

A survey on FinTech

Keke Gai^a, Meikang Qiu^{a,b,*}, Xiaotong Sun^a

^a Department of Computer Science, Pace University, New York, NY, 10038, USA

^b College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

Keywords:

FinTech
Cloud computing
Cyber security
Big data
Financial computing
Data-driven framework

ABSTRACT

As a new term in the financial industry, FinTech has become a popular term that describes novel technologies adopted by the financial service institutions. This term covers a large scope of techniques, from data security to financial service deliveries. An accurate and up-to-date awareness of FinTech has an urgent demand for both academics and professionals. This work aims to produce a survey of FinTech by collecting and reviewing contemporary achievements, by which a theoretical data-driven FinTech framework is proposed. Five technical aspects are summarized and involved, which include security and privacy, data techniques, hardware and infrastructure, applications and management, and service models. The main findings of this work are fundamentals of forming active FinTech solutions.

1. Introduction

As an emerging technical term, *Financial Technology* (FinTech) has been considered a distinguishing taxonomy that mainly describes the financial technology sectors in a wide range of operations for enterprises or organizations, which mainly addresses the improvement of the service quality by using *Information Technology* (IT) applications. A continuous growth of the investment has been powering the development of FinTech to advance on technologies breakthroughs in multiple areas, such as mobile networks (Wen et al., 2013; Zhang et al., 2013; Zhang and Soong, 2004; Gai et al., 2016a), big data (Yin and Gai, 2015), trust management (Zhang et al., 2016; Abawajy et al., 2016), mobile embedded systems (Zhang et al., 2011; Gai et al., 2017a), cloud computing (Castiglione et al., 2015; Gai et al., 2018), image processing (Castiglione et al., 2007), and data analytic techniques (Qiu et al., 2015a; Lee and Kim, 2015). Fig. 1 illustrates the growing trend of the FinTech investment during recent years. FinTech has become a hot term due to a number of driven forces, which include technical development, business innovation expectations (market), cost-saving requirements, and customer demands. It is reported that FinTech is considered one of the major investment for most competitive financial firms (Wigglesworth, 2016). However, a massive of implementations lead to a broad scope of utilizing FinTech solutions in various domains. A dramatically expanded expectation of using FinTech has caused a great challenge in its adoptions and planning, due to the intercrossed realms, complicated integrated systems, and distinctive demands. Therefore, an accurate and up-to-date awareness of FinTech has an urgent demand for both academics and professionals.

Moreover, recent researches have attempted in generating solutions to various FinTech problems. For instance, a quantity of recent investigations have emphasized the significance of security and privacy development in FinTech domains. Gartner's statistical reports (Morgan, 2015) present that the investment of cybersecurity is expected to turn into \$170 billion by 2020 globally. Only 35% companies that highly rely on the usage of technologies for their businesses are confident of their security, according to the statistics done by Silicon Valley Bank (Cybersecurity, 2015). The reality is that most modern *Financial Service Institutions* (FSIs) are applying IT-related techniques to support their financial services deliveries. This phenomenon has driven current researchers to solve the existing cyber threats at multiple layers throughout the technical process in the financial industry, from private to public sectors. Thus, gathering recent achievements of FinTech is a necessity.

Consider the driven force of applying FinTech, current development of FinTech is mainly matching the demands of financial service offerings. Crucial issues of FinTech can be categorized in various perspectives. From the technical perspectives, FinTech issues can be classified into five major technical dimensions include security and privacy, data techniques, hardware and infrastructure, applications and management, and service models. We provide a mapping structure of FinTech main issues in Fig. 2 showing the crucial cubes. As shown in the figure, we categorize the crucial aspects into five dimensions, including data-oriented, facility and equipment, applications, service models, and security and privacy dimensions. Our survey also follows these five dimensions in the succeeding sections.

As illustrated in Fig. 2, this paper concentrates on the crucial issues

* Corresponding author at: Department of Computer Science, Pace University, New York, NY 10038, USA.

E-mail addresses: kg71231w@pace.edu (K. Gai), mqiu@pace.edu, mqiu@szu.edu.cn (M. Qiu), xs43599n@pace.edu (X. Sun).

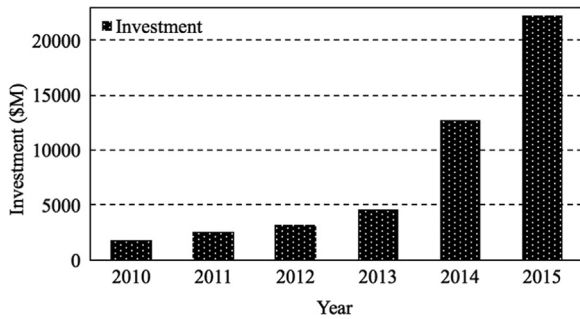


Fig. 1. Investment trend of FinTech from 2010 to 2015. Data are retrieval from Accenture Analysis on CB Insights Data [Shuttlewood et al. \(2016\)](#).

as well as the corresponding techniques or solutions in FinTech and formulates a presentation of the survey addressing the latest research achievements. The technical synthesis consists of five cubes, which include security and privacy, data techniques, hardware and infrastructure, applications and management, and service models. The selections of the issues were based on the contemporary development of the FinTech in recognitions, novelty, and demands, which also considers the value creations and improvements for current FSIs.

At the first dimension, we consider all operations on data for the purpose of financial services the data-oriented issues in FinTech, such as data analytics, data mining, and data deduplications. The research at this dimension is generally relevant with intelligent data usage or deep learning, which relies on utilizing data for value creations. In addition, the dimension of facility and equipment mainly refers to the infrastructure of financial service offerings as well as the corresponding systems. For example, many FSIs are adopting P2P business model over the networks, such that the configurations and establishments of entire system is a fundamental requirement for offering financial services. Next, a wide adoption of financial applications has been playing a dramatically important role in the contemporary financial industry. A variety of data-oriented applications have become supportive tools for improving financial services, such as SAS, Wealthfront, and Xero. Moreover, the deployments of the service model in FinTech are considered a wide scope of the research. The innovative service models are aligned with the enhancement of computing performances and network adoptions, such as smart city and cloud computing. Finally, one dimension penetrating all other dimensions is the issue of the security and privacy in FinTech. The challenges of security and privacy are restricting the adoptions of FinTech approaches and the corresponding solutions are required for ensuring the deliveries of other technical dimensions. In this paper, we aim to review recent achievements from these dimensions.

The main contributions of this work include:

1. This work concisely focuses on five crucial aspects of FinTech and completes a solid survey. The accomplishment offers an inclusive knowledge structure of FinTech for professional comprehensions and future researches.
2. The findings of this work highlight the foremost factors and concerns

of various cubes in FinTech. We find that data-driven applications and the associated hardware are critical factors of powering financial businesses due to its bring explored features and functionalities in multiple financial dimensions. The data-driven FinTech framework is proposed on the basis of our survey.

The rest of this paper is organized by the following order. [Section 2](#) summarizes the challenges of security and privacy in FinTech as well as potential or in progress solutions. In addition, [Section 3](#) reviews main techniques of data-oriented solutions in FinTech. Furthermore, [Section 4](#) represents major aspects of facility and equipment in FinTech and its corresponding restrictions and paradigms. Moreover, [Section 5](#) gathers recent achievements in FinTech applications and [Section 6](#) illustrates the service diversity of FinTech and recent relevant researches. Besides, in [Section 7](#), we propose our framework of FinTech based on the literature review aligning with discussions. Finally, we give our conclusions in [Section 8](#).

2. Security and privacy issues and solutions in FinTech

This section provides a review on the security and privacy issues as well as updated solutions in FinTech, which derive from our prior work ([Gai et al., 2016](#)).

2.1. Security and privacy issues in FinTech

We categorize the main issues of security and privacy in FinTech into three dimensions, as shown in [Fig. 3](#), which include business operations, outsourcing, and financial privacy. Detailed presentations on these dimensions are given in the following subsections.

Cyber concerns in the financial industry used to be a business operation issue at the early era of using electronic transactions and networking techniques ([Gai et al., 2016](#)). A variety of surveys had been done for forming solid IT security strategies ([Gai and Steenkamp, 2014, 2013](#)). One of the concerns for most financial firms was that the firms concerned about business operations using updated techniques, since the return of the investment on security was difficult to predicate and govern ([Farzan et al., 2013](#)). Classifying and understanding cyber incidents was a challenging task for most FSIs ([Gai et al., 2016](#)), which resulted in the difficulty in forming a strategic decision. To address this concern, a recent investigation ([Chai et al., 2011](#)) showed that the return of the security investment had a positive relationship with the level of the corresponding investment. Another study ([Liao et al., 2011](#)) further proved that perceiving privacy concerns and building up a trust mechanism were two critical tasks for securing transactions conducted electronically.

Meanwhile, some researches have addressed the threat sources in financial business operations. [Roumani et al. \(2016\)](#) completed a study on examining whether the records of the financial organizations are associated with the security vulnerabilities. This investigation considered the potential impacts caused by the assessed aspects, which included business scope, performance, markets, and sales. Using FinTech was an alternative of improving operations in various aspects ([Duan and Da, 2012](#)). An example of applying FinTech for improving

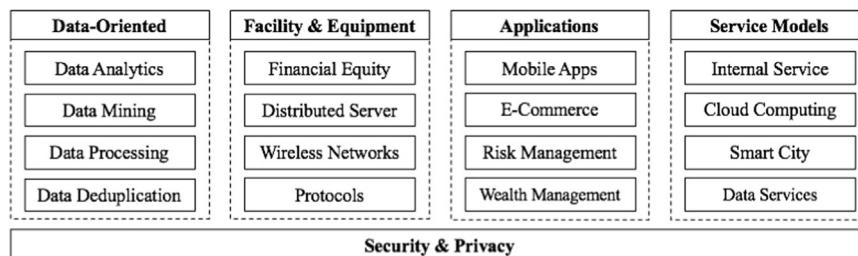


Fig. 2. Mapping main issues in FinTech.

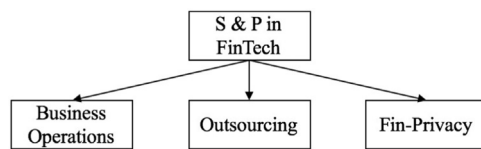


Fig. 3. Three dimensions of *Security and Privacy* (S & P) in FinTech.

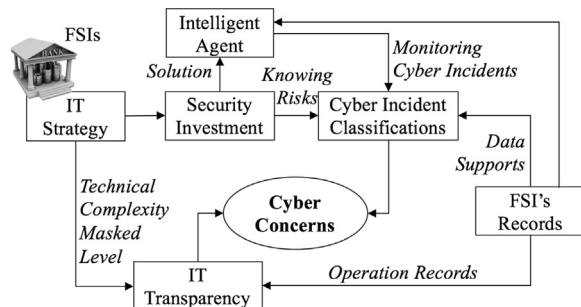


Fig. 4. Diagram of cyber concern generations deriving from business operations in the financial industry.

business operations was using intelligent agents to monitoring or predicting financial risks (Wang et al., 2002; Chen and Weiss, 2014; Zarandi et al., 2012). More work (Nussbaumer et al., 2012) have proved that the level of the IT transparency could impact on multiple business operations in financial advisory encounters from the perspective of the trustworthiness. On the whole, most concerns of the business operations mainly derive from unknown technical details, masked implementation process, and IT strategy-making. The hidden technical part can lead to concerns of business operations. Fig. 4 represents a diagram mapping the cyber concerns of the business in the financial industry. For FSI, most cyber concerns are arisen from the masked technical complexity and cyber incident classifications.

Moreover, as a popular Web-based service model, cloud computing has been widely accepted by the financial industry (Shim and Shin, 2016; Gai and Li, 2012). For example, *Bank of America* (BoA) has recently announced that the financial firm is collaborating with Microsoft enterprise to improve financial transactions by developing *Blockchain* technologies (Hernandez, 2016). The benefits of using cloud-based solutions are enabling financial businesses to closely connect to the target markets by increasing system performances (Gai et al., 2015), but this paradigm also introduces new threats due to outsourcing workloads. We converge major concerns of using cloud-based solutions as follows:

First, the masked complexity leaves FSIs a great concern due to lack of data controls in clouds (Tao et al., 2014; Gai, 2014), which makes private clouds a mainstream in the financial industry. A typical example is that FSIs may not know the physical server locations when using public clouds. Next, the data stored in remote cloud servers are still facing various threats, since the services settled on the complicated networks and intercrossed service participantships bring a lot of vulnerability opportunities. It is difficult for system designers and cloud vendors to fully predict or prevent the happenings of cyber risks in a cloud-based operating environment. Finally, the complicated and unanticipated communications between *Virtual Machines* (VM) can result in unpredictable vulnerabilities. Ni et al. (2014) pointed out that cloud data could be tampered due to the vulnerabilities of active protocols, even though the solution had been explored by the prior research. In summary, main threats of using cloud computing in FinTech derive from the complexity of the Web-based systems, lack of data controls, and uncertainty of technical details.

Furthermore, privacy protection is generally considered one of the most significant aspects in the financial security domain and preserving data carrying privacy is a critical task in producing a privacy protection strategy (Sánchez et al., 2012). A recent study evaluated the trade-off

between the data usage and privacy protection by implementing a machine learning-based method (Banu and Nagaveni, 2013). This work used a K-means clustering algorithm to discern data carrying privacy out of the multi-party clustering scenario. Additionally, location-based services usually carry users' movement privacy, such that the applications or devices have become common attack targets (Lu et al., 2012; Li et al., 2015). Despite the fine-grained approach can increase the privacy protection level, the problem of the latency time is restricting the application (Wang et al., 2014; Shao et al., 2014).

Finally, new financial services bring new concerns in security and privacy (Zhang et al., 2014). For example, implementing financial insurance is impacting on financial service organizations in making IT-related decisions, such as cybersecurity insurance (Elnagdy et al., 2016). Understanding cyber risks and discerning the relationships between the cyber incidents and the insurance covered items are challenging issues for many companies that are tightly attached to the Web-based applications. In addition, using electronic approaches for financial frauds (Sharma and Panigrahi, 2013) is another emerging issue in Fin-Tech. Traditional fraud detection methods usually relied on statistical methods (Ahmed et al., 2016; Chandola et al., 2012), which were insufficient to find out the continuous real-time deception happenings.

2.2. Solution synthesis

Many scholars have accomplished a great amount of researches in securing data. Some work exactly focused on the financial industry and some other studies addressed the universal solutions to security and privacy problems. In this section, we aim to synthesize the updated achievements that are either FinTech-oriented solutions or protection techniques that can be applied in the financial industry. Fig. 5 illustrates a structure graph mapping main techniques for security and privacy solutions. Our solution synthesis addresses the aspects mapped by the technical structure in the figure.

2.2.1. Risk detection explorations

The implementations of cloud computing have powered the cyber risk management by providing a flexible service deployment, either centralized or decentralized manners. For instance, one of the recent researches (Gai et al., 2016) has proved that cloud-based cyber risk management system could assist in classifying cyber-related information in the financial industry. This approach classifies information releases that can cause potential privacy leakage by using supervised learning techniques, which are combined with taxonomy. The work provides a feasible approach for classifying cyber incidents and align them with business items by using semantic techniques.

Next, financial frauds are frustrating FSIs as well. [Glancy and Yadav \(2011\)](#) developed a model for detecting financial frauds. The model

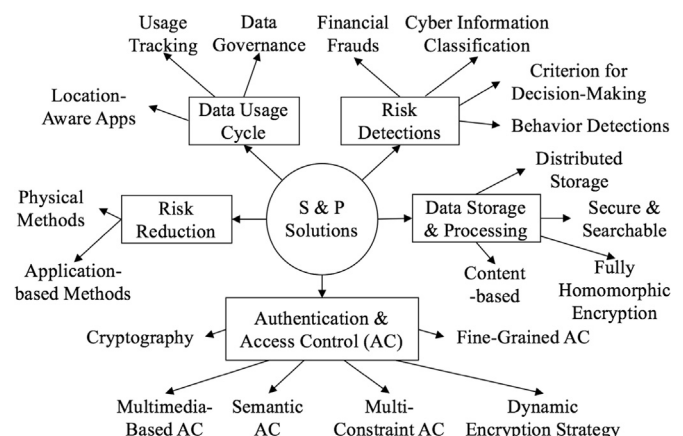


Fig. 5. Technical mapping for security and privacy solutions.

used quantitative analysis focusing on the frauds conducted by textual data. However, the quantitative approach could not ensure a stable detection accuracy, which meant the model could be only used as a supporting tool for distinguishing suspected transactions. An additional research (Kou et al., 2014) suggested a method that used multiple criterion decision making to select clustering algorithms for financial risk analysis. Moreover, analyzing correlation coefficients is an alternative approach for detecting abnormal operation behaviors, which has been proved by the prior work (Gai et al., 2015; Löhr et al., 2013). Improper activities can be detected when the correlation coefficient values deriving from multiple elements are different from the coefficients obtained from the clean dataset. In summary, most current work in financial risk detections intend to apply state analysis techniques.

2.2.2. Authentication and access control mechanisms

Many FSIs conduct authentications and access controls by using cryptography-based approaches. Besides the data encryptions for protecting financial information, a few novel security mechanisms are also alternatives for securing financial privacy. One solution direction was exploring the mechanism of strengthening access controls throughout multimedia (Li et al., 2016b). The semantic accesses are associated with the service requestors' features identified by ontology techniques, which are achieved by defining effective ontologies for access control configurations (Choi et al., 2014). Financial service acquisitions can be reached by creating semantic-based access controls that are supported by multimedia. Another ontology-based solution (Asamoah et al., 2016) was proposed to power up cybersecurity ecosystem by generating the knowledge graph that showed the inter-relations between cyber risks and their causes. However, ontology-based approaches usually has a limitation that the accuracy of access controls can hardly reach a perfect when the amount of ontologies in the system is large.

Furthermore, the fine-grained access control is one of the solutions to secure data storage (Ruj et al., 2012; Tang et al., 2012). This method can be also applied in cloud computing environment. Bugiel et al. (2013) have developed a fine-grained policy-based access control approach, which were designed to fit in and follow the diverse security and privacy policies on Android. The policy-based access control can be used in various needs scenarios, such as assured data deletions (Tang et al., 2012) and content sharing (Nabeel et al., 2013). A fine-grained access control approach can also achieve attribute-based keyword searching when multiple users independently encrypt and store the data in the cloud server (Sun et al., 2014). An owner-oriented access policy can be formed by utilizing the fine-grained-based approach.

Moreover, some previous work addressed developing multiple constraints for data accesses. One approach (Kanak and Sogukpinar, 2014) was proposed to address strengthening biometric authentication systems by considering three dimensional constraints in creating protection strategies, which included security, privacy, and trust. The researchers formulated the trade-offs from these three aspects in order to increase the efficiency of the protection when biometric technologies are applied. However, this approach is restricted in many application scenarios, since three dimensions in criterion, including security, privacy, and trust, have vague boundaries in practice. Adjusting the protection mechanism is also a challenging issue in that forming an implementation strategy takes a long time and frequently switching the mechanism of biometric authentication systems is not applicable for the demands of most financial services (Davis et al., 2013).

In addition, to improve the performance of the data encryptions, some scholars have developed a few approaches for dynamically determining the strategy of data protections. For example, one approach (Gai et al., 2016) was developed to selectively encrypt data based on the privacy classifications with privacy weights under the fixed timing constraint. This approach could produce optimal solutions to maximize the total privacy weight value. In summary, the main

trade-off of access security in FinTech is attached to the conflict between security and service performance. Most current FSIs intend to maximize the level of security to protect all transactions as well as financial customers' privacy.

2.2.3. Data usage cycle protections

FSIs generally emphasize the importance of data usage, since its performance has a direct relationship with the service quality. The data governance may become dramatically complicated when the size of the system increases or new functions are added (DeStefano et al., 2016). A research direction addresses the critical concern of financial data protections, which is to prevent data from malicious behaviors launched by the unexpected third party during the data usage cycles. Many previous studies have explored a variety of methods to reduce the risk level when data are shared, transmitted, or exchanged through different parties. As one of the approaches, tracking data usage (Enck et al., 2014) is an alternative for secure data governance. Chang and Ramachandran (2016) proposed a mechanism that used *Business Process Modeling Notation* (BPMN) to discern the processes and implementations of the data usage. The outcome of this approach can point at the period or data range that requires additional security protection operations based on the data usage simulations. Another research (Yu et al., 2014) also considered the research perspective of the business process but focused on payment systems. This work emphasized the vulnerabilities caused by poor business process management in e-commerce, such as improper control flows.

Furthermore, Xiao et al. (2013) proposed a novel security system for delivering location-aware services, which did not rely on techniques of key sharing. This approach utilized the proximity-based authentications and temporal location tags for establishing the session key. However, this work did not reach a non-error rate goal because of the radio propagation properties. Another research attempt (Bilogrevic et al., 2014) focused on developing a privacy protection strategy of grouping mobile users. By using this method, the data carrying location information are not needed to frequently share or transmit, since mobile users are grouped based on locations. Nevertheless, this method was hardly to be implemented because the inputs and geographical locations are generally dynamic for most location-oriented applications (Wang et al., 2014). Therefore, hazards in data usage cycle mainly come from a few dimensions, which include participations of the unexpected third party, unclear business processes, and large-range distributed usage.

2.2.4. Secure data storage and processing

Improving the security mechanism in data storage and processing is also vital research direction in FinTech and many researches in this field have been done in recent years. Gai et al. (2016) proposed a solution to securing data by applying a distributed data storage in cloud systems. This approach considers two potential adversaries that include both internal and external attackers. The data carrying sensitive data are divided into two parts before the data are sent out on the network, so that the privacy can be protected even though the transmissions are monitored by adversaries. This approach can be further strengthened by adding a judgement process by which the data requiring distributive storage are determined (Li et al., 2017). The improvement is an optimization of reducing workload when the volume of data is large.

Next, some other scholars looked into privacy-assured searchable data storage in cloud computing. One of the recent developed methods (Li et al., 2013) used symmetric-key encryption primitives and covered three functions, which included ranking results, identifying similarities, and searching structured data. This method allows users to share the encrypted data in the distributed service deployment. Moreover, prior privacy-related researches not only explored text-based file storage but also investigated the image maintenance in cloud computing. For example, one research (Xia et al., 2016) proposed a scheme

that used *Content-Based Image Retrieval* (CBIR) technique to encrypt images before they are stored in cloud servers. Nevertheless, the mathematical operations were not executable on the encrypted data in clouds. Achieving *Fully Homomorphic Encryption* (FHE) demanded more work.

In summary, FSIs intend to obtain a higher level data protection no matter what technologies are selected. Centralized data storage and processing can reduce the risks of the privacy release at server side, but it has a limited impact on improving security level during the data transmissions. Meanwhile, a decentralized data storage/processing is also facing challenges from various dimensions, such as monitoring communications and database abuses. Therefore, securing financial data is associated with protecting data within a network system and all threats existing in a web should be addressed in this field.

2.2.5. Risk reduction and prevention

In general, there are two basic types of methods for risk reductions and preventions, which include physical and application-based methods. The physical method refers to the approach of securing data by conducting operations on the physical infrastructure, such as preventing a jamming attack or avoiding a network damage (Gai et al., 2017b). An application-based method refers to the security solution achieved by cryptographic methods, such as creating a secure protocol or configuring access controls (Ma et al., 2016). For example, using logical authorization language is an alternative approach for forming an access control rule in social networks (Ma et al., 2016). Besides these two aspects, a number of new research directions in reducing financial cyber risks have been proposed along with the development of new technologies.

First, developing a proactive protection approach is an alternative research direction for reducing risks. One scheme was proposed to protect financial customers' privacy by using attribute-based access controls (Gai et al., 2015; Smari et al., 2014). This approach only allows those third parties that are configured as the trustable parties to decrypt their data either fully or partially. The approach was improved by introducing semantic web such that the relations between the data owners and unknown third parties can be clarified (Qiu et al., 2016). However, this approach needs configurations for defining trusted parties done by data owners, which means it may bring dramatical extra workloads when financial customers have the authentications for establishing their own trust parties.

Next, the financial production selection can be influenced by various elements when there exist a few alternative and available service choices. To address a proper decision-making on determining financial productions, some recent researches have tried to develop applicable solutions under certain constraints. For example, a research weighted all available financial services and measured coefficients in order to obtain the optimized solution to choosing services (Merigó and Gil-Lafuente, 2010). This method simply used an ordered sequence such that the weight values were not well addressed. Another research had similar research focus on weighting service items, which developed a method of classifying cyber risks (Elnagdy et al., 2016). This research had stronger contribution to cyber risk classifications by using semantic techniques to create knowledge representation graphs. The work was further improved (Gai et al., 2016c) by applying *Monte Carlo* (MC) simulations for efficient data analytics in forming security framework.

Moreover, many prior researches concentrated on reducing the risks related to physical locations. Ma et al. (2013) proposed a method of increasing the security level for *Radio-Frequency IDentification* (RFID)-based applications. This approach considered the geographical information-related data the constraints of the data accesses; thus, a financial transaction will be denied when an abnormal location-related movement is detected. The limitations of this approach are that the mechanism highly relies on RFID techniques and the accuracy of the adversaries detections is under debate. Another work (Wei et al., 2011) developed an authentication protocol of RFID that used a hash

function to reach mutual authentication between the data and data-bases. Using a hash function for the purpose of creating secure RFID protocols had been also used in other work (Sun and Zhong, 2012). The crucial part of using RFID techniques in securing financial data/information is to successfully protect hash tags.

In summary, it is difficult for FSIs and system administrators to perceive potential cyber hazards from emerging technologies or applications until the adversaries attack. Typical solution is to understand the active system and reduce the rate of attacks by discerning technical details and business processes.

3. Data-oriented techniques in FinTech

3.1. Big data analytics and data mining

Operating data mining within the big data context (Wu et al., 2014) is a mainstream of obtaining valuable information from the large volume of data pool. As a crucial tool of the information acquisition, quantitative tools have broadly implemented in operating financial data (Shi et al., 2017). The development of big data makes data represent more values than they used to be (Sagraves and Connors, 2017). New services are created for meeting the demands of the private clients by producing customized services, such as independent financial advisor or financial management. For instance, Jiang et al. (2016) developed an approach of creating personalized travel sequence recommendations by using photos and the corresponding heterogeneous metadata. This approach bridged up the travel preference with the routes, such that the sequence can be created based on the data mining on a group of parameters, such as features, costs, and time. Therefore, the data-oriented researches of FinTech have become active and popular over years.

The data-oriented researches of FinTech refer to any academic explorations or investigations of using data to obtain values from improving financial services or creating new service offerings. The core of this part is about information management and quick rational analyses; thus, data analytics and artificial intelligence are strongly attached to this aspect. For example, an advanced big data analytics can assist auditors in detecting frauds (Cao et al., 2015). Another research (Zhang et al., 2017) presents QuantCloud infrastructure that is designed for financial data analytics. This approach used a large-scale SSD-backed datastore in order to interconnect the technical section and data-driven research. There are a variety of approaches for value acquisitions when considering the type of the target data, such as intensive data (Chen and Zhang, 2014). In the financial industry, most FSIs are interested in mining data in deep to gain the value data pools.

The value of mining data assists FSIs to discern the truth of the financial occurrences, including the reasons, processes, potential impacts or results, and applicable solutions. Understanding the relationships between financial entities is a fundamental option for FSIs to govern financial incidents. Address this issue, developing a semantic solution to illustrating the relationships of entities in the system is an effective approach, which has been proved by recent researches in various domains, such as error information detections (Gai et al., 2015), incident classifications (Elnagdy et al., 2016), and data governance (DeStefano et al., 2016). The ontology-based approaches could be used to discover web services by semantically categorize the services, such as applying the *Universal Description Discovery and Integration* (UDDI) for offline semantic categorizations. It has also been proved that adding properties could increase the ranks of web services so that the discoverability could be enhanced (Hao et al., 2010). An exploration (Paliwal et al., 2012) further improve the performance of service discoveries by using cluster computing in order to increase the discovery accuracy. Another project (Ma et al., 2012) used similar ontology-based mechanism but focused on categorizing text-mining for the purpose of the decision making assistance.

Despite many attempts on ontology-based solutions, the accuracy

issue is restricting most prior means. The challenging part in this issue is to find out the method of generating semantic trees as well as defining ontologies. Zhang et al. (2015) present a novel data-driven paradigm in order to solve the mapping problems in parallel computing. This approach addresses the data modules by using network switch when new data modules are inputted. A task can be formed by related data modules as well as specific methods. However, the accuracy rate can hardly reach a perfect level due to the intercrossed platforms and the complicated structure of the subsystems. Some other scholars have paid attention to using machine learning techniques for assisting data analytics in recent years.

Machine learning techniques were applied in protecting privacy. Distinguish from traditional privacy-preserving solutions focusing on small-scale data sets, it has proved that using machine learning could have a better performance in securing privacy in a large sized data set based on the training dataset. One approach (Xu et al., 2015) was using data locality property and letting limited cryptographic operations allowed in the Reduce() procedures. Nevertheless, most training data are labeled for the purpose of machine learning, which means that those data may carry sensitive financial information. The major drawbacks of this type of approach include: (1) it is a time consuming task for accomplishing the training process; (2) it is hard to avoid privacy leakage during the training operations from the perspective of the internal attacks.

Furthermore, within the networking context, *Online Learning* (OL) has become a popular data mining approach for assisting FSIs in make decisions from resources on the Internet. A study (Wang et al., 2014) proved that combining *Cost-Sensitive Classification* (CSC) with OL was an effective method. This method applied online gradient descent techniques to maximize the total weight or minimize the weighted misclassification cost. In addition, it (Bernstein et al., 2005) has been demonstrated that using ontology-based methods could be applicable in producing solutions to CSC. The ontology operator creates a heuristic ranking deriving from validating the collected data mining processes. Other researches addressing CSC used other techniques, such as *K-Labelsets Ensemble* (KLE) (Lo et al., 2014) and Soft CSC (Jan et al., 2012). The studies above attempted in increasing the learning efficiency and online resource utilizations.

In summary, recent researches had paid sufficient attention to increasing the performance of data analytics using various techniques. The role of data mining was emphasized by both professionals and academics in FinTech. Besides the accuracy of the outcomes concerned by data analytics, the computation efficiency of the big data processing was also addressed by the recent researches as the other significant aspect of data-oriented techniques.

3.2. Big data processing

The implementation of big data is remarkably impacting quantitative finance, from business process modeling to back-end statistics (Fang and Zhang, 2016). The performance of big data processing is a crucial aspect of ensuring the quality service for current FSIs, since both data volume and data types become remarkably greater than that of few years ago. Numerous previous researches had explored the advance of data processing to snatch the innovations in the era of big data. The main trend of big data processing in FinTech is migrating from structured data and simple statistics work to unstructured data and complex computation work. An investigation (Casado and Younas, 2015) found that advanced big data processing could be located at a few directions, which mainly included unstructured data analyses, real-time data processing, batch processing, massive distributed processing, and data processing management tools. Thus, the development directions of big data process in FinTech turn into multi-dimensions.

Consider the big scope of the computation workload, many contemporary FSIs determine to use large sized datacenter for centralized data processing, which also raises the conflict between energy

efficiency and computation performance (Georgakoudis et al., 2016). The conflict turns into a remarkable effect as the size of financial service sector grows. Recent researches also addressed this issue and proposed a few solutions. Qiu et al. (2015b) designed a genetic algorithm for the optimization of *Phase-Change Memory* (PCM) in order to achieve green computing. Another research (Gai et al., 2016) focusing green memories developed a heuristic algorithm of data allocations in the heterogeneous memory. This approach used a novel operation, named the *Smart Switch*, in order to transfer a sub-optimal solution to an optimal solution at a high ratio. All these work emphasized the capacity of the memory and its enhancement; while some other researches focused on improving the computation efficiency from other perspectives.

One method (Li et al., 2016a) was proposed to increase the human behaviors' detections by using *Field-Programmable Gate Array* (FPGA)-based parallel architecture. This approach also addressed the optimization of processing heavy workload by minimizing computations and memory space and applying partial dynamic reconfiguration on FPGA. Moreover, some researches addressed analyzing massive semantic graphs for solving complex analytic problems. Each vertex or edge carries the attributes such that the queries can be determined based on the analytic goals. One research (Lugowski et al., 2015) solved the analytic queries by deploying a filter that filtrates unnecessary vertices and edges; thus, the customized graph can be used for analytic purposes.

In summary, the improvement of data processing in the perspective of hardware was highly concerned by recent studies. The concept of big memories was driving the increase of the big data processing; therefore, some memory-related researches had become popular recently, such as data allocations and memory designs.

4. Facility and equipment explorations of FinTech

Deploying advanced facility and equipment is a fundamental requirement for contemporary FSIs to retain the competitiveness. The deployment requires not only advanced level performances but also scalable and flexible adoptions. Employing virtual memories in the distributed manner was considered an alternative solution to increasing computation performances. The big memory servers also encounter various challenges. For example, graph analytics may consume a great portion of computation workload on *Translation-Lookaside Buffers* (TLB) misses (Basu et al., 2013). Recent studies also addressed the computing utilities. Kousiouris et al. (2014) accomplished an investigation on the resource provisions within the distributed environment. This approach applied a two-layer generic algorithm and provided the estimates of the resource attributes for applications. The forecasted attributes were requirements of the applications on both user and application sides.

Moreover, the distributed and scalable computing deployment generally results in the distributed data storage. However, this data storage manner introduces the challenges of data mining within a distributed computing environment. For solving the training problem in distributed data mining, one solution (Lu et al., 2008) was using *Distributed Parallel Support Vector Machine* (DPSVM) training mechanism that switched support vectors within the networks. The outcome of this solution could ensure different servers synchronously processed distributed data when consider both cost and efficiency. Another research (Yu et al., 2015) presented an approach addressing the issue of *High Availability* (HA) in database, which covered three dimensions, including cost, performance, and availability. This study also emphasized the significance of integrating *Database Management System* (DBMS) with three crucial concerns, which were storage, performance, and security. The results of the study had shown that the *Input/Output Operation per Second* (IOPS) could be 27 times faster than the traditional method.

Furthermore, a wide adoption of the mobile computing has driven

the movement of some financial operations from wired networks to the wireless. This trend has brought a quantity of benefits, such as improving financial business processes, increasing customer relationship management, and diversifying financial service deliveries. Nevertheless, implementing wireless networks and communications are generally associated with security issues. Data transferred over the networks require a higher-level protections throughout the communications in the financial industry. In the big data context, the conflict between the data stream authenticity and the medium of communications becomes greater than that of before, since the continuous authenticities may cause great latencies. It has been considered that the dynamic security verification or authenticity system for communications could be an alternation. One scheme was proposed for big data stream processing using dynamic shared key periodical updates (Puthal et al., 2015). Another approach (Gai et al., 2016) was proposed to use multi-channel communications to balance the conflict caused by efficiency and security. This approach used dynamic programming to produce the optimal solutions to obtaining the high privacy protections when assigning the data packages to certain communication channels under the configured timing constraint.

In summary, the major trend of exploring facility and equipment in FinTech is implementing scalable hardware deployment over the Internet. The distributed scalable computing leads the financial business to more value creation directions; however, it also brings challenges in both financial applications and data exchanges.

5. FinTech applications

5.1. Mobile finance and E-Commerce

The expeditious development of Web-based technologies had driven the integration of e-commerce and social networks (Geslevich-Packin and Lev-Aretz, 2016). The power of current social networks enriched the ways of creating or transferring the personal financial recommendations and solutions, such as commercial reviews and making payment over the social networks or media. One of the crucial issues in mobile finance is to ensure the mobile devices can efficiently search inclusive information from tons of Web objects. Lee et al. (2011) designed a mobile web navigation using routed directed trees to increase the efficiency of valuable information acquisitions. This approach emphasized the value of the information that was determined by the ranks of the relevant Web sites. Another research (Gai, 2015) considered the heterogeneous resources in cloud computing and dynamically changed the source usage based on the predictions of the workload and Web capacity. A dynamic implementation for achieving real-time services had become a mainstream of the mobile financial applications.

Moreover, in the mobile context, the wireless bandwidth had a direct relationship with the performances of mobile financial apps. Dynamically shifting channels and redistributions could be an alternative solution for those apps requiring high-performance in communications, such as stock trading applications. Address the *Quality-of-Service* (QoS), an approach (Misra, 2014) was developed to combine both shifting and redistributions for the purpose of QoS-guaranteed bandwidth. This approach considered the bandwidth redistribution the utility maximization problem, such that the shifting and redistributions were determined by the bandwidth constraints between the mobile nodes. Another similar research (Qiu et al., 2016) proposed an optimal solution to optimizing the usage of the bandwidth. This method considered three variables, which were time, success transmission probabilities, and networking capacities. The outcome was the task assignment that could maximize the utilization of the networking bandwidth.

Next, implementations of FinTech applications also brought some business operational issues. For instance, the anti-counterfeiting issue in e-commerce is remarkable crucial since the counterfeiting deluge

could have a strong impact on the business reputations. Liang and Gai (2015) proposed an approach using the Internet-based data collection for analyzing the patterns of counterfeits. This model emphasized the cost of counterfeiting and aimed to assist the government to form anti-counterfeiting strategies. In addition, an approach (Gai et al., 2015) focusing on detecting counterfeits in e-commerce used correlation coefficients to estimate whether the target could be a suspect counterfeiting provider by comparing it with the real commodities. Most updated work in anti-counterfeits used the pattern recognition techniques to detect the suspicious behaviors. However, the anti-counterfeiting accuracy still needed to be increased, since the behavior-based data mining could not fully uncover the matter of counterfeits. Many elements could influence the outcomes of data mining, such as training dataset selections and sizes.

In summary, the improvement of FinTech applications in mobile finance mainly concentrated on the mechanism of the complicated computing resource sharing within the Web-based environment. Many researches considered the advancement the optimizations of increasing computing resource utilizations in the distributed manner. Some issues caused by the implementations of FinTech applications were introduced to the public, such as counterfeiting issues in e-commerce. Next section presented the recent achievements of FinTech in management.

5.2. FinTech in management

We consider applying FinTech in management the approach of solving the complicated business problems that use the advanced computing techniques, such as optimizations or machine learning. Li and Hoi (2014) completed a survey on *Online Portfolio Selections* (OPS) that were considered a fundamental issue in the domain of financial computations. They pointed out that the OPS problem could be formulated as a sequential decision problem in the perspective of online machine learning. Their main findings were categorizing OPS solutions into five groups, including benchmark-based, *Follow-the-Winner* (FW), *Follow-the-Loser* (FL), pattern matching-oriented, and *Meta-Learning Algorithm* (MLA) approaches. Additionally, they also presented that the accuracy of predictions in OPS was still an unsolved issue, even though multiple researches had addressed this realm.

Moreover, fuzzy algorithms had been explored in job scheduling optimization fields. Liu et al. (2010) proposed a fuzzy *Particle Swarm Optimization* (PSO) algorithm to increase the performance of job scheduling in computational grids. Their studies showed that fuzzy PSO approach was superior to the general genetic algorithm and *Simulated Annealing* (SA) method. Another similar work (Xhafa and Abraham, 2010) revealed the complexity of computation grid scheduling and discussed the adoptability of using meta-heuristic approaches. The challenge of using PSO-based solution was to determine the sets of swarms, which could directly influence the results in various cases.

Furthermore, Wang (2015) designed a dynamical model that used fuzzy systems theory to transfer the common adopted trading rules into excess demand functions. The model was mainly applied in stock operations and utilized nonlinear dynamic paradigms that followed two trading strategies, including *Follow-the-Big-Buyer* (FBB) and *Ride-the-Mood* (RM). A similar research (Sanz et al., 2015) using fuzzy rule-based classification systems was proposed to deal with the financial crisis predictions. Another research (Wang et al., 2013) in stock predictions combined *Decision Tree* (DT) with *Support Vector Machine* (SVM) algorithms. There were two phases in this approach. The first phase used DT algorithm to filter noises and the second phase applied SVM algorithm to achieve high performance. Additionally, as a tool of machine learning, *Artificial Neural Networks* (ANN) was also applied in predicting stock markets, which was integrated with genetic fuzzy systems (Hadavandi et al., 2010). Nevertheless, similar to other data-driven solutions, selecting proper datasets is a challenging issue for training data as the results could be dramatically distinct deriving from different data source.

In summary, data-driven solutions began to be popular in solving financial business problems, such as strategy-making and stock market predictions. Prior researches attempted on using machine learning techniques to explain the business issues or achieve intelligent analyses, which highly relied on the data. Recent studies also demonstrated that the accuracy of using machine learning techniques needed to be increased for matching the requirements of FSIs.

6. Service diversity of FinTech

6.1. Internal services and information flow optimization

From the perspective of the service scope, current FSIs can be categorized into two classes, including large-size enterprises and middle/small sized enterprises. In general, large sized FSIs establish their business ecosystems (Basole et al., 2013) for reaching distinct market targets through the complicated and global networks, such that the internal connections establishment and maintenance are challenging. A few dimensions include data visualizations, application streaming, trading data, and real-time services.

The data visualization is an approach for enterprises to formulate their strategies and business planning based on the data acquisitions (Kandel et al., 2012). One solution (Satyanarayan et al., 2016) was proposed to deliver the declarative visual and interaction designs for data visualizations. This method created a data flow graph in order to classify the first-class streaming data source and re-wrote the data flow at runtime based on the demands. Other data visualization techniques address distinct concentrations. Some researches attempted to visualize data in a trajectory manner (Tominski et al., 2012; Wang et al., 2013). However, technical challenges of applying data visualizations (Liu et al., 2014) were fourfold: (1) there generally exists a gap between visualized data and enterprises' demands; (2) challenges occur when integrating heterogeneous data; (3) results may highly rely on the dataset's size and state; (4) inaccurate information acquisition from visualization can mislead enterprises.

In addition, application streaming is an emerging distribution method for on-demand software, which supports the partial application installations in terms of the real-time needs (Wu et al., 2016). The operations of application streaming generally need to cover application states, codes selections, and user-defined configurations within the parallel computation context. The challenge is that data parallelism needs to be configured by application developers or developed by the compilers, such that the historical data cannot be applied for general stream processing. One solution (Schneider et al., 2015) was examining the semantics of the transformed programs by comparing the data-parallelization with the original sequential data streaming.

Moreover, trading data has been considered the crucial aspect of improving the usage of big data as well. The methods of managing data trades have been explored by recent studies. For example, one approach (Mashayekhy et al., 2014) used a two-sided mechanism that traded computing resources for data-oriented applications. This work mainly focused on determining prices for the deployment of cloud-based services. For data buyers, the challenging part was effectively integrate data from various sources. Data cleaning (Wang et al., 2014; Shepperd et al., 2013) is a fundamental procedure for improving efficiency of data integrations and increasing data quality when processing a large size of data.

Furthermore, for achieving a real-time service, data gathering is a challenging issue for those services that require high workload data analytics, such as risk analysis. A method (Yin et al., 2015) improved *Partial Least Squares* (PLS) by fixing the problems of the fault diagnosis for the performance indicators. Two aspects were covered by this approach, which included performance test statistics and the corresponding maintenance action generations. Ding et al. (2016) developed an algorithm for real-time big data collections, by which the clustering data transmission structure based on the signal strength indicators could be developed.

6.2. Cloud computing and smart city

The development of cloud computing and smart city has brought diverse service models, which impacted on FSIs from the external perspective. Liu et al. (2014) proposed a mechanism that could simultaneously support authorized auditing and fine-grained update requests. This study addressed two limitations of using cloud-based storage services. The first limitation was missing authorizations/authentication between auditors and clouds. The other limitation was that coarse-grained updates required redundant computations such that it could result in additional storage and communication costs.

Next, as one of the crucial parts in smart city, *Knowledge Discovery in Databases* (KDD) is struggling with the dramatically growing amounts of data. The core mechanism of discovering knowledge is to obtain low-dimensional information from a high-dimensional data in which the computation complexity is generally high. One approach (Chen et al., 2015) of dealing with high-dimensional data is combining the communication theory with information theory to modelize the data transfers over the networks and create an equalizer for data transmissions.

Moreover, a study (Islam et al., 2012) concentrating on the online transactions between enterprises pointed out that the execution time of initializing new virtual instances could cause great latencies in computing resource allocation. The common solution to this problem was to develop an optimizer that could predict computing resources for allocations. Some prior studies attempted on the optimizations of the whole service life cycles (Ferrer et al., 2012). However, most previous work had specific focuses considering the computer science discipline. One research (Gai et al., 2016b) on allocating multimedia big data to cloud-based memories applied a genetic algorithm to produce sub-optimal solutions. This method emphasized the weights of the amounts of the Read and Write. Another work (Gai et al., 2016) addressed the task assignment rather than data allocations.

In summary, service diversity of FinTech mainly influences FSIs in two aspects. The first aspect is that updated FinTech solutions can strengthen FSIs' problem-solving abilities, such as visualizing data and increasing data usage. The other aspect is that external development, such as cloud computing and smart city, is changing the financial business environment. FSIs can outsourcing technical services from the third party.

7. Discussions

7.1. Main findings

This survey reviewed five crucial aspects of FinTech, from data-driven applications to cybersecurity, even though there might be more aspects attaching to FinTech. For example, FinTech was also with financial social networks, business process improvements, or strategy reforms. The emerging technologies were driving an exceptional enrichment of the financial service offering, which also brought a variety of challenges from the technical perspective. such as security and privacy concerns. The continuously instantaneous financial business environment demanded an on-demand and quick-action technical supports as well as a secure service platform. We summarized our main findings addressing major appearing challenges and solutions as given in the following statements.

- An explicit perceptibility of financial business processes was a critical task for formulating secure data flows and predicating user scopes. Cyber hazards could be caused by the participations of unknown, untrusted, or unexpected data users. Using multimedia was a trend to strengthen the access controls in that multiple constraints could be applied in validations. The logical restrictions defining role authentications could increase the protections in network-based financial systems.

- Mobile financial services highly relied on the utilizations of mobile devices and networks, such that cyber risks attached to distributed networking systems should be addressed, such as monitoring communications and hacking cloud storage. Meanwhile, new technologies could bring unanticipated cyber risks although the technologies were desired to introduce benefits. The challenges were generally attached to technical vulnerabilities of new systems, uncertain business process designs, and high complexity of governance.
- Data analytic techniques in FinTech are facing both opportunities and challenges. The opportunities deriving from data-oriented approaches include: (1) FSIs have a great demand of using data analytic techniques to create/improve their business values and most FSIs have abundant data accesses. (2) The scope of data analytics covers more technical dimensions than it was before. A few emerging or popular techniques are exceptionally driving the development of FinTech, such as computations on heterogeneous computing, ontology-based information retrievals, and machine learning.

Concurrently, the challenging issues involve a variety of problems caused by the data limitations, algorithm designs, and hardware restrictions. The determination of the target dataset is demanding in the big data context, by which the data mining results can be influenced. The enhancement of the computation efficiency can be addressed by both algorithm improvements and hardware upgrades.

- In the perspective of facility and equipment, FinTech is dealing with the emerging challenges created by the new execution environment, such as wireless networks and mobile computing. System integrations, multimedia communications, and distributed data storage are three pivotal aspects of the successful financial service deliveries. Most contemporary work pay attention to the optimizations of the facility implementations.
- FinTech applications are playing a facilitator role in assisting financial businesses. The extent of FinTech applications comprises both the financial service improvement and financial management assistance. Web-based technologies have propelled FinTech applications to a complicated and integrated operating environment, which consists of online transactions, data tracking, real-time services, merchandise frauds, communication quality management, market predictions, and decision-making supports.
- The expansion of FinTech brings a variety of service models to FSIs. On one side, from the perspective of the internal impact, modern FinTech spreads enable large sized FSIs to establish an expected business ecosystem for the purpose of broadening markets and increasing customer-reaching. A few techniques include data visualizations and application streaming. One challenge is integrating data from distinct sources and operating data mining on different platforms. On the other side, FinTech also has external impacts. Emerging service models from cloud computing or smart city aid FSIs to obtain higher level capabilities in knowledge discovery and business awareness.

7.2. Data-driven FinTech framework

Stemming from our survey, we found that the critical section of FinTech is mostly data-related, from using data for value creations to formulating data trading mechanisms, from designing data mining algorithms to developing financial privacy protection methods. This phenomenon is associated with the feature of the financial industry, which is broadly offering electronic financial services in which data provide service representations with content sources. Attached to this characteristic, we propose the *Data-Driven FinTech Framework* (DF2) to locate the crucial knowledge structure of FinTech and guide the executions for both practitioners and academics. Fig. 6 represents the workflow of the proposed DF2.

As illustrated in the figure, we consider the framework the computing reference and formulate four dimensions of FinTech that

are driven by the data usage, which include *Efficiency*, *Accuracy*, *Energy*, and *Security and Privacy* (S & P). The major role of data in FinTech has two sides. The first role is the financial service itself that delivers data to financial customers. The other role is the tool of assisting FSIs to improve their businesses by either upgrading the existing services or creating new offerings. Data-driven techniques can provide FSIs with the reference plan, such as using data analytic techniques to predict market trends, which are generally associated with the practical problem formulations. Formulating problems need to place one of the dimensions and consider the service deployment. Some popular service deployments include cloud computing, big data, and mobile computing, by which the manner of the data utilizations can be impacted.

In addition, the meanings of these four dimensions are: (1) The efficiency dimension refers to the efficiency of the data processing so that it is generally associated with optimizations or artificial intelligence. The efficiency can be influenced by both software and hardware. In addition, considering the feature of the financial industry, information management is one of the cores for value acquisitions; therefore, data transmissions and information retrievals are two crucial aspects in this dimension. (2) The accuracy dimension addresses the results obtained from data processing, which determines the service quality and FSIs' resilience. It can be either the financial service offerings, such as financial transactions, or the additional value creation, such as data mining on the customer data. In general, the accuracy issue is associated with two aspects, which are proper data collections and appropriate quantitative finance model or artificial intelligent algorithms. (3) The energy dimension focuses on the energy-saving solutions, which can cover both applications and infrastructure, too. Energy-related data collections and data-driven resource management are two fundamental aspects of achieving green finance. A number of novel techniques for alternatives of the financial system designs are being developed, such as centralized computing, fog computing, and edge computing. (4) The S & P dimension covers two aspects, which are protecting data and using data to protect systems. This dimension is attached to one of the core values of the financial service, which is protecting financial customers' privacy. The security and privacy issues exist at multiple layers and the challenges become more daunting when emerging technologies are employed.

Moreover, we propose a few research directions on the basis of our survey to guide our future work.

1. Design optimization algorithms for speeding up the efficiency of financial big data processing in various scenarios or distinct execution phases. This advance can tightly bridge up FSIs with their customers from quick information acquisitions and accurate quantitative finance. For example, the optimization can locate at either the task scheduling on-premise or the task assignment in cloud computing. A few challenges required to be solved include operating large sized data integrations, increasing the effectiveness of the training process for deep learning, and developing optimal task scheduling approaches at the application layer.
2. Develop financial-purpose hardware to increase the execution efficiency of big data processing. The specific-designed infrastructure for the financial service purpose needs to meet the characteristics of the financial big data processing, such as parallel processing. The development of FinTech hardware should cover a few technical directions, such as the heterogeneous computing, *General-Purpose computing on Graphics Processing Units* (GPGPU), *Massive Distributed Storage* (MDS), and high-performance communication mechanism.
3. Further improve the accuracy of the data analytics by sharpening data analytic algorithms at distinct financial operation layers. The outcomes of the data analytics should have a higher capability to assist FSIs in making decisions on market predictions, finance awareness, investment management, and strategy-making.
4. Study and propose energy-aware algorithms for different business

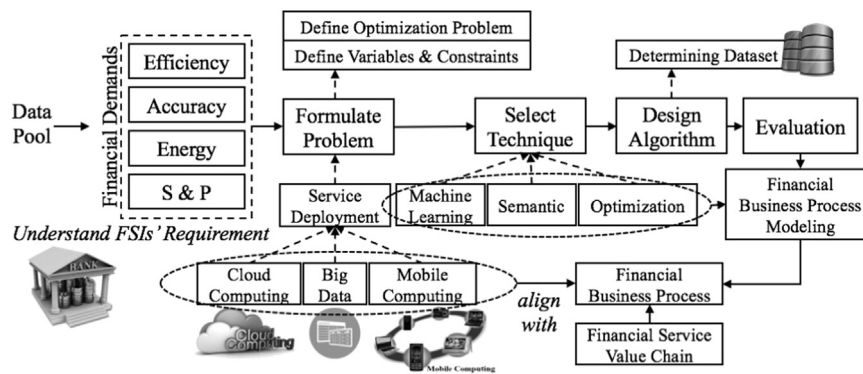


Fig. 6. Workflow for the Data-Driven FinTech Framework (DF2).

processes and operations in FinTech. The dynamic task assignment mechanism is an alternative solution to building up energy-aware centralized data centers. Meanwhile, the optimized data transmissions over the wireless networks can reduce the energy cost, such as using *Cognitive Wireless Networks* (CWN).

5. Protect financial data and privacy throughout multiple security layers, from data transmissions to data processing. In cloud computing, the security mechanism needs not only protect data from the outsider attacks but also should protect data from insider attacks. Developing a *Fully Homomorphic Encryption* (FHE) has an urgent demand for secure cloud-based financial solutions. In the wireless networking environment, data transmissions need to deal with multiple adversaries even though the size of the transmissions has become substantial. Application-enabled encryption mechanisms that selectively encrypt data should be designed for certain financial data transfers.

In summary, four major aspects of FinTech are involved in our proposed framework. Specifically speaking, the core of FinTech is about data and security. On one hand, contemporary FSIs seek additional values from a large pool of data in a wide scope, such as the strategy-making assistance, environmental-friendly solution, service improvement, business collaboration support, risk prediction, and financial operation aid. FinTech is playing a critical value creator in the value chain for most current FSIs. On the other hand, FSIs also need to ensure that the data are used in a correct manner all the time, which introduces security and privacy concerns when applying FinTech in the financial industry. This is not only a vital concern of FSIs but also a key issue for all financial customers. A FinTech solution can be deployed only when the developers take action to prevent predictable threats and establish a secure mechanism for those unpredictable risks.

8. Conclusions

This paper completed a survey on five key technical aspects of FinTech for understanding contemporary development of the discipline and guiding future researches. Five technical aspects included data-oriented techniques, facility and equipment development, application designs, service models placement, and security and privacy protections. We proposed the *Data-Driven FinTech Framework* (DF2) to facilitate and standardize future FinTech researches and technical deployments. Finally, we suggested a few research directions of FinTech deriving from our main findings.

References

- Abawajy, J., Wang, G., Yang, L., Javadi, B., 2016. Trust, security and privacy in emerging distributed systems. *Future Gener. Comput. Syst.* 55, 224–226, (C).
- Ahmed, M., Mahmood, A., Islam, R., 2016. A survey of anomaly detection techniques in financial domain. *Future Gener. Comput. Syst.* 55, 278–288, (C).

- Asamoah, C., Tao, L., Gai, K., Jiang, N., 2016. Powering filtration process of cyber security ecosystem using knowledge graph. In: *Proceedings of the 2nd IEEE International Conference of Scalable and Smart Cloud*, pp. 240–246.
- Banu, R., Nagaveni, N., 2013. Evaluation of a perturbation-based technique for privacy preservation in a multi-party clustering scenario. *Inf. Sci.* 232, 437–448.
- Basole, R., Clear, T., Hu, M., Mehrotra, H., Stasko, J., 2013. Understanding interfirm relationships in business ecosystems with interactive visualization. *IEEE Trans. Vis. Comput. Graph.* 19 (12), 2526–2535.
- Basu, A., Gandhi, J., Chang, J., Hill, M., Swift, M., 2013. Efficient virtual memory for big memory servers. In: *Proceedings of the 40th Annual International Symposium on Computer Architecture*, pp. 237–248, Tel-Aviv, Israel. ACM.
- Bernstein, A., Provost, F., Hill, S., 2005. Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification. *IEEE Trans. Knowl. data Eng.* 17 (4), 503–518.
- Bilogrevic, I., Jadhwal, M., Joneja, V., Kalkan, K., Hubaux, J., Aad, I., 2014. Privacy-preserving optimal meeting location determination on mobile devices. *IEEE Trans. IFS* 9 (7), 1141–1156.
- Bugiel, S., Heuser, S., Sadeghi, A., 2013. Flexible and fine-grained mandatory access control on android for diverse security and privacy policies. In: *Proceedings of the 22nd USENIX Security Symposium (USENIX Security 13)*, Washington, DC, USA, pages 131–146, .
- Cao, M., Chyckyla, R., Stewart, T., 2015. Big data analytics in financial statement audits. *Account. Horiz.* 29 (2), 423–429.
- Casado, R., Younas, M., 2015. Emerging trends and technologies in big data processing. *Concurr. Comput.: Pract. Exp.* 27 (8), (2078?2091).
- Castiglione, A., De Santis, A., Soriente, C., 2007. Taking advantages of a disadvantage: Digital forensics and steganography using document metadata. *J. Syst. Softw.* 80 (5), 750–764.
- Castiglione, A., Pizzolante, R., De Santis, A., Carpentieri, B., Castiglione, A., Palmieri, F., 2015. Cloud-based adaptive compression and secure management services for 3D healthcare data. *Future Gener. Comput. Syst.* 43, 120–134.
- Chai, S., Kim, M., Rao, H., 2011. Firms' information security investment decisions: stock market evidence of investors' behavior. *Decis. Support Syst.* 50 (4), 651–661.
- Chandola, V., Banerjee, A., Kumar, V., 2012. Anomaly detection for discrete sequences: a survey. *IEEE Trans. Knowl. Data Eng.* 24 (5), 823–839.
- Chang, V., Ramachandran, M., 2016. Towards achieving data security with the cloud computing adoption framework. *IEEE Trans. Serv. Comput.* 9 (1), 138–151.
- Chen, S., Weiss, G., 2014. An intelligent agent for bilateral negotiation with unknown opponents in continuous-time domains. *ACM Trans. Auton. Adapt. Syst. (TAAS)* 9 (3), 16.
- Chen, C., Zhang, C., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Inform. Sci.* 275, 314–347.
- Chen, K., Huang, S., Zheng, L., Poor, H., 2015. Communication theoretic data analytics. *IEEE J. Sel. Areas Commun.* 33 (4), 663–675.
- Choi, C., Choi, J., Kim, P., 2014. Ontology-based access control model for security policy reasoning in cloud computing. *J. Supercomput.* 67 (3), 711–722.
- SVB. Cybersecurity report 2015, 2015. Retrieve from url=(https://www.svb.com/uploadedFiles/Content/Trends_and_Insights/Reports/Cybersecurity_Report/cybersecurity-report-2015.pdf).
- Davis, M., Kumiega, A., Van, V., 2013. Ethics, finance, and automation: a preliminary survey of problems in high frequency trading. *Sci. Eng. Ethics* 19 (3), 851–874.
- DeStefano, R., Tao, L., Gai, K., 2016. Improving data governance in large organizations through ontology and linked data. In: *Proceedings of the 2nd IEEE International Conference of Scalable and Smart Cloud*, pages 279–284. IEEE.
- Ding, X., Tian, Y., Yu, Y., 2016. A real-time big data gathering algorithm based on indoor wireless sensor networks for risk analysis of industrial operations. *IEEE Trans. Ind. Inform.* 12 (3), 1232–1242.
- Duan, L., Da, X., 2012. Business intelligence for enterprise systems: a survey. *IEEE Trans. Ind. Inform.* 8 (3), 679–687.
- Elnagdy, S., Qiu, M., Gai, K., 2016. Cyber incident classifications using ontology-based knowledge representation for cybersecurity insurance in financial industry. In: *Proceedings of the 2nd IEEE International Conference of Scalable and Smart Cloud*, pages 301–306.
- Elnagdy, S., Qiu, M., Gai, K., 2016. Understanding taxonomy of cyber risks for cybersecurity insurance of financial industry in cloud computing. In: *Proceedings of*

- the 2nd IEEE International Conference of Scalable and Smart Cloud, pages 295–300.
- Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B., Cox, L., Jung, J., McDaniel, P., Sheth, A., 2014. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Trans. Comput. Syst.* 32 (2), 5.
- Fang, B., Zhang, P., 2016. Big data in finance. In: *Big Data Concepts, Theories, and Applications*, pp. 391–412. Springer.
- Farzan, F., Lahiri, S., Kleinberg, M., Gharieh, K., Farzan, F., Jafari, M., 2013. Microgrids for fun and profit: The economics of installation investments and operations. *IEEE Power Energy Mag.* 11 (4), 52–58.
- Ferrer, A., Hernández, F., Tordsson, J., Elmroth, E., Ali-Eldin, A., Zsigri, C., Sirvent, R., Guitart, J., Badia, R., Djemame, K., et al., 2012. OPTIMIS: a holistic approach to cloud service provisioning. *Future Gener. Comput. Syst.* 28 (1), 66–77.
- Gai, K., Li, S., 2012. Towards cloud computing: a literature review on cloud computing and its development trends. In: *Proceedings of the 4th IEEE International Conference on Multimedia Information Networking and Security*, Nanjing, China. pages 142–146.
- Gai, K., Steenkamp, A., 2013. 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.
- Gai, K., Steenkamp, A., 2014. A feasibility study of Platform-as-a-Service using cloud computing for a global service organization. *J. Inf. Syst. Appl. Res.* 7, 28–42.
- Gai, K., Qiu, M., Zhao, H., Dai, W., 2015. Anti-counterfeit schema using monte carlo simulation for e-commerce in cloud systems. In: *Proceedings of the 2nd IEEE International Conference on Cyber Security and Cloud Computing*, New York, USA. IEEE, pages 74–79.
- Gai, K., Qiu, M., Zhao, H., Tao, L., Zong, Z., 2015. Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. *J. Netw. Comput. Appl.* 59, 46–54.
- Gai, K., Qiu, M., Chen, L., Liu, M., 2015. Electronic health record error prevention approach using ontology in big data. In: *Proceedings of the 17th IEEE International Conference on High Performance Computing and Communications*, New York, USA pages 752–757.
- Gai, K., Qiu, M., Thuraisingham, B., Tao, L., 2015. Proactive attribute-based secure data schema for mobile cloud in financial industry. In: *Proceedings of IEEE International Symposium on Big Data Security*, New York, USA, pp. 1332–1337.
- Gai, K., Qiu, M., Tao, L., Zhu, Y., 2016a. Intrusion detection techniques for mobile cloud computing in heterogeneous 5G. *Secur. Commun. Netw.* 9 (16), 3049–3058.
- Gai, K., Qiu, M., Zhao, H., 2016b. Cost-aware multimedia data allocation for heterogeneous memory using genetic algorithm in cloud computing. *IEEE Trans. Cloud Comput.* 99, 1–11.
- Gai, K., Qiu, M., Qiu, H., Liu, M., 2016. Energy-aware optimal task assignment for mobile heterogeneous embedded systems in cloud computing. In: *Proceedings of the 3rd IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)*, pages 198–203, Beijing, China.
- Gai, K., Qiu, M., Elnagdy, S., 2016. A novel secure big data cyber incident analytics framework for cloud-based cybersecurity insurance. In: *Proceedings of the 2nd IEEE International Conference on Big Data Security on Cloud*, New York, USA, pages 171–176.
- Gai, K., Qiu, M., Zhao, H., Xiong, J., 2016. Privacy-aware adaptive data encryption strategy of big data in cloud computing. In: *Proceedings of the 2nd IEEE International Conference of Scalable and Smart Cloud (SSC 2016)*, Beijing, China. IEEE, pages 273–278.
- Gai, K., Qiu, M., Zhao, H., Dai, W., 2016. Privacy-preserving adaptive multi-channel communications under timing constraints. In: *Proceedings of IEEE International Conference on Smart Cloud 2016*, New York, USA. IEEE, pages 190–195.
- Gai, K., Qiu, M., Hassan, H., 2016c. Secure cyber incident analytics framework using Monte Carlo simulations for financial cybersecurity insurance in cloud computing. *Concurr. Comput. Pract. Exper.* 29 (7), 1.
- Gai, K., Qiu, M., Sun, X., Zhao, H., 2016. Security and privacy issues: A survey on fintech. In: *Proceedings of IEEE International Conference on Smart Computing and Communication*, Shenzhen, China. Springer, pages 236–247.
- Gai, K., Qiu, M., Zhao, H., 2016. Security-aware efficient mass distributed storage approach for cloud systems in big data. In: *Proceedings of the 2nd IEEE International Conference on Big Data Security on Cloud*, New York, USA, pages 140–145.
- Gai, K., Qiu, M., Elnagdy, S., 2016. Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data. In: *Proceedings of the 2nd IEEE International Conference on Big Data Security on Cloud*, New York, USA, pages 197–202.
- Gai, K., Qiu, M., Zhao, H., Qiu, L., 2016. Smart energy-aware data allocation for heterogeneous memory. In: *Proceedings of the 18th IEEE International Conference on High Performance Computing and Communications*, Sydney, Australia, pages 136–143.
- Gai, K., Qiu, M., Chen, M., Zhao, H., 2017a. SA-EAST: security-aware efficient data transmission for ITS in mobile heterogeneous cloud computing. *ACM Trans. Embed. Comput. Syst.* 16 (2), 60.
- Gai, K., Qiu, M., Ming, Z., Zhao, H., Qiu, L., 2017b. Spoofing-jamming attack strategy using optimal power distributions in wireless smart grid networks. *IEEE Trans. Smart Grid* 8 (5), 2431–2439.
- Gai, K., Qiu, M., Zhao, H., 2018. Energy-aware task assignment for mobile cyber-enabled applications in heterogeneous cloud computing. *J. Parallel Distrib. Comput.* 111, 126–135.
- Gai, K., 2014. A review of leveraging private cloud computing in financial service institutions: value propositions and current performances. *Int. J. Comput. Appl.* 95 (3), 40–44.
- Gai, K., Du, Z., Qiu, M., Zhao, H., 2015. Efficiency-aware workload optimizations of heterogeneous cloud computing for capacity planning in financial industry. In: *Proceedings of the 2nd IEEE International Conference on Cyber Security and Cloud Computing*, New York, USA, pages 1–6.
- Georgakoudis, G., Gillan, C., Sayed, A., Spence, I., Faloan, R., Nikolopoulos, D., 2016. Methods and metrics for fair server assessment under real-time financial workloads. *Concurr. Comput.: Pract. Exp.* 28 (3), 916–928.
- Geslevich-Packin, N., Lev-Aretz, Y., 2016. Big data and social netbanks: what happens when tech companies become financial companies? *ACM SIGCAS Comput. Soc.* 46 (1), 36–40.
- Glancy, F., Yadav, S., 2011. A computational model for financial reporting fraud detection. *Decis. Support Syst.* 50 (3), 595–601.
- Hadavandi, E., Shavandi, H., Ghanbari, A., 2010. Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowl.-Based Syst.* 23 (8), 800–808.
- Hao, Y., Zhang, Y., Cao, J., 2010. Web services discovery and rank: An information retrieval approach. *Future Gener. Comput. Syst.* 26 (8), 1053–1062.
- Hernandez, P., 2016. Microsoft, Bank of America announce Blockchain collaboration, 2016. Retrieve from url=(<http://www.eweek.com/cloud/microsoft-bank-of-america-announce-blockchain-collaboration.html>).
- Islam, S., Keung, J., Lee, K., Liu, A., 2012. Empirical prediction models for adaptive resource provisioning in the cloud. *Future Gener. Comput. Syst.* 28 (1), 155–162.
- Jan T., Wang, D., Lin, C., Lin, H., 2012. A simple methodology for soft cost-sensitive classification. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, China. ACM, pages 141–149.
- Jiang, S., Qian, X., Mei, T., Fu, Y., 2016. Personalized travel sequence recommendation on multi-source big social media. *IEEE Trans. Big Data* 2 (1), 43–56.
- Kanak, A., Sogukpinar, I., 2014. BioPSM: a formal model for privacy, security, and trust in template-protecting biometric authentication. *Secur. Commun. Netw.* 7 (1), 123–138.
- Kandel, S., Paepcke, A., Hellerstein, J., Heer, J., 2012. Enterprise data analysis and visualization: An interview study. *IEEE Trans. Vis. Comput. Graph.* 18 (12), 2917–2926.
- Kou, G., Peng, Y., Wang, G., 2014. Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Inf. Sci.* 275, 1–12.
- Kousiouris, G., Menychtas, A., Kyriazis, D., Gogouvitis, S., Varvarigou, T., 2014. Dynamic, behavioral-based estimation of resource provisioning based on high-level application terms in cloud platforms. *Future Gener. Comput. Syst.* 32, 27–40.
- Löhr, S., Mursajew, O., Rösch, D., Scheule, H., 2013. Dynamic implied correlation modeling and forecasting in structured finance. *J. Futures Mark.* 33 (11), 994–1023.
- Lee, T., Kim, H., 2015. An exploratory study on Fintech industry in Korea: Crowd funding case. In: *Proceedings of 2nd International Conference on Innovative Engineering Technologies*, Bangkok, Thailand, pages 58–64.
- Lee, W., Leung, C., Lee, J., 2011. Mobile web navigation in digital ecosystems using rooted directed trees. *IEEE Trans. Inf. Electron.* 58 (6), 2154–2162.
- Li, B., Hoi, S., 2014. Online portfolio selection: a survey. *ACM Comput. Surv.* 46 (3), 35.
- Li, M., Yu, S., Ren, K., Lou, W., Hou, Y., 2013. Toward privacy-assured and searchable cloud data storage services. *IEEE Netw.* 27 (4), 56–62.
- Li, Y., Dai, W., Ming, Z., Qiu, M., 2015. Privacy protection for preventing data over-collection in smart city. *IEEE Trans. Comput.* 65, 1339–1350.
- Li, Y., Gai, K., Qiu, M., Dai, W., Liu, M., 2016a. Adaptive human detection approach using FPGA-based parallel architecture in reconfigurable hardware. *Concurr. Comput.: Pract. Exp.* 29 (14), 1.
- Li, Y., Gai, K., Ming, Z., Zhao, H., Qiu, M., 2016b. Intercrossed access control for secure financial services on multimedia big data in cloud systems. *ACM Trans. Multimed. Comput. Commun. Appl.* 12, 67, (4s).
- Li, Y., Gai, K., Qiu, L., Qiu, M., Zhao, H., 2017. Intelligent cryptography approach for secure distributed big data storage in cloud computing. *Inf. Sci.* 387, 103–115.
- Liang, H., Gai, K., 2015. Internet-based anti-counterfeiting pattern with using big data in China. In: *Proceedings of the IEEE International Symposium on Big Data Security on Cloud*, New York, USA, pp. 1387–1392.
- Liao, C., Liu, C., Chen, K., 2011. Examining the impact of privacy, trust and risk perceptions beyond monetary transactions: an integrated model. *ECRA* 10 (6), 702–715.
- Liu, H., Abraham, A., Hassanien, A., 2010. Scheduling jobs on computational grids using a fuzzy particle swarm optimization algorithm. *Future Gener. Comput. Syst.* 26 (8), 1336–1343.
- Liu, C., Chen, J., Yang, L., Zhang, X., Yang, C., Ranjan, R., Kotagiri, R., 2014. Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates. *IEEE Trans. Parallel Distrib. Syst.* 25 (9), 2234–2244.
- Liu, S., Cui, W., Wu, Y., Liu, M., 2014. A survey on information visualization: recent advances and challenges. *Vis. Comput.* 30 (12), 1373–1393.
- Lo, H., Lin, S., Wang, H., 2014. Generalized k-labelsets ensemble for multi-label and cost-sensitive classification. *IEEE Trans. Knowl. Data Eng.* 26 (7), 1679–1691.
- Lu, Y., Roychowdhury, V., Vandenbergh, L., 2008. Distributed parallel support vector machines in strongly connected networks. *IEEE Trans. Neural Netw.* 19 (7), 1167–1178.
- Lu, R., Lin, X., Liang, X., Shen, X., 2012. A dynamic privacy-preserving key management scheme for location-based services in VANETs. *IEEE Trans. Intell. Transp. Syst.* 13 (1), 127–139.
- Lugowski, A., Kamil, S., Buluç, A., Williams, S., Duriakova, E., Oliker, L., Fox, A., Gilbert, J., 2015. Parallel processing of filtered queries in attributed semantic graphs. *J. Parallel Distrib. Comput.* 79, 115–131.
- Ma, J., Xu, W., Sun, Y., Turban, E., Wang, S., Liu, O., 2012. An ontology-based text-mining method to cluster proposals for research project selection. *IEEE Trans. Syst., Man, Cybern.-Part A: Syst. Hum.* 42 (3), 784–790.

- Ma, D., Saxena, N., Xiang, T., Zhu, Y., 2013. Location-aware and safer cards: enhancing RFID security and privacy via location sensing. *IEEE TDSC* 10 (2), 57–69.
- Ma, L., Tao, L., Gai, K., Zhong, Y., 2016. A novel social network access control model using logical authorization language in cloud computing. *CCPE* 99, 1.
- Ma, L., Tao, L., Zhong, Y., Gai, K., 2016. RuleSN: Research and application of social network access control model. In: *Proceedings of the IEEE International Conference on Intelligent Data and Security*, New York, USA. pages 418–423.
- Mashayekhy, L., Nejad, M., Grosu, D., 2014. A two-sided market mechanism for trading big data computing commodities. In: *Proceedings of International Conference on Big Data*, Washington, DC. pages 153–158.
- Merigó, J., Gil-Lafuente, A., 2010. New decision-making techniques and their application in the selection of financial products. *Inf. Sci.* 180 (11), 2085–2094.
- Misra, S., Das, S., Khatua, M., Obaidat, M., 2014. QoS-guaranteed bandwidth shifting and redistribution in mobile cloud environment. *IEEE Trans. Cloud Comput.* 2 (2), 181–193.
- Morgan, S., 2015. Cybersecurity market reaches \$75 billion in 2015; expected to reach \$170 billion by 2020.
- Nabeel, M., Shang, N., Bertino, E., 2013. Privacy preserving policy-based content sharing in public clouds. *IEEE Trans. Knowl. Data Eng.* 25 (11), 2602–2614.
- Ni, J., Yu, Y., Mu, Y., Xia, Q., 2014. On the security of an efficient dynamic auditing protocol in cloud storage. *IEEE Trans. Parallel Distrib. Syst.* 25 (10), 2760–2761.
- Nussbaumer, P., Matter, I., Schwabe, G., 2012. Enforced vs. casual transparency-findings from IT-supported financial advisory encounters. *ACM Trans. Manag. Inform. Syst.* 3 (2), 11.
- Paliwal, A., Shafiq, B., Vaidya, J., Xiong, H., Adam, N., 2012. Semantics-based automated service discovery. *IEEE Trans. Serv. Comput.* 5 (2), 260–275.
- Putthal, D., Nepal, S., Ranjan, R., Chen, J., 2015. DPBSV-an efficient and secure scheme for big sensing data stream. In *Trustcom/BigDataSE/ISPA*, volume 1, pp. 246–253, Helsinki, Finland. IEEE.
- Qiu, M., Cao, D., Su, H., Gai, K., 2015a. Data transfer minimization for financial derivative pricing using Monte Carlo simulation with GPU in 5G. *Int. J. Commun. Syst.* 11 (16), 2364–2374.
- Qiu, M., Zhong, M., Li, J., Gai, K., Zong, Z., 2015b. Phase-change memory optimization for green cloud with genetic algorithm. *IEEE Trans. Comput.* 64 (12), 3528–3540.
- Qiu, L., Gai, K., Qiu, M., 2016. Optimal big data sharing approach for tele-health in cloud computing. In: *Proceedings of the IEEE International Conference on Smart Cloud 2016*, pages 184–189, New York, USA. IEEE.
- Qiu, M., Gai, K., Thuraisingham, B., Tao, L., Zhao, H., 2016. Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry. *Future Gener. Comput. Syst.*, PP 1.
- Roumani, Y., Nwankpa, J., Roumani, Y., 2016. Examining the relationship between firm's financial records and security vulnerabilities. *Int. J. Inf. Manag.* 36 (6), 987–994.
- Ruj, S., Stojmenovic, M., Nayak, A., 2012. Privacy preserving access control with authentication for securing data in clouds. In: *Proceedings of the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, Ottawa, Canada. pages 556–563.
- Sánchez, R., Almenares, F., Arias, P., Díaz-Sánchez, D., Marín, A., 2012. Enhancing privacy and dynamic federation in IdM for consumer cloud computing. *IEEE Trans. Consum. Electron.* 58 (1), 95–103.
- Sagraves, A., Connors, G., 2017. Capturing the value of data in banking. *Appl. Mark. Anal.* 2 (4), 304–311.
- Sanz, José A., Bernardo, D., Herrera, F., Bustince, H., Hagrás, H., 2015. A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data. *IEEE Trans. Fuzzy Syst.* 23 (4), 973–990.
- Satyanarayan, A., Russell, R., Hoffswell, J., Heer, J., 2016. Reactive vega: a streaming dataflow architecture for declarative interactive visualization. *IEEE Trans. Vis. Comput. Graph.* 22 (1), 659–668.
- Schneider, S., Hirzel, M., Gedik, B., Wu, K., 2015. Safe data parallelism for general streaming. *IEEE Trans. Comput.* 64 (2), 504–517.
- Shao, J., Lu, R., Lin, X., 2014. FINE: a fine-grained privacy-preserving location-based service framework for mobile devices. In: *Proceedings of IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, Toronto, Canada. IEEE. pages 244–252.
- Sharma, A., Panigrahi, P., 2013. A review of financial accounting fraud detection based on data mining techniques. *arXiv:1309.3944*.
- Shepperd, M., Song, Q., Sun, Z., Mair, C., 2013. Data quality: some comments on the NASA software defect datasets. *IEEE Trans. Softw. Eng.* 39 (9), 1208–1215.
- Shi, X., Zhang, P., Khan, S., 2017. Quantitative data analysis in finance. *Handbook of Big Data Technologies*, pp. 719–753.
- Shim, Y., Shin, D., 2016. Analyzing China's fintech industry from the perspective of actor – network theory. *Telecommun. Policy* 40 (2), 168–181.
- Shuttlewood, P., Volin, M., Wozniak, L., 2016. Global fintech investment growth continues in 2016 driven by europe and asia, accenture study finds. [url=\(https://newsroom.accenture.com/news/global-fintech-investment-growth-continues-in-2016-driven-by-europe-and-asia-accenture-study-finds.htm\)](https://newsroom.accenture.com/news/global-fintech-investment-growth-continues-in-2016-driven-by-europe-and-asia-accenture-study-finds.htm).
- Smari, W., Clemente, P., Lalande, J., 2014. An extended attribute based access control model with trust and privacy: application to a collaborative crisis management system. *Future Gener. Comput. Syst.* 31, 147–168.
- Sun, D., Zhong, J., 2012. A hash-based RFID security protocol for strong privacy protection. *IEEE Trans. Consum. Electron.* 58 (4), 1246–1252.
- Sun, W., Yu, S., Lou, W., Hou, Y., Li, H., 2014. Protecting your right: Attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud. In: *Proceedings of IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, Toronto, Canada. IEEE. pages 226–234.
- Tang, Y., Lee, P., Lui, J., Perlman, R., 2012. Secure overlay cloud storage with access control and assured deletion. *IEEE Trans. Dependable Secur. Comput.* 9 (6), 903–916.
- Tao, F., Zuo, Y., Da, X., Zhang, L., 2014. IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing. *IEEE Trans. Ind. Inform.* 10 (2), 1547–1557.
- Tominski, C., Schumann, H., Andrienko, G., Andrienko, N., 2012. Stacking-based visualization of trajectory attribute data. *IEEE Trans. Vis. Comput. Graph.* 18 (12), 2565–2574.
- Wang, H., Mylopoulos, J., Liao, S., 2002. Intelligent agents and financial risk monitoring systems. *Commun. ACM* 45 (3), 83–88.
- Wang, D., Liu, X., Wang, M., 2013. A DT-SVM strategy for stock futures prediction with big data. In: *Proceedings of the 16th International Conference on Computational Science and Engineering*, Sydney, Australia. pp. 1005–1012.
- Wang, Z., Lu, M., Yuan, X., Zhang, J., Wetering, H., 2013. Visual traffic jam analysis based on trajectory data. *IEEE Trans. Vis. Comput. Graph.* 19 (12), 2159–2168.
- Wang, J., Zhao, P., Hoi, S., 2014. Cost-sensitive online classification. *IEEE Trans. Knowl. Data Eng.* 26 (10), 2425–2438.
- Wang, L., Da, X., Bi, Z., Xu, Y., 2014. Data cleaning for RFID and WSN integration. *IEEE Trans. Ind. Inform.* 10 (1), 408–418.
- Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., Liu, H., 2014. E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures. In *Proceedings of the 20th Annual International Conference on MCN*, pages 617–628, Maui, Hawaii, USA. ACM.
- Wang, B., Li, B., Li, H., 2014. Oruta: privacy-preserving public auditing for shared data in the cloud. *IEEE Trans. cloud Comput.* 2 (1), 43–56.
- Wang, L., 2015. Dynamical models of stock prices based on technical trading rules - part iii: Application to hong kong stocks. *IEEE Trans. Fuzzy Syst.* 23 (5), 1680–1697.
- Wei, C., Hwang, M., Chin, A., 2011. A mutual authentication protocol for RFID. *IT Prof. Mag.* 13 (2), 20.
- Wen, S., Zhou, W., Zhang, J., Xiang, Y., Zhou, W., Jia, W., 2013. Modeling propagation dynamics of social network worms. *IEEE Trans. Parallel Distrib. Syst.* 24 (8), 1633–1643.
- Wigglesworth, R., 2016. Fintech: Search for a super-algo, January. [url=\(https://www.ft.com/content/5eb91614-bee5-11e5-846f-79b0e3d20eaf\)](https://www.ft.com/content/5eb91614-bee5-11e5-846f-79b0e3d20eaf).
- Wu, X., Zhu, X., Wu, G., Ding, W., 2014. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* 26 (1), 97–107.
- Wu, T., Dou, W., Wu, F., Tang, S., Hu, C., Chen, J., 2016. A deployment optimization scheme over multimedia big data for large-scale media streaming application. *ACM Trans. Multimed. Comput., Commun., Appl.* 12, 73, (5s).
- Khafa, F., Abraham, A., 2010. Computational models and heuristic methods for grid scheduling problems. *Future Gener. Comput. Syst.* 26 (4), 608–621.
- Xia, Z., Wang, X., Zhang, L., Qin, Z., Sun, X., Ren, K., 2016. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Trans. IFS* 11 (11), 2594–2608.
- Xiao, L., Yan, Q., Lou, W., Chen, G., Hou, Y., 2013. Proximity-based security techniques for mobile users in wireless networks. *IEEE Trans. IFS* 8 (12), 2089–2100.
- Xu, K., Yue, H., Guo, L., Guo, Y., Fang, Y., 2015. Privacy-preserving machine learning algorithms for big data systems. In: *Proceedings of the 35th International Conference on Distributed Computing Systems*, Columbus, OH, USA, pages 318–327.
- Yin, H., Gai, K., 2015. An empirical study on preprocessing high-dimensional class-imbalanced data for classification. In: *Proceedings of the IEEE International Symposium on Big Data Security on Cloud*, New York, USA, pp. 1314–1319.
- Yin, S., Zhu, X., Kaynak, O., 2015. Improved PLS focused on key-performance-indicator-related fault diagnosis. *IEEE Trans. Ind. Electron.* 62 (3), 1651–1658.
- Yu, W., Yan, C., Ding, Z., Jiang, C., Zhou, M., 2014. Modeling and validating e-commerce business process based on petri nets. *IEEE Trans. Syst., Man, Cybern.: Syst.* 44 (3), 327–341.
- Yu, K., Gao, Y., Zhang, P., Qiu, M., 2015. Design and architecture of dell acceleration appliances for database (DAAD): A practical approach with high availability guaranteed. In: *Proceedings of the 17th IEEE International Conference on High Performance Computing and Communications*, New York, USA, pages 430–435.
- Zarandi, M., Hadavandi, E., Turksen, I., 2012. A hybrid fuzzy intelligent agent-based system for stock price prediction. *Int. J. Intell. Syst.* 27 (11), 947–969.
- Zhang, Y., Soong, B., 2004. Performance evaluation of GSM/GPRS networks with channel re-allocation scheme. *IEEE Commun. Lett.* 8 (5), 280–282.
- Zhang, Y., Yu, R., Xie, S., Yao, W., Xiao, Y., Guizani, M., 2011. Home M2M networks: architectures, standards, and QoS improvement. *IEEE Commun. Mag.* 49 (4), 44–52.
- Zhang, J., Chen, C., Xiang, Y., Zhou, W., Xiang, Y., 2013. Internet traffic classification by aggregating correlated naive Bayes predictions. *IEEE Trans. Inf. Forensics Secur.* 8 (1), 5–15.
- Zhang, L., Luo, Y., Tao, F., Li, B., Ren, L., Zhang, X., Guo, H., Cheng, Y., Hu, A., Liu, Y., 2014. Cloud manufacturing: a new manufacturing paradigm. *Enterp. Inf. Syst.* 8 (2), 167–187.
- Zhang, P., Liu, L., Deng, Y., 2015. A data-driven paradigm for mapping problems. *Parallel Comput.* 48, 108–124.
- Zhang, Q., Yang, L., Chen, Z., 2016. Privacy preserving deep computation model on cloud for big data feature learning. *IEEE Trans. Comput.* 65 (5), 1351–1362.
- Zhang, P., Yu, K., Yu, J., Khan, S., 2017. QuantCloud: big data infrastructure for quantitative finance on the cloud. *IEEE Trans. Big Data*, PP 99, 1.