Cyber Incident Classifications Using Ontology-Based Knowledge Representation for Cybersecurity Insurance in Financial Industry

Sam Adam Elnagdy[1], Meikang Qiu[2]*, Keke Gai[3]

***Abstract*—As a recent emerging industry, cybersecurity insurance has been growing ambitiously fast, which mainly serves the financial industry and assists financial firms to reduce cybersecurity risks. Understanding the risk classification is an important hemisphere for operating cybersecurity insurance. However, the classification representation will be complicated when the service system becomes large. Improper presentation of the risks can result in financial loss or operational mistakes. This paper addresses this concern and proposes an approach using ontology-based knowledge representation for cybersecurity insurance. The approach is named as *Semantic Cyber Incident Classification* (SCIC) model, which uses knowledge representation deriving from semantic techniques. Our approach is specifically designed for targeting at cybersecurity insurance domain, which has been assessed by our experiments.**

***Index Terms*—Cyber incident classification, ontology, knowledge representation, cybersecurity insurance, financial industry**

I. Introduction

Current booming changes in Web-related technologies have a remarkable impact on the financial industry in empowering value creations and improving business processes [1], [2]. The implementations of the Internet-based solutions are considered a mainstream of efficient business model [3]. Numerous financial services strongly depend on the applications of web services, such as international trade, e-commerce [4], and mobile payment. Despite current enormous benefits of using Internet-based solutions, the *Financial Service Institutions* (FSIs) are still facing a variety of challenges from the cybersecurity domain, such as malicious attacks, data breaches, and networking abuse. *Cybersecurity Insurance* (CI) is an option for FSIs to mitigate the risks of cyber incidents and increase the success rate of defending cyber attacks.

However, there are many unsolved problems in CI since it is remaining at an exploring stage. One of the problems in CI is that insurance vendors can hardly accurately categorize cyber incidents [5], [6]. The complex relations between insurance items often cause the repetitive workloads or uncovered aspects. This problem also results in confusions while CI clients determine the insurance items or CI service providers create service offerings. Misunderstandings and unclear definitions can also lower down the effect of defending cyber incidents. Whether or not this problem can be solved is a crucial aspect of spreading the CI implementations in practice.

We consider that one of the solutions to the proposed problem is classifying cyber incidents by identifying all relations involved in the system. Therefore, we propose a new model, named *Semantic Cyber Incident Classification* (SCIC), which uses semantic techniques to provide an consistent and convinced knowledge representation for mapping all entities in CI system. There are mainly three relations addressed in our model, which include *Incident to Incident* (I2I), *inSurance item to Incident* (S2I), and *inSurance item to inSurance item* (S2S). The proposed model targets at an effective approach that can accurately illustrate these three relations in the manner of knowledge representation within one system.

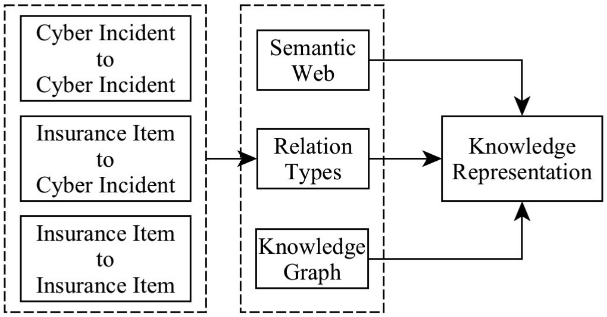
Cyber Incident to Cyber Incident
Insurance Item to Cyber Incident
Insurance Item to Insurance Item
Semantic Web
Relation Types
Knowledge Graph
Knowledge Representation

Fig. 1: The architecture of *Semantic Cyber Incident Classification* (SCIC).

Fig. 1 shows the diagrammatic architecture of SCIC model. For reaching the goal of accurate knowledge repre-

[1] S. Elnagdy is with Department of Computer Science, Pace University, New York, NY, 10038, se85420n@pace.edu;
[2] M. Qiu is with Department of Computer Science, Pace University, New York, NY 10038, USA, mqiu@pace.edu;
[3] K. Gai is with Department of Computer Science, Pace University, New York, NY 10038, USA, kg71231w@pace.edu;
* M. Qiu is the corresponding author of this paper. Email address: mqiu@pace.edu.

** This work is supported by NSF CNS-1457506 and NSF CNS-1359557. (Prof. M. Qiu).

978-1-5090-0946-6/16 $31.00 © 2016 IEEE
DOI 10.1109/CSCloud.2016.45

sentation, three vital parts are involved in our model, which are semantic web, knowledge graph, and relation types. As shown in the figure, three main relations are identified to support relation types. In our work, we mainly focus on the method using knowledge graph in semantic Web for the purpose of knowledge representation.

The main contributions of this work are:

1) We accomplish a research work of classifying cyber incidents in the cybersecurity insurance domain. The proposed method uses knowledge representation gained from ontology-based methods.
2) Our work has concentrated on the cybersecurity taxonomy issues in the financial industry, which explores the novel method of representing relations in CI.

The rest of this paper is written in the following order. Section II reviews and summarizes main recent work in CI and cyber incidents. Next, Section III illustrates the main concepts in our proposed model and describes the mechanism of SCIC. Furthermore, we provide a use case study as well as some discussions about future work in Section IV . Finally, we give our conclusions in Section V.

II. Related Work

A. Cybersecurity Insurance and Cyber Incidents

As an emerging industry, CI is a derivative domain of the Internet and cyber technologies. The concept of CI is an insurance business model that provide a number of service offerings focusing on reducing the negative impacts on business from cyber incidents. According to the *Department of Homeland Security National Protection and Programs Directorate* (NPPD), current CI industry is seeking cybersecurity protections services in main two aspects [7]. First, CI service providers explore more cyber incident evaluation approaches in order to increase the amount of coverage items. It implies that the scope of the preventive measures has a positive relationship with Second, the return of using CI services need to be better than self-protection done by CI customers. The return can be any interest dimensions, such as efficiency or financial costs.

Compare with traditional insurances, CI has a specific insurance focus that is not normally covered by other insurance domains. Most commercial liability or property insurance policies do not usually consider cyber incidents or risks the coverage items due to various reasons [8]. The technical burden is one of the constraints for insurance companies to cover cyber risks in their current policies. The constraints derive from a variety of reasons, such as the complexity of taxonomy, definition of service scope, and dynamic developing techniques [9]. The appearance of CI has enabled such insurance services by developing required professional techniques.

Moreover, a few recent researches have addressed the issue of improving CI service quality and performance [10], [11]. The research concentrations are varied due to the different commercial demands. We summarize three crucial research concentrations explored by recent researches, which are cyber incident knowledge sharing, cyber incident impacts and consequences, and the corresponding risk management for cyber incidents. These three research directions have formed three vital steps in generating effective CI policies. Fig. 2 illustrates a diagram of three crucial research dimensions for CI industry.

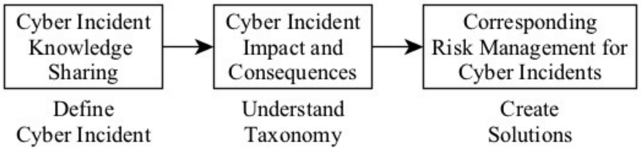
Cyber Incident Knowledge Sharing
Cyber Incident Impact and Consequences
Corresponding Risk Management for Cyber Incidents
Define Cyber Incident
Understand Taxonomy
Create Solutions

Fig. 2: Diagram of three crucial research dimensions of cybersecurity insurance industry.

At the stage of cyber incident knowledge sharing, current practitioners are looking for sharing incident shareable data for the purpose of improving CI framework or models [6]. This is a significant fundamental step for all improvement activities because the outcomes of this step have a direct impact on the succeeding work. The major sharing part is in knowledge sharing is building up a constructive data repository, which has been also suggested by *Homeland Security Department* (HSD) [7], [12]. The expected data repository is designed for the purpose of risk analysis, which is accessible for multiple parties. Therefore, the input data are anonymous cyber incident data and the output will be the results of data analysis.

However, one challenge of implementing this method is that classifying cyber incidents is a great challenge. Contemporary insurance policy generation is based on weighting both risks and costs [13]. It is difficult for service providers to select or gain proper parameters to accomplish the weights. This is because the relations between different cyber incidents within multiple parties are intercrossed, which is dramatically hard for mapping services deriving from a correct risk classification.

Therefore, our research addresses this problem and proposes a novel approach for solving the problem of cyber incident classifications. This problem was rarely addressed by the prior researches, even though there were a few unsuccessful attempts.

B. New Technologies in Cybersecurity Insurance

This section emphasizes the impacts deriving from emerging technologies, such as semantic web, cloud computing, *Internet-of-Things* (IoT), and big data. The importance of understanding these technologies' influences is great because many new cyber risks are caused by implementing new techniques. Classifying cyber risks needs to be aware of the causes.

First, the semantic technique is an approach that allows data reuse and exchange on the Web [14]. A few recent researches have attempted using semantic-based solutions in the financial industry. For example, a novel approach was proposed to proactively protect financial users' data, which used semantic web techniques to enable attribute-based access controls [15], [16]. In this research, financial customers' attributes are defined as individual ontologies and the access control policies are formed by linking the access requesters' attributes and data owners' configurations. The feasibility of using semantic techniques in the financial access controls has been explored in using multimedia big data in cloud computing [17]. However, these researches mainly focused on using semantic techniques to improve security level of financial services. The method of classifying cyber risks have been rarely addressed yet.

In addition, ontology is one of important component in semantic Web system, which is an effective approach for generating methods of knowledge representations. Many recent work has also explored in this field. For instance, it has been proved that using ontology-based knowledge representation can secure self-diagnosis in cloud computing [18]. Another research explored that ontology-based approach could prevent errors in *Electronic Health Record* (EHR) systems [19]. All these work done in the tele-health field have shown the feasibility of using semantic technique to represent knowledge. Despite many prior work addressed using ontology-based solution for knowledge representation, the research concentrating on classifying cyber incident for financial CI purposes has not been explored.

Moreover, as one of current most popular techniques, cloud computing has been broadly accepted by current FSIs [20]. The implementations of cloud computing in different industries have been investigated by numerous prior researches in order to make cloud-based solutions more adaptive and operative [21]–[25]. The cyber risks brought by cloud computing were also highlighted in the prior researches [26], [27]. For example, using cloud-based services may result in privacy information over collections [1]. Cloud users are not notified while the personal information is released on the cloud side [28]. Meanwhile, web services in cloud computing have empowered the indexing and searching capability [29]. For an enterprise, feasible services can be also used for improving business processes, which have been proved by many prior work [30]. However, cyber incidents of CI in clouds were rarely researched in spite of many done research work in cloud computing.

Furthermore, current data generation speed has become dramatically high, which drives financial data management to being a big data issue. Data mining is considered an effective method for assisting FSIs to determine a proper decision-making, such as data analysis in risk management or data governance [31], [32]. For example, classification is an important method in data mining that depends on the feature selections for the target dataset such that its optimizations were developed by many previous work [33], [34]. These prior work has provided abundant methods of feature extractions and data classifications.

In summary, we focus on extracting features of cyber incident and develop a model of classifications in this research. Knowledge representation is applied in our proposed model, which illustrates the relations between entities. The cyber risks caused by new technologies will also be considered, by which the relation generation will be defined. Following sections give the description about the proposed model.

III. Concepts and the Proposed Model

We provide statement of the main research problem as well as the model descriptions in this section. Section III-A gives the definition of the research problem. Section III-B demonstrates operating principle of the proposed model.

A. Definition

Our main research problem is defined in Definition III.1.

Definition III.1. Cyber Incident Relation Mapping (CIRM) Problem: *Inputs include the number of coverage item N_C, the number of pre-defined cyber incident N_{CI}, a set of ontologies for all coverage items $\mathbb{C}$=$\{C_1, C_2, \ldots, C_n\}$, a set of ontologies for all cyber incidents $\mathbb{CI}$=$\{CI_1, CI_2, \ldots, CI_n\}$. Outputs include a knowledge representation $\mathbb{K}$. Our proposed problem is find outing the method of generating knowledge representation by which classifications can be accurately generated for showing relations between coverage items.*

In our research work, we aim to find out the solution to identify the method of classifying cyber incidents. The complexity of this goal is that many incidents have overlaps with other incidents. The genetic relation is also a common relation in this field. For example, attacks from a website and attacks launched via emails have an overlap since they synchronously exist sometimes. A website launches the attack via sending an email to the attack objective covers both adversarial manners.

Concerning this issue, we define the main inputs include the number of coverage items N_C and a set of ontologies for these coverage items, represented as $\mathbb{C}$=$\{C_1, C_2, \ldots, C_n\}$. Meanwhile, considering cyber incidents, the inputs also include a set of ontologies for cyber incidents, $\mathbb{CI}$=$\{CI_1, CI_2, \ldots, CI_n\}$, and the number of these cyber incidents N_{CI}. We utilize ontologies to define constraints and construct knowledge graph. The output of our model $\mathbb{K}$ is a knowledge representation that can assist CI practitioners to obtain the knowledge via semantic Web.

B. Semantic Cyber Incident Classification (SCIC) Model

SCIC model is designed to use semantic web to generate knowledge representation by linking all ontologies. The benefits of semantic web is providing flexible solutions for describing web contents. The semantic language we used in our experiment is *Web Ontology Language* (OWL). Fig. 3 illustrates a diagram of the mechanism for SCIC model.

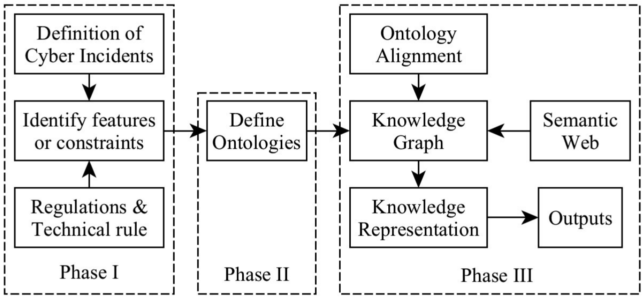

Fig. 3: Diagram of the mechanism for semantic cyber incident classification model.

As illustrated in the figure, there are mainly three phases. Phase I consists of a few activities for generating inputs that are used for defining ontologies in the succeeding Phase II. Three vital activities are involved in Phase I, which are defining cyber incidents, understanding regulations and technical rules, and identifying features or constraints of each cyber incident. Among these activities, defining cyber incidents and identifying features or constraints are two precedent tasks of feature/constraints identifications.

In addition, ontologies will be defined at Phase II. This is a crucial step in our model because the objects will be linked based on the relations and constraints. We consider both cyber incidents and cybersecurity insurance coverage items the ontology objects. OWL will be applied at this phase to define ontologies. For example, at this phase, type of security controls, level of security, cyber incident progressions, attack goal, threat model, and adversarial effect will be considered for generating a cyber incident ontology by identifying their relations between the object and these mentioned entities.

Finally, Phase III generates the knowledge representations that result from the outcomes of the Phase II. Three vital techniques used at this phase include ontology alignments, semantic Web, and knowledge graph. Among these, ontology alignments are used to converge and integrate defined ontologies. Semantic Web provides supports for data sharing and resource descriptions. Knowledge graph combines the results of ontology alignments with interconnections supported by semantic Web to represent the knowledge to users. We provide a use case as an example in Section IV in order to clarify the proposed model's implementations.

IV. Experiment and Discussions

A. Use Case Study

In this section, we provided a use case that is designed as a simple application example showing our proposed method. This example illustrated a multiple classifications that used different layers to show the relations between entities. The entities involved in this example included *System Failures*, *System*, *Hardware*, *Software*, *CybersecurityInsurancePolicy #*, *Performance*, *Capacity*, *Availability*, *Capacity Planning*, *Cloud Computing*, *Web Bottleneck*, and *Operation Problem*.

The first step was to identify features and constraints for these entities. We found a few crucial features and constraints below:

1) *System Failures* could be caused by three potential reasons, which included *System*, *Hardware*, and *Software*.
2) *Hardware* had a few dimensions for checking the system failure reasons, including *Performance*, *Capacity*, and *Availability*. Moreover, *Hardware* was involved in *Cloud Computing* as an important component.
3) One solution to *Capacity* was *Capacity Planning*.
4) As a solution, *Capacity Planning* required *Cloud Computing* as a technological support. Two main problems were addressed in *Capacity Planning*, which were *Web Bottleneck* and *Operation Problem*.
5) The problem caused by *Web Bottleneck* can be covered by the *Cyber Security Insurance Policy #1*.
6) The problem caused by *Availability* can be covered by the *Cyber Security Insurance Policy #2*.

Next, we developed ontologies according to these features and constraints. Fig. 4 illustrated a knowledge graph that showed the interconnections and interrelations between entities that were mentioned above. As shown in the figure, all entities were displayed and connected by arrow lines that showed the relations. The italic phrases showed the relation values.

Furthermore, we generated knowledge representation by using OWL. Two listing examples were given in Listing 1 and 2. Listing 1 showed an example of *Hardware* ontology expressed in OWL. Listing 2 expressed an example of *CapacityPlanning* ontology expressed in OWL.

```
<owl:Class rdf:ID=‘‘Hardware”>
  <rdfs:subClassOf rdf:resource=‘‘#SystemFailures
     ”/>
  <rdfs:label xml:lang=‘‘en”>Hardware</rdfs:label>
  <rdfs:comment xml:lang=‘‘en”>Device−
     Infrastructure</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource=‘‘#
     A_Failure_Reason”/>
      <owl:someValuesFrom>
        <owl:Class>
```

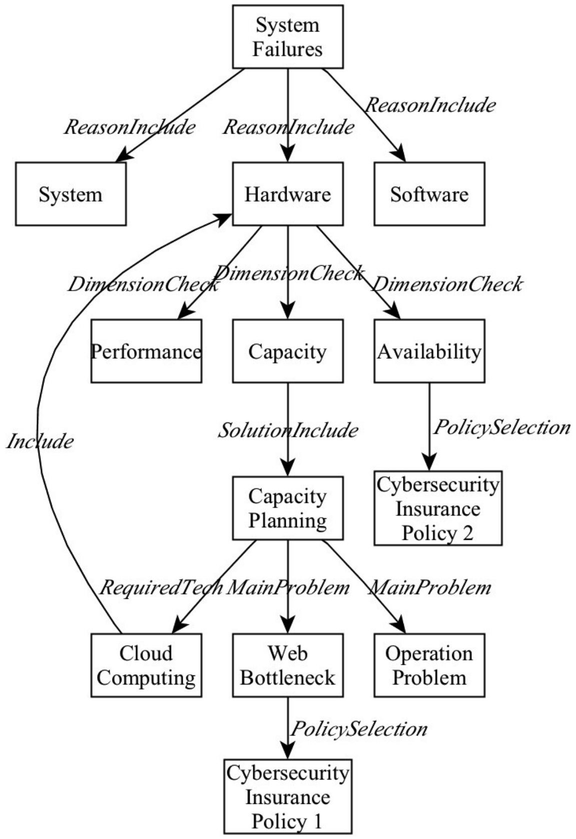

Fig. 4: Example of knowledge graph connecting entities by using ontology-based mapping.

```
          <owl:oneOf rdf:parseType=‘‘
    ReasonCollection”>
            <owl:Thing rdf:about=‘‘#System”/>
            <owl:Thing rdf:about=‘‘#Software”/>
          </owl:oneOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource=‘‘#
    CloudComputing”/>
    </owl:Restriction>
  </rdfs:subClassOF>
</owl:Class>
```

Listing 1: Example of *Hardware* ontology expressed in OWL.

```
<CapacityPlanning rdf=ID=‘‘Solution”>
  <hasRequiredTech rdf:resource=‘‘#CloudComputing
    ”/>
  <hasMainProblem rdf:resource=‘‘#WebBottleneck”/>
  <hasMainProblem rdf:resource=‘‘#OperationProblem
    ”/>
  <hasSolutionTo rdf:resource=‘‘#Capacity”/>
</CapacityPlanning>
```

Listing 2: Example of *CapacityPlanning* ontology expressed in OWL.

B. Main Findings

According to the results of current research stage, we have a few main findings below:

1) The knowledge representation is an effective approach for representing the complex relations between entities from various entity sources.
2) The ontology-based solution is an effective approach to connect cybersecurity insurance coverage items and cyber incidents or risks.
3) Semantic web techniques can be implemented for attaching various entities from different data sources in the cybersecurity insurance industry.

C. Discussions and Future Work

Our research has focused on developing a practical model of architecturalizing cyber incidents and risks in order to link them with security insurance policies. HoweverThere are two main limitations at current stage, which will be addressed in our future research work. (1) Our use cases explorations have not covered all relations in the cybersecurity insurance field yet. Most relations need to be defined in order to establish a comprehensive interrelationship network for representing all possible entities in cybersecurity insurance domains. (2) The proposed model needs to be evaluated in a practical operating environment. Contemporary research mainly focuses on the feasibility studies that are given in a smaller-sized implementation scope.

Therefore, considering these limitations, our future work includes:

1) We will further explore SCIC model by building the architectural structure and classifications detailedly.
2) Most use cases concerning the real-world applications will be explored in our future research as well. Various relations will be involved and SCIC model will be evaluated.

V. Conclusions

In this paper, we proposed an approach of classifying cyber incidents in order to solve the problem caused by the complicated relations between entities in cybersecurity insurance fields. The technique used in our proposed approach is ontology-based knowledge representation that derived from the classified information with interconnected relations. An empirical study was given in this paper and some future research work was defined.

References

[1] Y. Li, W. Dai, Z. Ming, and M. Qiu. Privacy protection for preventing data over-collection in smart city. *IEEE Transactions on Computers*, PP:1, 2015.

[2] K. Gai. A review of leveraging private cloud computing in financial service institutions: Value propositions and current performances. *International Journal of Computer Applications*, 95(3):40–44, 2014.

[3] K. Gai and A. Steenkamp. A feasibility study of Platform-as-a-Service using cloud computing for a global service organization. *Journal of Information System Applied Research*, 7:28–42, 2014.

[4] K. Gai, M. Qiu, H. Zhao, and W. Dai. Anti-counterfeit schema using monte carlo simulation for e-commerce in cloud systems. In *The 2nd IEEE International Conference on Cyber Security and Cloud Computing*, pages 74–79, New York, USA, 2015. IEEE.

[5] K. Gai, M. Qiu, and S. Elnagdy. Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data. In *The 2nd IEEE Int'l Conference on Big Data Security on Cloud*, pages 197–202, New York, USA, 2016.

[6] K. Gai, M. Qiu, and S. Elnagdy. A novel secure big data cyber incident analytics framework for cloud-based cybersecurity insurance. In *The 2nd IEEE International Conference on Big Data Security on Cloud*, pages 171–176, New York, USA, 2016.

[7] NPPD. Cybersecurity insurance industry readout reports. url=https://www.dhs.gov/cybersecurity-insurance.

[8] D. Garrie and M. Mann. Cyber-security insurance: Navigating the landscape of a growing field, 31 j. marshall j. info. tech. & privacy l. 379 (2014). *J. Marshall J. Info. Tech. & Privacy L.*, 31:i, 2014.

[9] L. Gordon, P. Loeb, W. Lucyshyn, and L. Zhou. The impact of information sharing on cybersecurity underinvestment: a real options perspective. *J. of Accounting and Public Policy*, 34(5):509–519, 2015.

[10] L. Shen. The NIST cybersecurity framework: Overview and potential impacts. *Scitech Lawyer*, 10(4):16, 2014.

[11] R. Kuhlman and J. Kempf. SEC issues cybersecurity exam observations. *Journal of Investment Compliance*, 16(2):44–46, 2015.

[12] DHS. Cybersecurity insurance, 2015. Retrieve from Forbes.com at http://www.dhs.gov/cybersecurity-insurance.

[13] L. Ma, L. Tao, Y. Zhong, and K. Gai. RuleSN: research and application of social network access control model. In *IEEE International Conference on Intelligent Data and Security*, pages 418–423, New York, USA, 2016.

[14] L. Tao, S. Golikov, K. Gai, and M. Qiu. A reusable software component for integrated syntax and semantic validation for services computing. In *9th Int'l IEEE Symposium on Service-Oriented System Engineering*, pages 127–132, San Francisco Bay, USA, 2015.

[15] M. Qiu, K. Gai, B. Thuraisingham, L. Tao, and H. Zhao. Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry. *Future Generation Computer Systems*, PP:1, 2016.

[16] K. Gai, M. Qiu, B. Thuraisingham, and L. Tao. Proactive attribute-based secure data schema for mobile cloud in financial industry. In *The IEEE International Symposium on Big Data Security on Cloud; 17th IEEE International Conference on High Performance Computing and Communications*, pages 1332–1337, New York, USA, 2015.

[17] Y. Li, K. Gai, Z. Ming, H. Zhao, and M. Qiu. Intercrossed access control for secure financial services on multimedia big data in cloud systems. *ACM Transactions on Multimedia Computing Communications and Applications*, PP(99):1, 2016.

[18] K. Gai, M. Qiu, S. Jayaraman, and L. Tao. Ontology-based knowledge representation for secure self-diagnosis in patient-centered telehealth with cloud systems. In *The 2nd IEEE Int'l Conf. on Cyber Security and Cloud Computing*, pages 98–103, New York, USA, 2015. IEEE.

[19] K. Gai, M. Qiu, L. Chen, and M. Liu. Electronic health record error prevention approach using ontology in big data. In *17th IEEE International Conference on High Performance Computing and Communications*, pages 752–757, New York, USA, 2015.

[20] M. Qiu, M. Zhong, J. Li, K. Gai, and Z. Zong. Phase-change memory optimization for green cloud with genetic algorithm. *IEEE Transactions on Computers*, 64(12):3528 – 3540, 2015.

[21] L. Chen, Y. Duan, M. Qiu, J. Xiong, and K. Gai. Adaptive resource allocation optimization in heterogeneous mobile cloud systems. In *The 2nd IEEE International Conference on Cyber Security and Cloud Computing*, pages 19–24, New York, USA, 2015. IEEE.

[22] K. Gai and S. Li. Towards cloud computing: a literature review on cloud computing and its development trends. In *IEEE Fourth Int'l Conf. on Multimedia Information Networking and Security*, pages 142–146, Nanjing, China, 2012. IEEE.

[23] K. Gai, M. Qiu, H. Zhao, L. Tao, and Z. Zong. Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. *Journal of Network and Computer Applications*, 59:46–54, 2016.

[24] H. Zhao, M. Qiu, K. Gai, J. Li, and X. He. Maintainable mobile model using pre-cache technology for high performance android system. In *The 2nd IEEE International Conference on Cyber Security and Cloud Computing*, pages 175–180, New York, USA, 2015. IEEE.

[25] H. Zhao, M. Chen, M. Qiu, K. Gai, and M. Liu. A novel pre-cache schema for high performance Android system. *Future Generation Computer Systems*, 2015.

[26] K. Gai, M. Qiu, L. Tao, and Y. Zhu. Intrusion detection techniques for mobile cloud computing in heterogeneous 5G. *Security and Communication Networks*, pages 1–10, 2015.

[27] K. Thakur, M. Qiu, K. Gai., and M. Ali. An investigation on cyber security threats and security models. In *The 2nd IEEE International Conference on Cyber Security and Cloud Computing*, pages 307–311, New York, USA, 2015. IEEE.

[28] K. Gai, M. Qiu, and H. Zhao. Security-aware efficient mass distributed storage approach for cloud systems in big data. In *The 2nd IEEE International Conference on Big Data Security on Cloud*, pages 140–145, New York, USA, 2016.

[29] K. Gai, Z. Du, M. Qiu, and H. Zhao. Efficiency-aware workload optimizations of heterogenous cloud computing for capacity planning in financial industry. In *The 2nd IEEE Int'l Conf. on Cyber Security and Cloud Computing*, pages 1–6, New York, USA, 2015. IEEE.

[30] K. Gai and A. Steenkamp. Feasibility of a Platform-as-a-Service implementation using cloud computing for a global service organization. In *Proceedings of the Conference for Information Systems Applied Research ISSN*, volume 2167, page 1508, San Antonio, USA, 2013.

[31] H. Jean-Baptiste, M. Qiu, K. Gai, and L. Tao. Meta meta-analytics for risk forecast using big data meta-regression in financial industry. In *The 2nd IEEE International Conference on Cyber Security and Cloud Computing*, pages 272–277, New York, USA, 2015. IEEE.

[32] H. Liang and K. Gai. Internet-based anti-counterfeiting pattern with using big data in china. In *The IEEE International Symposium on Big Data Security on Cloud*, pages 1387–1392, New York, USA, 2015. IEEE.

[33] H. Yin, K. Gai, and Z. Wang. A classification algorithm based on ensemble feature selections for imbalanced-class dataset. In *The 2nd IEEE International Conference on High Performance and Smart Computing*, pages 245–249, New York, USA, 2016.

[34] H. Yin and K. Gai. An empirical study on preprocessing high-dimensional class-imbalanced data for classification. In *The IEEE International Symposium on Big Data Security on Cloud*, pages 1314–1319, New York, USA, 2015.