enforcement through policy extension," in *The 35th IEEE Military Communications Conference*, pp. 1096–1100, 2016.

[4] A. Chen, H. Xing, K. She, and G. Duan, "A dynamic risk-based access control model for cloud computing," in *The 6th IEEE International Conference on Big Data and Cloud Computing*, pp. 579–584, 2016.

[5] A. Chattopadhyay, A. Nag, and K. Majumder, "Secure data outsourcing on cloud using secret sharing scheme," *International Journal of Network Security*, vol. 19, no. 6, pp. 912–921, 2017.

[6] Z. Chunfeng, *Research and Application on Unstructured Big Data Storage and Processing for Vehicle Network*, University of Science and Technology of China, Master Thesis, 2018.

[7] D. R. dos Santos, R. Marinho, G. R. Schmitt, C. M. Westphall, and C. B. Westphall, "A framework and risk assessment approaches for risk-based access control in the cloud," *Journal of Network and Computer Applications*, vol. 74, pp. 86–97, 2016.

[8] S. Hou, *The Research and Application of Access Control Model Based on Attribute in the Web Services*, Beijing University Of Technology, Master Thesis, 2017.

[9] Z. Hui, H. Li, M. Zhang, and D. Feng, "Risk-adaptive access control model for big data in health care," *Journal on Commu- nications*, vol. 36, no. 12, pp. 190–199, 2015.

[10] M. S. Hwang, T. H. Sun, and C. C. Lee, "Achieving dynamic data guarantee and data confidentiality of pub- lic auditing in cloud storage service," *Jou- rnal of Circuits Systems and Computers*, vol. 26, no. 5, pp. 1–17, 2017.

[11] K. P. Kibiwott, Z. Fengli, O. A. Anyembe, and D. Adu-Gyamfi, "Secure cloudlet- based ehealth big data system with fine-grained access control and outsourcing decryption from ABE," *International Journal of Network Security*, vol. 20, no. 6, pp. 1149–1162, 2018.

[12] K. Y. Lee, A. Kim, Y. E. Jeon, J. J. Kim, Y. S. Im, G. S. Choi, S. B. Park, Y. S. Lim, and J. J. Kang, "Spatio-temporal xacml: The expansion of xacml for access control," *International Journal of Security and Networks*, vol. 10, no. 1, pp. 56–63, 2015.

[13] J. Li, C. Peng, Y. Zhu, and H. Ma, "Risk access control model for hadoop," *Chinese Journal of Network and Information Security*, vol. 2, no. 1, pp. 46–52, 2016.

[14] A. Sara, T. Yassine, and M. Abdellatif, "Secure confidential big data sharing in cloud computing using KP-ABE," in *The 2nd International Conference on Big Data Cloud and Applications (BDCA'17)*, 2017. ISBN: 978-1-4503-4852-2.

[15] Z. Song, H. Wang, H. Zhao, Y. and Chen, "Method and application for multi-scenario hybrid risk decision making based on utility-risk entropy," *Systems Engineering and Electronics*, vol. 40, no. 12, pp. 2751–2757, 2018.

[16] J. Xu, Z. Liu, S. Li, B. Qiao, and G. Tan, "A cloud-user behavior assessment based dynamic access control model," *Inter- national Journal of Systems Assurance Engi- neering and Management*, vol. 8, pp. 1966–1975, 2017.

[17] Y. Yang, *Design and Implementation of Security Enhanced Cloud Storage Access Control System Based on Attributes*, Beijing Institute of Technology, Master Thesis, 2016.

Biography

Peng-shou Xie was born in Jan. 1972. He is a professor and a supervisor of master student at Lanzhou University of Technology. His major research field is Security on Internet of Things. E-mail: xiepsh_lut@163.com.

Hong-jin Fan was born in Mar. 1993. He is a master student at Lanzhou University of Technology. His major research field is Security on big data of vehicular network. E-mail:fan_hjin@163.com.

Tao Feng was born in Dec. 1970. He is a professor and a supervisor of Doctoral student at Lanzhou University of Technology. His major research field is modern cryptography theory, network and information security technology. E-mail: fengt@lut.cn.

Yan Yan was born in Oct. 1980. She is a associate professor and a supervisor of master student at Lanzhou University of Technology. Her major research field is privacy protection, multimedia information security. E-mail: yanyan@lut.cn.

Guo-qiang Ma was born in Jun. 1992. He is a master student at Lanzhou University of Technology. His major research field is network and information security. Hwang2005ijnsaesE-mail: magq1514@163.com.

Xue-ming Han was born in Jan. 1990. He is a master student at Lanzhou University of Technology. His major research field is network and information security. E-mail: hxmhan@163.com.