Hindawi
Complexity
Volume 2017, Article ID 6254842, 11 pages
https://doi.org/10.1155/2017/6254842

Research Article

Covering the Monitoring Network: A Unified Framework to Protect E-Commerce Security

Lirong Qiu and Jie Li

School of Information Engineering, Minzu University of China, Beijing, China

Correspondence should be addressed to Lirong Qiu; qiu_lirong@126.com

Received 7 July 2017; Revised 13 August 2017; Accepted 28 August 2017; Published 5 December 2017

Academic Editor: Daniela Paolotti

Multimedia applications in smart electronic commerce (e-commerce), such as online trading and Internet marketing, always face security in storage and transmission of digital images and videos. This study addresses the problem of security in e-commerce and proposes a unified framework to analyze the security data. First, to allocate the definite security resources optimally, we build our e-commerce monitoring model as an undirected network, where a monitored node is a vertex of the graph and a connection between vertices is an undirected edge. Moreover, we aim to find a minimal cover for the monitoring network as the optimal solution of resource allocation, which is defined as the network monitoring minimization problem (NMM). This problem is proved to be NP-hard. Second, by analyzing the latent threats, we design a novel and trusted monitoring system that can integrate incident monitoring, data analysis, risk assessment, and security warnings. This system does not touch users' privacy data. Third, we propose a sequential model-based risk assessment method, which can predict the risk according to the text semantics. Our experimental results on web scale data demonstrate that our system is flexible enough when monitoring, which also verify the effectiveness and efficiency of our system.

1. Introduction

The rapid development of electronic commerce (e-commerce) has turned it into an indispensable aspect of the Chinese lifestyle and has thoroughly penetrated people's lives and work. This shows that the distribution of digitized images and videos is more and more popular. For example, online shops in Taobao, JD, and Amazon begin to use images and videos, not only text, to introduce the features and usage of commodities. Multimedia data, such as text, images, and videos, can be easily switched between the Internet and users. Mass of multimedia data is produced daily and the privacy becomes an increasingly major concern for users. Moreover, multimedia data is always sent through wireless networks owing to the convenience of mobile devices. However, business objects are more and more easily attacked than before since data increases exponentially and such the potential threat does not get enough attentions. It is important to ensure secure and reliable multimedia data.

Although multiple Internet security technologies have been applied to e-commerce, strengthening e-commerce security is still challenging [1]. For example, service leaks can be detected by hackers, and existing Internet security algorithms perform poorly due to the increasing number of newly created e-commerce services. Similar systems, such as SOA-based systems [2], have been built to protect e-commerce. In 2014, Luhach et al. [3] proposed a logical security framework for e-commerce systems. Accordingly, they attempted to apply an easy, flexible, and recyclable application for e-commerce security. Massa and Valverde designed a fraud detection system based on the detection systems of anomaly intrusion for e-commerce applications [4]. These researchers used different anomaly detection techniques to predict computer intrusion attacks in e-commerce web applications. To the best of our knowledge, existing e-commerce security systems only consider one or two aspects of e-commerce protection despite the insights generated by prior studies. Moreover, these security systems may have difficulty in obtaining suitable results in real-time applications. The current study attempts to overcome these research limitations by redesigning and implementing a new

system that integrates incident monitoring, data analysis, risk assessment, and security warnings.

Since customer and company data are private and sensible, privacy problems caused by data leaks have become primary concerns. Thus, customers and companies alike can hardly trust a third party. Furthermore, data are produced massively and rapidly. Hence, scheduling resources dynamically by adding and deleting devices may be considerably difficult. Accordingly, updating strategies may not easily adapt to changing network attacks. The rapid rise and development of e-commerce has brought new challenges. Since users would not like to disclose their own unique and private information including multimedia contents to others, especially strangers, privacy preservation is of paramount importance, where the contents include identities, locations, preferences, and social relationships. However, sometimes, users are not even aware of the disclose of their privacy and do not intentionally protect themselves.

In this paper, we apply cloud computing to our system to enable it to be configurable and resource-shared. In addition, this system considers data, services, and devices as normalized resources that could be accessed by the unified interfaces. The sensible input data will also be encrypted and not be stored. We treat input data as streams and extract-only features. We make the following contributions.

First, we formulate the problem of monitoring minimization in e-commerce (NMM-problem), which aims to use the fewest resources to monitor the whole e-commerce network. The e-commerce network is modeled as an undirected graph, where a monitored e-commerce object is a vertex of the graph and a link between two objects is an edge. By reducing from the set cover problem, we prove that NMM-problem is NP-hard. We also give a greedy algorithm to allocate the optimal resources.

Second, we build a novel and unified system to handle emerging e-commerce security risks. This system is configurable and resource-shared because it integrates incident monitoring, data analysis, risk assessment, and security warnings. Moreover, this system considers data, services, and devices as normalized resources that could be accessed by the uniform interfaces.

Third, we propose a sequential model-based risk assessment method, which can predict the risk according to the text semantics. With the help of the built text semantic knowledge base, we can know the emotional tendency of words. Given a positive minimal integer, we can mine the negative patterns and determine whether the negative pattern is frequent, which can also be treated as a latent risk.

Fourth, we implement several data analysis algorithms for different applications and construct knowledge bases for different types of data. As a case study, we provide a malicious analysis model to illustrate the working mechanism. This case study also verifies the effectiveness and efficiency of our system.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 introduces our problem formulation. Section 4 provides threat analysis and our system structure for e-commerce security. Section 5 provides several efficient mechanisms. Section 6 provides an experimental study. Section 7 presents the conclusions.

2. Related Work

This work is related to e-commerce security, multimedia, and cloud computing security. In this section, we briefly review the most related work about multimedia, e-commerce, cloud computing, and security.

2.1. E-Commerce Security. E-commerce security is a complex problem that is not limited to network security. This type of security also involves the following specific aspects.

(1) Management Issues. Most e-commerce sites have yet to establish a unified management and evaluation standards. Most of the security risk management capabilities of these sites are weak. The lack of the ability to resist hacking sites commonly leads to server paralysis, thereby affecting the credibility of the site.

(2) Technical Problems. Although many e-commerce security products exist, the real product certification is limited. The main reasons for this condition are as follows. First, the global network security has not formed a complete system. Second, the intensity of security technology is weak.

(3) Environmental Issues. E-commerce can also be affected by the social environment, particularly the impact of legal systems. Thus, we must improve the relevant laws to ensure the development of e-commerce construction.

The successful functioning of e-commerce security depends on a complex interrelationship between several applications, namely, development platforms, database management systems, systems software, and network infrastructure [5]. A general model has been introduced by Schmid, who identified three phases for most e-commerce processes. (1) During the information phase, the parties attempt to find partners, compare them, clarify their trade relation, and specify the products to be exchanged. These actions are not legally binding. (2) In the contracting phase, the parties decide on their partners based on their decision criteria and then work out and sign a contract about their trade relations. (3) In the delivery phase, payment and delivery are completed and a new transaction is eventually prepared [6].

Several studies that discuss the security aspects in e-commerce have been presented.

E-commerce software packages should work with secure electronic transfer, secure socket layer (SSL), public key infrastructure, and secure E-commerce protocol [7] technologies for the encryption of data transmissions. E-commerce operates on the Internet or intranet. The main e-commerce transaction models are B2B and B2C. Public key infrastructure (PKI) offers the ideal approach to identify or authenticate the identity of the other party on the Internet. Several security services based on PKI can be implemented to secure e-commerce transactions.

Digital signatures provide the requirement for authentication and integrity. A sending message is run through a hash

function and a new value (i.e., message digest) is eventually generated. The message digest and plain text are encrypted using the recipient's public key and sent to the recipient. The recipient decrypts the message using his private key and passes the message through the supplied hash algorithm. A digital certificate is also used for security purposes. An algorithm provides the capability to generate and verify signatures. Signature generation uses a private key to generate a digital signature [8].

SSL was developed by Netscape to provide secure communication between web servers and clients. The information is broken into packets, numbered sequentially, and attached with an error control. Individual packets are sent through different routes [6]. SSL is extensively applied on the Internet, particularly for interactions that involve exchanging confidential information, such as credit card numbers [9]. SSL protects the communication between a client and a server, as well as providing authentication to both parties for secure communication [10]. SSL provides point-to-point security [11]. That is, a message is encrypted only during transmission over the network, and other security mechanisms are required to handle the message security in an application or disk. SSL is used to securely exchange secret keys between communication parties.

PGP was developed by Phil Zimermann and provides secure communication in an unsecured electronic environment. PGP provides authentication, confidentiality, compression, and segmentation services for email security. Moreover, PGP provides confidentiality and authentication services that can be used for electronic mail and file storage applications [8]. PGP is extensively used for email security.

The effectiveness of enforced countermeasures is subject to several vulnerabilities. Firewall systems, which are the most preferred form of security, also have several design and configuration challenges. Intrusion prevention systems suffer from vulnerabilities, such as the underestimation of security capabilities of prevention and detection, focus on performance instead of security, and nondefined management policies that include design and implementation [12]. A study conducted on the effectiveness of evasion techniques against intrusion prevention systems revealed that IPS systems are vulnerable to evasion techniques and combinations [10]. The majority of the detection rates were over 95%; however, old evasion techniques could penetrate even the most sophisticated of systems. The aforementioned study concludes that the default configurations were insufficiently strict to block the attacks masked with evasions. A study conducted by Imperva on the effectiveness of antivirus software collected 40 of these products and tested them against 82 malware types. The conclusion of this study was that antivirus products are effective at malware detection that redistributes rapidly in large samples; however, new strains still leave a window for attacks. The window of attack creates a blind spot because security teams are commonly unaware of its existence. Accordingly, a proposed security model, which includes monitoring access to servers, databases, and files, has been recommended to handle such blind spot [13]. A strong password policy can also protect against repeated attacks on a user's account, constant speculative attacks on all user accounts, and specific attacks with guesses based on the user details. However, a study conducted to determine the effectiveness of passwords concludes that users are misguided to employ strong passwords because they do not offer any defense against password stealing attacks. Therefore, an effective lockout system (i.e., three unsuccessful logins) would be an ideal control measure in a small institution [14].

The current study constructs a new model and framework to overcome the limitations of prior methods for e-commerce protection. This model will be discussed in the next section.

2.2. Multimedia and Cloud Computing Security. Generally, the multimedia security solution is based on using hiding techniques, such as encryption [9], key management [15], authentication [16], and watermarking [10], which were applied on multimedia data protection. However, traditional security measures did not work well when applied on cloud directly since it was difficult to deal with mass of data. Researchers are focusing on the more efficient and effective approaches to handle the multimedia security problem on cloud platforms, due to the widely used cloud computing techniques.

Cloud computing can be expressed as a combination of software-as-a-service, which refers to a service delivery model [17]. Since it provides flexible support, multimedia data is always stored on the cloud platform. Zhu et al. [18] described the overview of multimedia cloud computing and it gave addressed multimedia cloud computing from multimedia-aware cloud. How to protect multimedia data stored in cloud environment is still an open problem. A comprehensive review on recent multimedia security research activities in association with cloud computing is provided by [19].

Security systems in cloud computing were built. Jain and Kaur [20] presented a data security system in cloud computing using DES algorithm. This cipher block chaining system was secure for clients and server. The security architecture of the system was designed by using DES cipher block chaining, which eliminated the fraud that occurred daily with stolen data. Ren et al. [21] proposed an effective and flexible scheme, which aimed to enhance multimedia security in cloud computing environment using crossbreed algorithms. A user can access cloud services as a utility service and begin to use them almost instantly. Yadav et al. [22] proposed a privacy-preserving and copy-deterrence content-based image retrieval (CBIR) scheme in cloud computing, which supported CBIR over encrypted images without leaking the sensitive information to the cloud server. They extracted feature vectors to represent the corresponding images and the prefilter tables were constructed by locality-sensitive hashing to increase search efficiency. The security analysis and the experimental results demonstrated the security and efficiency of the proposed scheme. With the widely deployed mobile devices such as smartphones, a significant question about how can mobile users trust the media services provided by the media cloud service providers occurs. Wang et al. [19] proposed using both secure sharing and watermarking schemes to protect user's data in the media cloud. The secure sharing scheme allowed users to upload multiple data

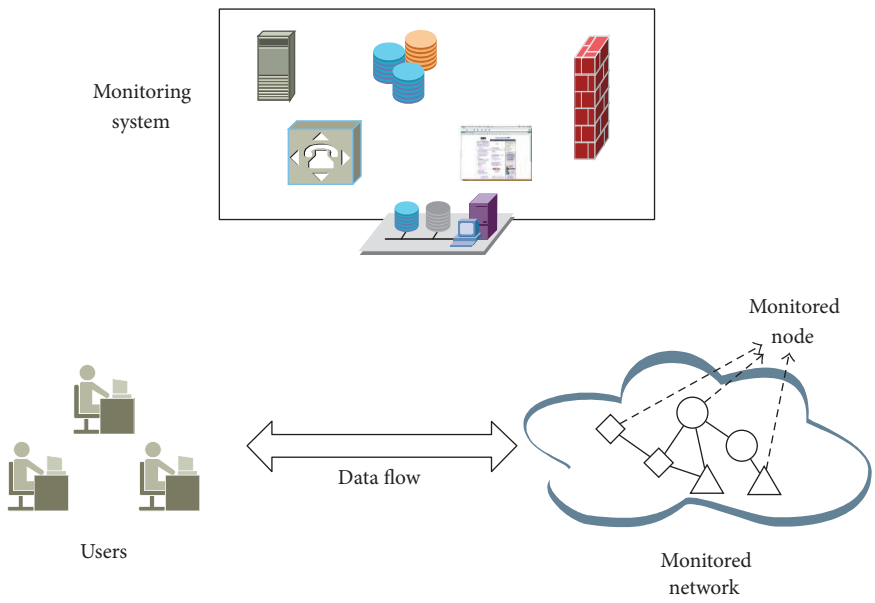

Figure 1: Cloud monitoring service architecture.

pieces to different clouds, making it impossible to derive the whole information from any one cloud. In addition, the proposed scalable watermarking algorithm can be used for authentications between personal mobile users and the media cloud. The proposed approach not only achieved good security performance, but also can enhance media quality and reduce transmission overhead.

However, the above schemes or systems only focus on how to protect multimedia data. They did not touch the resource assignment problem and had lack of monitoring. Thus, our method is quite different from the mentioned existing ones.

3. Problem Statement

In this section, we first describe the monitoring scenario practically and then illustrate our model preliminaries.

3.1. E-Commerce Monitoring Scenario. The representative network architecture for cloud monitoring service architecture (CMS) is illustrated in Figure 1. The entities of the architecture can be described as follows.

(1) Monitored Network (MN). The monitored network is formulated as an undirected graph, where a monitored object is a vertex of the network and an edge is a connection between two vertices. In practical, data is transmitted through the connected links, that is, edges. Intuitively, monitoring one node should monitor not only the object itself, but also the connected links to the object. Thus, a monitored network can be divided into several parts, where each part includes a monitored object and links being incident to the object. A monitored object can be a website, an information system, and so on.

(2) User. A user is an entity that visits the MN and has data to be stored in the MN and can be either enterprise or individual customers.

(3) Monitoring System (MS). A monitoring system is a unified scheme based on cloud computing. It consists of three parts: management domain, monitoring domain, and data domain. Management domain manages MS and provides friendly user interfaces for user visiting. With such interfaces, MS managers can obtain real-time dynamics. Monitoring domain provides monitoring functions for MN and assigns resources dynamically. Data domain realizes data processing. The data, including malicious codes and logs, is stored on cloud in relational or NoSQL databases. The data would be extracted as features while storing.

3.2. Preliminaries

3.2.1. Monitored Network. A monitored network is an undirected graph, denoted by $G = (V, E)$, where V is a set of vertices and $E = \{(u, v) \mid u, v \in V\} \subseteq V \times V$ is a set of undirected edges. We denote the set of neighbors of a vertex v by $N(v)$ and the degree of v by $d(v) = |N(v)|$. Corresponding to the CMS model, v is denoted as a monitored object and (u, v) is denoted by the connection between u and v. We denote a monitored entity (ME) as T_v, where T_v contains a root vertex v and any edge of G belongs to one and only one T_v. Φ is a set of monitored entities, where $\Phi = \{T_v \mid v \in V\}$.

We now give an example to explain the above notations.

Example 1. Figure 2 show a graph G and two different set of monitored entities, where $\Phi = \{T_2, T_3, T_4\}$ and $\Phi' = \{T_2', T_3', T_4'\}$. G contains 6 vertices and 8 edges. T_4 contains a root vertex v_4 and 4 edges which is incident to v_4, where such edges are (v_4, v_2), (v_4, v_3), (v_4, v_5), and (v_4, v_6).

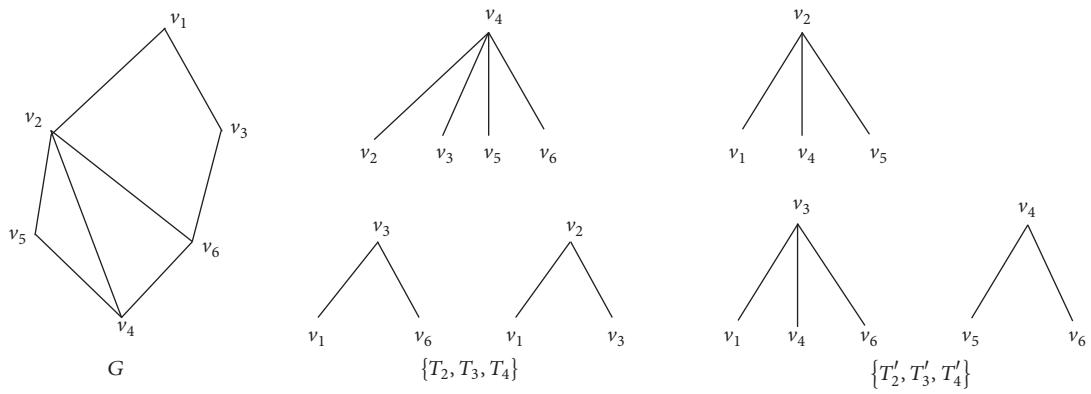

Figure 2: Monitored entity decomposition.

3.2.2. Sequential Model. Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of items. An itemset X is a subset of I: that is, $X \subseteq I$. A sequence s is an ordered list of itemsets, donated by $s = \langle s_1 s_2 \cdots s_l \rangle$, where s_j is an itemset. A sequence with length l is called an l-sequence. A sequence $\alpha = \langle a_1 a_2 \cdots a_n \rangle$ is called a subsequence of another sequence $\beta = \langle b_1 b_2 \cdots b_m \rangle$ and β a supersequence of α, donated as $\alpha \sqsubseteq \beta$, if there exist integers $1 \leq j_1 < j_2 < \cdots < j_n \leq m$ such that $a_1 \subseteq b_{j_1}, a_2 \subseteq b_{j_2}, \ldots, a_n \subseteq b_{j_n}$.

A sequence database S is a set of tuples $\langle \text{sid}, s \rangle$, where sid is a sequence_id and s a sequence. A tuple $\langle \text{sid}, s \rangle$ is said to contain a sequence s, if α is a subsequence of s. The support of a sequence α in a sequence database S is the number of tuples in the database containing α: that is, $\sup_S(\alpha) = |\{\langle \text{sid}, s \rangle \mid (\langle \text{sid}, s \rangle \in S) \wedge (\alpha \sqsubseteq s)\}|$.

We then give an example to illustrate the above concepts, as shown in Table 1.

Example 2. We denote the marked sequence dataset S and the minimum support min_sup = 2. The dataset contains 4 sequences and 8 items. The length-1 patterns are $\langle A \rangle$, $\langle B \rangle$, $\langle D \rangle$, and $\langle F \rangle$ and the length-2 patterns are $\langle AB \rangle$, $\langle AF \rangle$, $\langle (B)(A) \rangle$, $\langle BF \rangle$, $\langle (D)(A) \rangle$, $\langle (D)(B) \rangle$, $\langle (D)(F) \rangle$, and $\langle (F)(A) \rangle$.

Table 1: A toy database.

SID	Sequence
(1)	$\langle (CD)(ABC)(ABF)(ACDF) \rangle$
(2)	$\langle (ABF)E \rangle$
(3)	$\langle ABF \rangle$
(4)	$\langle (DHG)(BF)(AGH) \rangle$

4. Threat Analysis and Efficient Mechanisms

In this section, we first analyze the existing threats of e-commerce and then design several efficient mechanisms. We also conclude the system design goals thereafter.

4.1. Threat Analysis. E-commerce network security inevitably hides dangers, such as network user information theft, information tampering, fake information, malicious computer viruses, malicious damage, and system security. E-commerce faces the following threats.

(1) E-commerce websites may be counterfeit, thereby possibly damaging consumer rights: e-commerce provides an excellent consumer shopping process. Although this trading approach facilitates the people's economic lives, many people still lack general knowledge of security (e.g., they cannot identify which shopping websites are real). Clicking on false websites will undoubtedly mislead customers.

(2) E-commerce lacks standardized management: people are rarely concerned about e-commerce security: trading processing involves different nonstandard protocols because no unified national standard exists. Furthermore, China lacks any network-related legislation; thus, people can do whatever they want do, thereby resulting in serious damage to the public Internet environment.

(3) Information storage security is relatively weak: two main forms of storage threats exist in e-commerce, namely, "unauthorized users" and "view information." When companies connect to the Internet, problems in the processes of e-commerce operations will significantly affect enterprises. External threats to enterprises include external attacks, unauthorized access, and information theft. By contrast, internal threats involve unauthorized access to the information.

(4) Data cardinality increases: numerous customers and enterprises are joining in e-commerce activities; thus, data are increasing exponentially. This scenario may lead to losses in anomaly detection because the existing computing resources cannot handle the dramatic increase in data. A directed approach is to add computing resources. However, existing systems that cannot support adding or deleting dynamically are another emerging problem.

4.2. Cloud Computing-Based E-Commerce System. In this section, we propose to design our e-commerce security system based on the threat analysis. Firstly, to ensure the security and dependability for data under the aforementioned model, we aim to achieve the following goals:

(1) Correctness: to ensure users that their data are indeed stored appropriately and kept intact all the time in the cloud

(2) Dynamic resource support: to maintain the same level of monitoring even if users modify, delete, or append

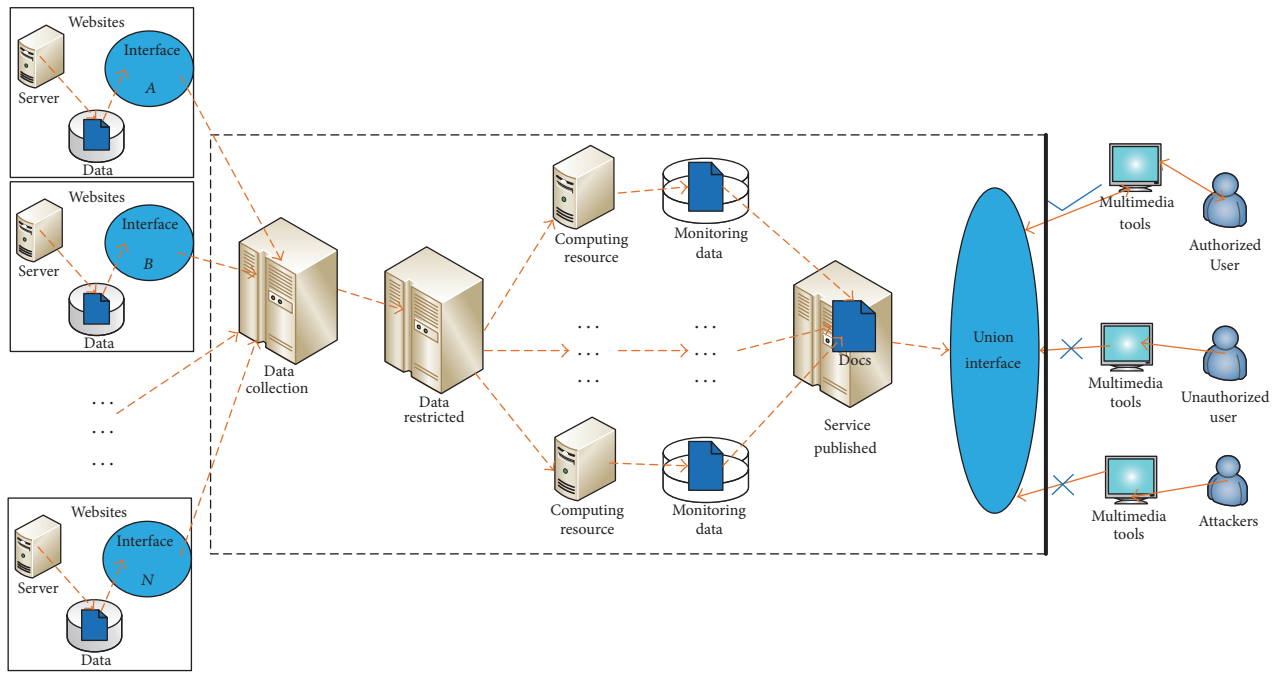

Figure 3: Work flow of the e-commerce system.

their data in the cloud and to maintain the scalability when increasing or decreasing resources

(3) Online analysis: to analyze incidents online in real time and publish early warning in time

(4) Handle big data: to deal with mass of streaming data

We now illustrate the details of our e-commerce security system. We employ and implement our applications on an open-source cloud computing model as the basic platform to support adding and deleting sources dynamically. Only authorized users can visit this system through unified interfaces designed for different security strategies. We do not determine the user data and only provide interesting analysis results to substantially protect data privacy. Figure 3 shows the work flow.

The e-commerce system works as follows. We have initially built knowledge bases for multiple tasks, such as anomaly detection, frequent data mining, and dark chain identification. This process can be downloaded offline. We deploy our system on the Internet and connect to websites, thereby enabling us to monitor security incidents in real time. First, we obtain data streams from monitoring data by using the selected interfaces. We only record data information and not store them in our database in this process. Second, we reconstruct streaming data in formats that are suitable for our input. Third, we provide dynamic computing resources for the data analysis. Multiple algorithms may analyze the data for different statistical characters in such case. Thus, we can obtain several analysis results. Lastly, we output the results to the authorized users through unified interfaces. We isolate users so they can only know their own analysis results and suggestions. The data are encrypted throughout the entire process to avoid data leaks.

The next section provides efficient mechanisms for the system.

5. Efficient Mechanisms

In this section, we propose several efficient mechanisms, including network monitoring minimization and sequential feature based risk assessment.

5.1. Network Monitoring Minimization. It is clear that MS has limited resources and we should assign them reasonably. As we have formulated above, we would like to ensure the whole network G is monitored by MS. Thus, we can define a problem as follows.

Problem 3. Given an MS, a monitored network G, and a set of monitored entities $\Phi = \{T_1, \ldots, T_n\}$, we call Φ a monitoring cover of G. The problem is to find minimal monitoring cover of G.

The following theorem shows the hardness of finding the minimal monitoring cover.

Theorem 4. *The minimum network cover is polynomial equivalent to the minimum set cover problem.*

Proof. We prove the theorem by reducing from the NP-hard set cover problem [20]. Give a ground set and a collection of sets whose union equals the ground set, the set cover problem is to a set covering that uses the fewest sets.

Given an instance of the set cover problem, let S be a minimum network cover of G. Let V be the set of the root

```
(1)  Φ = ∅; S = ∅;
(2)  while G has more edges do
(3)      if S = ∅ then
(4)          Pick an edge (u, v) such that d(u) + d(v) is the largest;
(5)      else
(6)          Pick an edge (u, v) such that v ∈ S and d(u) + d(v) is the largest;
(7)      T_v ← the ME rooted at v;
(8)      Add T_v into Φ;
(9)      S ← S ∪ neighbor(v);
(10)     Remove edges in T_v from G;
(11)     if d(u) > 0 then
(12)         T_u ← the ME rooted at u;
(13)         Add T_u into Φ;
(14)         Remove edges in T_u from G;
(15)         S ← S ∪ neighbor(u);
(16)     Remove u and v and all nodes with degree 0 from G;
(17) Return Φ
```

Algorithm 1: Monitoring minimization.

nodes of the monitored entities (MEs). It is clear that $|V| = |S|$ and V is a vertex cover of G; that is, any edge of G is incident to at least one node in V. Next, we show that V is a minimal vertex cover. Assume there exists another vertex cover V' such that $|V'| < |V|$. We construct MEs S' from V' as follows. For each vertex $v \in V'$, we form an ME which consists of v and all edges incident to v. Then, we randomly delete edges in the MEs until each edge belongs to one ME. Let S' be the set of MEs that have at least one edge. Clearly, $|S'| \leq |V'| < V = |S|$, which means S is not a minimum network cover. Obviously, S' does not exist. It is a contradiction. Thus, the minimum network cover is polynomial equivalent to the minimum set cover problem. □

In the proof of Theorem 4, we showed that we can construct monitoring cover from a set cover in polynomial steps. There exists a 2-approximate algorithm [4] for the set cover problem. In our approach, we propose an optimal processing order, which ensures that the MEs generated by the algorithm can be a more favorable MEs processing order. We propose two rules to guide edge selection.

(1) Select edges that already connect to previously selected edges.

(2) Select edges incident to vertices with high selectivity. The selectivity of ME is determined by $d(v)$.

Algorithm 1 outlines the method that combines monitoring covering and selection. We first select an edge with the largest $d(u) + d(v)$. For vertices u and v, we build T_u and T_v and remove the edges in T_u and T_v to ensure that one edge can only belong to one ME. This process ends till all edges of G are removed. We then give an example to explain the process.

Example 5. We first calculate the degrees for vertices and the edge with the largest $d(u)+d(v)$ is (v_2, v_4). Then, we construct the MEs rooted at v_2 and v_4, respectively. The edges (v_1, v_3) and (v_3, v_6) are left and we construct the ME rooted at v_3. All of the edges are removed now and (T_2, T_3, T_4) is returned.

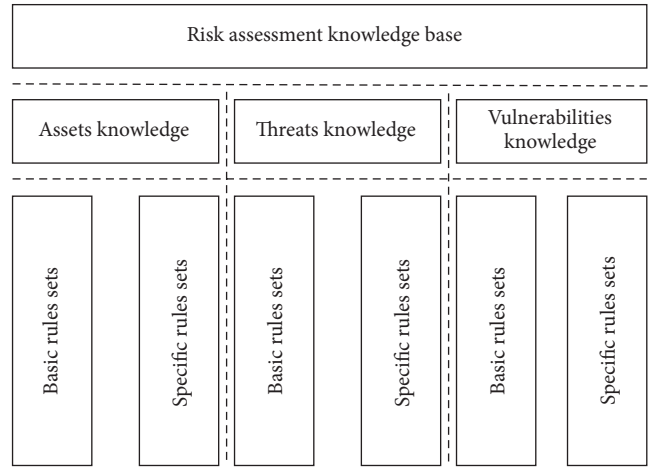

Figure 4: The architecture of risk assessment knowledge base.

Regarding the time complexity of the algorithm, we note that computing degrees and sorting edges have $O(n \log n)$ cost, where n is the number of vertices in the network. In each round, at least two vertices are removed from G, so the iteration needs $O(n)$ steps. Thus, the time complexity of Algorithm 1 is $O(n^2 \log n)$.

5.2. Sequential Model-Based Risk Assessment. In this section, we propose how to use sequential features to recognize risk. Notice that our risk assessment is significantly different from information security risk since our method considers the sentiment of the sentences. Our method can obtain the sentiment of reviews so that customers can determine which shop is better. We now show the basic model of information security risk, which uses knowledge of assets, threats, and vulnerabilities. The architecture of risk assessment knowledge base is shown as Figure 4.

(1) Initialize the sequence set $\Theta \leftarrow \varnothing$;
(2) for each text t do
(3) Extract affective features, statistical features and semantic dependency features for t;
(4) Remove neutral features of t;
(5) Construct s based on t, where each item of s is a word of t;
(6) $\Theta \leftarrow \Theta \cup s$;
(7) end for
(8) Apply sequential pattern mining techniques on Θ to obtain Q;
(9) Return Q.

Algorithm 2: Mine word patterns.

Risk assessment knowledge is divided into assets knowledge, threats knowledge, and vulnerabilities knowledge. Assets knowledge defines the confidentiality, availability, and integrity of assets. Threats knowledge defines the level of threats by the frequencies of faults. Generally, the higher frequencies produce the bigger threats. Vulnerabilities knowledge is actually defined as the impact on assets, in which vulnerabilities are utilized by threats. Rules, including basic rules and specific rules, are defined by users to analyze latent threats, respectively.

As we all know, the mentioned knowledge based risk assessment needs a lot of user defined rules, which is always hard to be fixed. Different from it, we try to apply semantic analysis on risk knowledge. The reason is that traditional risk assessment methods cannot recognize threats from short text since they can just work on structural words or phrases. How to design a rational evaluation system is still an open problem.

Given a set of reviews, we are interested in extracting the following features.

(1) Affective Features. The so-called affective factors are negative words, degree words, and related words. The appearance of such words often affects the emotional changes or emotional strength in a sentence. For example, "although we cannot succeed immediately, but if you work hard, you will eventually complete" contains the emotional impact factors "not," "but," and so on. These emotional words can determine the user's emotional trends.

(2) Statistics of Words. According to the statistics and research, we found that a single part of speech or several consecutive parts of the combination contain subjective information and objective information. The N-POS (part of speech) model is a corpus-based statistical natural language model. When N is 3, the three consecutive words are combined into a pattern. This paper treats the sequences of three consecutive parts as emotional characteristics.

The three-POS features of the sentence are: nouns-adjectives-pronouns, adjectives-pronouns-lattice, pronouns-lattice-verb, lattice-verb-name markup, verb-name markup-lattice, name markup-lattice-noun, lattice-noun-verb, and noun-verb–symbol.

(3) Semantic Dependency Features. We use the semantic relation between words to reveal the syntactic structure of sentences. The semantic dependency is the main element of the syntactic structure of the grammatical grammar. It refers to the binary relation of the word pair in the sentence. One is called the central word and the other is called the subsidiary word. The dependency expresses a semantic dependency between the central word and the subsidiary word. By exploring the interdependence between the central words in the sentence and the subsidiary words of the central word, we can obtain effective emotional characteristics. If there exists a dependency relationship between two words, we treat the two words as an entity, which represents an itemset in a sequence.

Given a set of text, after we finish extracting the above features, we actually obtain a set of sequences, where a word is an item or dependent words set is an itemset. Fixing the minimal support min_sup, we can mine frequent sequential patterns and determine if there exist negative emotional patterns. By comparing the negative emotional patterns and positive emotional patterns, we can obtain the text sentiment. Algorithm 2 gives the details of this method.

6. Experiments

In this section, we analyze the effectiveness and efficiency of the proposed system and the optimization techniques. The algorithms were implemented in Java and compiled with JDK 8. All experiments were conducted on a cluster with 8 machines, where each machine has 32 GB DDR3 RAM and one 3.10 Ghz Intel Xeon E3-1220 v2 CPU with 4 cores and 4 threads. The operating system is Linux Ubuntu 14.04.

6.1. Datasets and Evaluation Metrics

6.1.1. Datasets. We use 5 publicly available real-worlds networks, which is published in SNAP [23], to evaluate the monitoring minimization. The network statistics are shown in Table 2. Vertices # and edges # are denoted as the number of vertices and edges, respectively.

We also collected malicious Uniform Resource Locator (URL) data from the Internet. Malicious websites are a cornerstone of Internet criminal activities [5]. Users must implicitly evaluate the associated risks each time they decide whether to click on an unfamiliar URL [17]. We select the malicious data for the following reasons. First, malicious data have many samples in different e-commerce areas, including

Table 2: Network statistics.

Name	Vertices #	Edges #	Description
Citation	12,008	118,521	Collaboration network of Arxiv High Energy Physics
Amazon	334,863	925,872	Amazon product network
Email	36,692	183,831	Email communication network from Enron
Youtube	1,134,890	2,987,624	Youtube online social network
Gowalla	196,591	950,327	Gowalla location based online social network

Table 3: Malicious data statistics.

File format	Total number	Total size (in megabytes)
.axd	84	3.7
.bin	20	1.5
.class	205	2.7
.css	5,983	102.8
.gz	2,080	3210.0
.html	180,342	7286.3
.js	20,869	516.8
.obj	6,771	608.0
.php	7,711	112.8
.swf	3,580	1996.9
.tff	6	0.2
.txt	22	1.6
Others	85	5.3

text, images, videos, and other multimedia data. Hence, we can learn significant knowledge about malicious features and thereafter construct a knowledge base. Second, abundant knowledge enables us to easily deal with new input data. In the data analysis, substantial data is considerably beneficial for learning models. Thus, we can increase the accuracy and recall the ratios of classification and recommendation tasks with sufficient probability. Validating is also convenient because our data come from popular malicious websites, which have been marked as suitable or not. Third, malicious data have clear categories; hence, they can be easily cleaned and analyzed.

We designed a crawler and performed statistical analysis on the features. Thus, we also obtained the feature vectors when we completed the analysis. The total number of documents is 259,137, and their total size is 13.6 GB. Table 3 lists the descriptive statistics of malicious data.

6.1.2. Evaluation Matrices. We evaluate the performance of our proposed methods in the following aspects.

(1) "Time Cost": it measures the average total runtime of each method.

(2) "Precision" and "Recall": "Precision" is the number of correct results divided by the number of all returned results. "Recall" is the number of correct results divided by the number of results that should have been returned. Notice that, this metric is only used on the malicious URL detection.

(3) "#ME": it measures the number of MEs, which is the solution of the NMM-problem.

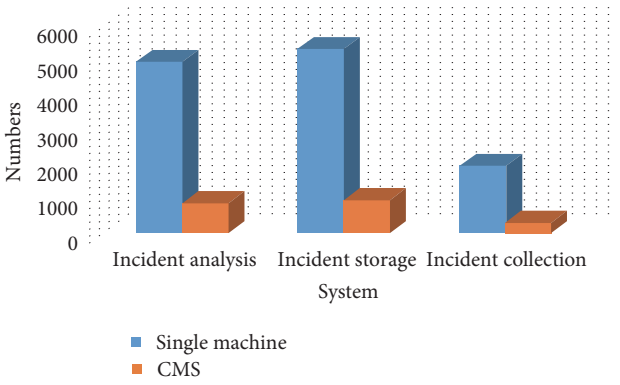

Figure 5: The comparison of the abilities on different systems.

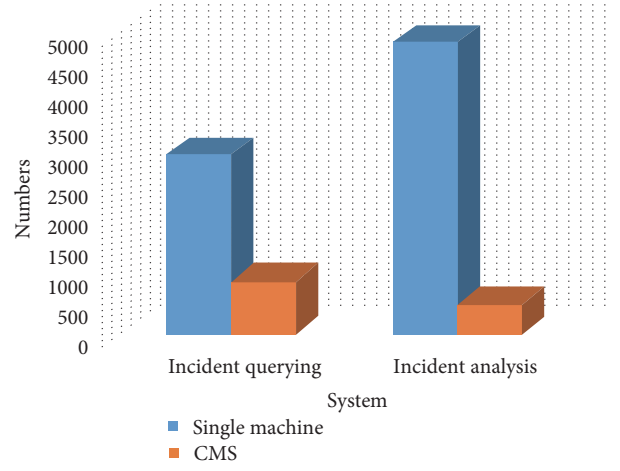

Figure 6: The comparison of the time cost on different systems.

6.2. Results on System Performance. Figures 5 and 6 show the performance of the different systems. Obviously, our cloud-based system achieves better performance in incident analysis, incident storage, and incident collection. In summary, our system can monitor 95 important websites and simultaneously collect data from a minimum of 20 data sources. Our system can gather at least 2,000 data records and handle at least 5,000 data records every second. We decrease the running time and improve efficiency by 7.9 times by deploying our cloud computing platform. Although this algorithm is only a miniature version of the proposed system, it still demonstrates the effectiveness and efficiency of this system.

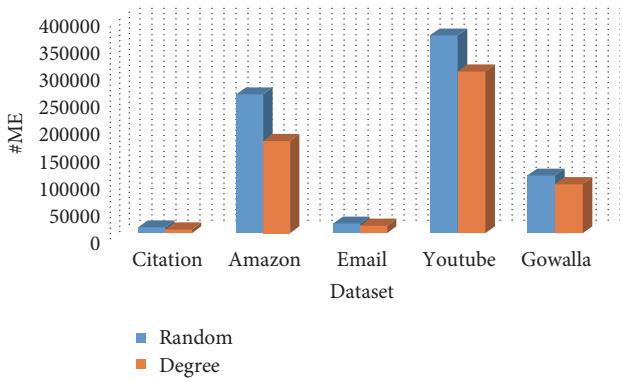

Figure 7: #ME of the different NMM methods.

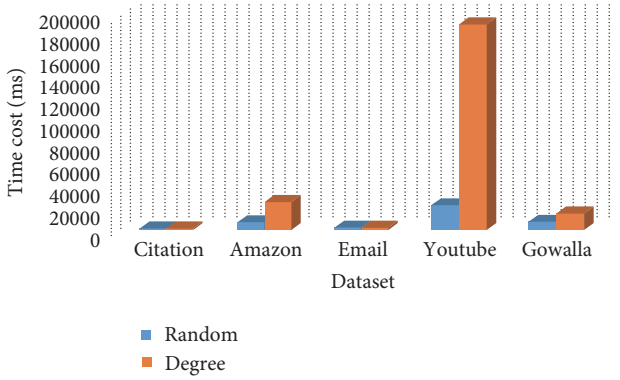

Figure 8: The time cost of the different NMM methods.

6.3. Results on NMM-Problem. Figures 7 and 8 show the performance of the network monitoring minimization methods. The baseline method is called "Random," which selects a root vertex randomly to construct an ME. Our proposed method Algorithm 1 is called "Degree." From Figure 7, the number of MEs decreases and shows that Algorithm 1 can obtain fewer MEs, which leads to fewer monitoring resources. From Figure 8, the time cost of "Degree" is larger than that of "Random" since "Degree" sorts the vertices firstly. The dataset Youtube has the largest #ME and time cost since this dataset is the largest one.

6.4. Results on Malicious URL Detection. We select several features to form the data vectors. Thereafter, these features can represent malicious code snippets. Given that constant page tags divide the web pages, removing irrelevant information and leaving truly beneficial text are easy. The context words are also nearly irrelevant. Thus, we do not consider the semantics and only obtain the features from statistics. Table 4 describes the selected features.

Figure 9 shows the precision and recall with the number of samples increasing. We can know that the precision and recall is larger than 80%, which are better results. This shows that our system is effective on the task of malicious URL detection.

6.5. Case Study. In this section, we report a case study about malicious URL detection and risk prediction, which come from the monitoring data.

Table 4: Selected features.

No.	Description
(1)	Number of "eval" strings
(2)	Number of unescape strings
(3)	Number of ".exe" or ".EXE" strings
(4)	Frequencies of special characters
(5)	Number of unicode strings
(6)	Frequency of the blank space character
(7)	Frequencies of upper case letters
(8)	Frequencies of numerical characters
(9)	Number of long strings (a string has at least 150 characters)
(10)	Average string length
(11)	Frequency of "%*u*"
(12)	Frequency of readable strings (a string contains over 70% letters)

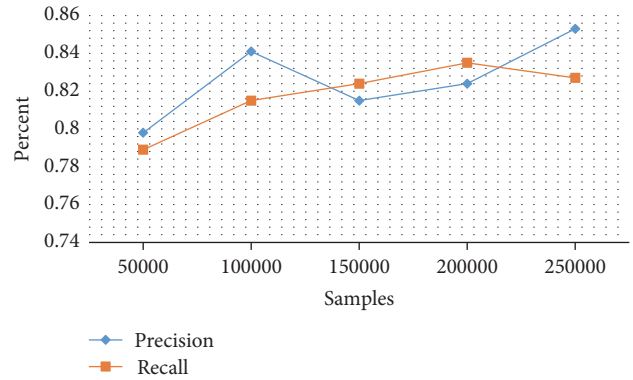

Figure 9: The precision and recall on malicious URL detection.

We use a fake shopping website, "http://www.taobao1.com," to explain the case study about malicious URL detection. After crawling the webpage, we identify a code snippet embedded an executable file "XX.exe." When a user clicks on the hyperlink, this executable file will pop up and mislead the user to input his/her username and password. Thereafter, the user's private information will be leaked.

We then report a risk prediction example, which is ever detected by our system. When we dealt with a security report from WWW, our system found that this report had heavily negative sentiment. We obtained that this report was about a leak of Apache Tomcat. Then we predicted a high risk for assets which had Apache Tomcat.

7. Conclusion

This study designs a unified framework to analyze the security data. Given the proposed framework, we build a novel and trusted system that integrates incident monitoring, data analysis, risk assessment, and security warnings. We run the malicious URL detection oriented to e-commerce security to verify the effectiveness and efficiency of our system. We

present valuable and beneficial measures for companies to avoid losses caused by Internet attacks by analyzing the results of the specific task. However, our research has several limitations. First, we do not focus on the internal security of e-commerce systems, which is also a security threat that requires complete rights management. Second, we should also improve the selection of the proper data analysis because different algorithms may return similar results.

The limitations of our research indicate our future research direction. First, an improved approach to the trade-off between accuracy and efficiency should be formulated. Second, internal security threats should be addressed. Third, management strategies should be properly established.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Nature Science Foundation of China (no. 61672553) and Project of Humanities and Social Sciences (the Ministry of Education in China) (no. 16YJCZH076).

References

[1] K. Ahmad and M. S. Alam, "E-commerce Security through Elliptic Curve Cryptography," *Procedia Computer Science*, vol. 78, pp. 867–873, 2016.

[2] A. K. Luhach, S. K. Dwivedi, and C. K. Jha, "Applying SOA to an E-commerce system and designing a logical security framework for small and medium sized E-commerce based on SOA," in *Proceedings of the 5th IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2014*, pp. 1–6, December 2014.

[3] A. K. Luhach, S. K. Dwivedi, and C. K. Jha, "Designing A Logical Security Framework for E-Commerce System Based on SOA," *International Journal on Soft Computing*, vol. 5, no. 2, pp. 1–10, 2014.

[4] D. Massa and R. Valverde, "A fraud detection system based on anomaly intrusion detection systems for E-commerce applications," *Computer and Information Science*, vol. 7, no. 2, article 117, 2014.

[5] S. Kesh, S. Ramanujan, and S. Nerur, "A framework for analyzing e-commerce security," *Information Management & Computer Security*, vol. 10, no. 4, pp. 149–158, 2002.

[6] M. S. Mokbel and L. Jiajin, "Integrated security architecture for web services and this challenging," *JATIT*, pp. 518–525, 2005.

[7] J. Buchmann, *Introduction to Cryptography*, Springer, 2013.

[8] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, New York, NY, USA, 1996.

[9] G. Jakimoski and K. P. Subbalakshmi, "Cryptanalysis of some multimedia encryption schemes," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 330–338, 2008.

[10] O. Tayan and Y. M. Alginahi, "A review of recent advances on multimedia watermarking security and design implications for digital Quran computing," in *Proceedings of the 2014 4th International Symposium on Biometrics and Security Technologies, ISBAST 2014*, pp. 304–309, August 2014.

[11] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594–2608, 2016.

[12] P. S. Lokhande, "E-commerce applications: Vulnerabilities, attacks and countermeasures," *IJARCET*, pp. 499–509, 2013.

[13] J. Oberheide, E. Cooke, and F. Jahanian, "Cloudav: N- version antivirus in the network cloud," in *Proceedings of the USENIX Security*, pp. 91–106, 2008.

[14] D. Florncio, C. Herley, and B. Coskun, "Do strong web passwords accomplish anything?" *HotSec*, article 7, 2007.

[15] S. Guleria and D. S. Vatta, "To enhance multimedia security in cloud computing environment using crossbreed algorithm," *IJAIEM*, pp. 562–568, 2013.

[16] S. Riaz and S.-W. Lee, "A robust multimedia authentication and restoration scheme in digital photography," *Multimedia Tools and Applications*, vol. 73, no. 3, pp. 1291–1321, 2014.

[17] M. Cusumano, "Cloud computing and SaaS as new computing platforms," *Communications of the ACM*, vol. 53, no. 4, pp. 27–29, 2010.

[18] W. Zhu, C. Luo, J. Wang, and S. Li, "Multimedia cloud computing," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 59–69, 2011.

[19] H. Wang, S. Wu, M. Chen, and W. Wang, "Security protection between users and the mobile media cloud," *IEEE Communications Magazine*, vol. 52, no. 3, pp. 73–79, 2014.

[20] N. Jain and G. Kaur, "Implementing des algorithm in cloud for data security," *VSRD-IJCSIT*, pp. 316–321, 2012.

[21] Y. Ren, J.-C. Chen, J.-C. Chin, and Y.-C. Tseng, "Design and analysis of the key management mechanism in evolved multimedia broadcast/multicast service," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8463–8476, 2016.

[22] D. Yadav, D. Malwe, K. S. Rao, P. Kumari, P. Yadav, and P. Deshmukh, "Intensify the security of one time password using elliptic curve cryptography with fingerprint for e- commerce application," *IJES*, article 5480, 2017.

[23] SNAP, "Stanford large network dataset collection," http://snap.stanford.edu/data/index.html.