International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Optimizing Data Misuse Detection by Identifying Guilty Agent without Causing Disturbance and Inconvenience to Trusted Agent

Sangramsinh Deshmukh

JSPM"s Imperial College of Engineering and Research, Wagholi, Pune

Abstract: ***Now a day's in fact engaged data rich environment, prevention of data misuse is a challengeable field for data holders. A Data misuse is the way toward preventing delicate data by unauthorized people. There are distinctive ways for identifying the data misuse from untrusted agents. There are some diverse systems and techniques received for recognizing data misuse. To defeat the impediments of conventional frameworks, data portion systems are utilized to enhance the likelihood of recognizing guilty outsiders, without changing the first data. The proposition is to disperse the data with insight to agents in view of express data demand and test data demand keeping in mind the end goal to enhance the shot of identifying the guilty agents. On the off chance that any operator is leaks a part of conveyed data that agent is called guilty agent. The heuristic calculations are utilized to execute the blame on agent detection by including fake objects that enhance the distributor possibility of distinguishing guilty agents. Exploratory assessments demonstrate that specified proposed methods are effective and enhance the likelihood for finding the guilty agents in expansive observation database s gathered by frameworks.***

Keywords: Data misuse, Data monitoring, honeytokens and Information security.

1. Introduction

DUE to today"s competitive technologies related to big data, many companies outsource certain business processes and associated activities to a third party. This allows companies to focus on their core competency by subcontracting other activities to specialists, so that reduction in operational costs and increased productivity. In general, to carry out any services, the service providers must access to a company"s intellectual property and other confidential information. For example, for bank account, payroll done by an outsourcer, conditions that must have the salary and customer bank account numbers. The main security problem is that the service provider may not be fully trusted or may not be securely administered. Business agreements for try to regulate how the data will be handled by service providers, but in administrative domains, it is almost impossible to truly enforce or verify business policies. Because of their digital nature, relational databases are easy to duplicate and in case of financial incentives, a service provider may have to fail to handle redistributed data properly. Hence, we need powerful techniques that can detect and determine such dishonest. The aim of the system is to detect when the distributor"s sensitive data has been leaked by trusted and untrusted agents, and to identify the guilt agent [1].

To design guilt agent detection system which takes the data objects as input and returns guilt agents by analysing the search results. The purpose of this goal is that to optimizing data misuse detection by identifying guilty agent without causing disturbance and inconvenience to trusted agent. Our method suggests monitoring only a subset of data objects that are selected in such a way that detection rate is maximized and the monitoring effort is minimized. So that model this selection process with two independent optimization problems. Then present two heuristic algorithms for solving these problems and illustrate their efficiency in three scenarios [2]. The heuristic algorithms achieved either optimal or near optimal solutions for the optimization problems. In addition , illustrated the effect of the ratio of the trusted and untrusted agents, and the number of monitored objects on the expected detection rate.

To develop a system application for assess the “guilt” of agents, the algorithms for distributing objects to agents is used such that, in a way that improves our chances of identifying an agents which leaks information. Finally, consider the option of adding fake objects to the distributed set of agents. These fake objects do not correspond to real entities but appear realistic to the agents. Without modifying any individual members, it turns out an agent was given one or more fake objects were leaked confidential information and the distributor can be more confident that agent was guilty.

The objective of this paper is to how result of data leakages with the help of data allocation strategies from given the agents and also to identify the guilty party who leaked the data by adding fake data records without causing disturbance and inconvenience to trusted agents.

2. Literature Review

Recent research and surveys show that data misuse incidents can cause enormous damage due to privacy violations, identity theft, bank account exploitation, corporate data theft, and exposure of confidential information. According to Verizon reports [Baker et al. 2009; Baker et al. 2010], data breaches by insiders increased from 20% in 2008 to 48% in 2010, and data breaches by trusted partners increased from 32% in 2008to 38% in 2010. Moreover, attacks involving misuse of privileges increased from 22%in2008 to 48% in 2010. Statistics from the Data Loss DB Web site (www.datalossdb.org)show that in 2010, 45% of the

incidents reported on this site were caused by insiders. Koch [2011] claims that in reality the statistics are much higher, because many incidents involving insiders and partners are not published due to the firms concern with the loss of its reputation. The guilt detection approach presents about a data attribution problem [4] [6]. Tracing the extraction [5][12] set of objects imply effectively the detection of the guilty agents. And take for granted some prior information on the way a data view is formed out of data sources. Sometimes sensitive data is leaked [8] and found in unauthorized places. The distributor must evaluate probability of the leaked data came from the one or more agents.

There are also many research works on mechanisms that allow only authorized users to access sensitive data through access control policies [13][14]. Such approaches prevent in some sense data leakage by sharing information only with trusted parties. However, these policies are restrictive and may make it impossible to satisfy agents‟ requests. All of the agents in Papadimitriou and Garcia Molina [2010] are considered to be untrusted [2]. To identify , with a high probability, the source of a data leakage. This is done by finding an optimal allocation of the data objects among the agents and not by actively monitoring the interaction of agents with the data. The Decoy Document Distributor (D3) system planted into files (Honey files) this method investigate the use of such trap based mechanisms for the detection of masquerade attacks [7].

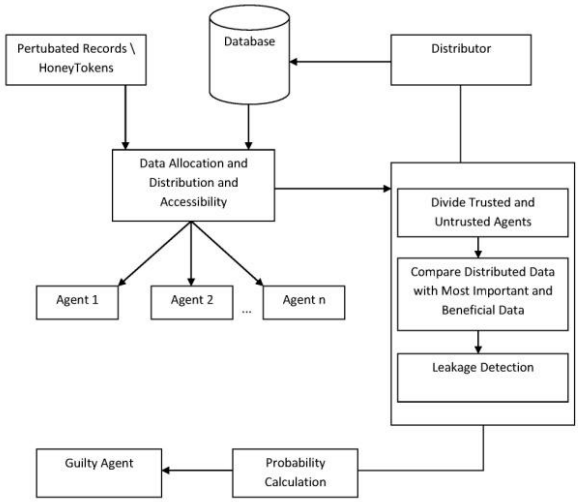
Existing solutions are difficult to detect faults in Web applications deployed in a large scale dynamic cloud computing environment due to the following reasons.

- Algorithm used to distribute the objects to agents that improve the chances of identifying a leaker.
- Realistic but fake objects are injected to the distributed set.
- Leakers cannot argue that they did not leak the confidential data, because this system traces leakers with good amount of evidence.
- In some applications the original sensitive data cannot be perturbed.
- Watermarks can be very useful in some cases but involve some modification of the original data.
- In some cases it is important not to alter the original distributors‟ data.

3. Proposed Methodology and Design

- **System Design**

A data distributor will give sensitive data to a set of supposedly trusted agents. When some of the data is found in an unauthorized place, the distributor may assess the likelihood that the leaked data came from one or more agents. We call the owner of the data the distributor and divide data in to trusted and untrusted agents, then the supposedly trusted third parties the agents. The guilty agent is one who leaks a portion of distributed data. And the one who receives the data from the guilt agent is an unauthorized person. Figure .1 shows the basic architecture of the proposed system.



In this figure, distributor distributes data to multiple agents. Before distributing data, distributor adds fake objects to original data and sends it to multiple trusted third parties i.e. agents. The distributor distributes data to multiple agents before distributing data; distributor adds fake objects to original data and sends it to multiple trusted third parties i.e. agents. After that, data distribution, distributor view in the database, to see how much and what data is sent to which agents. Once he done with checking the data, he finds out the probability for leaked data. Finally , list out the guilty agents. This proposed system consists of four modules:

3.1.1Framework Building for Data Distributor

Before distributing data, distributor adds fake objects to original data and sends it to multiple trusted third parties i.e. agents.

Planting Honeytokens: In object or data list for each agent, regard as the selection of planting fake objects, this process is also known nowadays as honeytokens. In our scenario, honeytokens can be preferred in an approach that each untrusted agents list will contain at least a predefined number of honeytokens while minimizing the probability so as these honeytokens are included in the trusted agents lists as well.

3.1.2Data Allocation Module

The most important focus of our proposed system is to handle data allocation problem as how can the distributor with intelligence give data or information to agents in order to improve the chances of detecting a guilty agent. From the entire set of objects, we are exploring, a list of accessible data objects is defined for each agent i. e. for both trusted and untrusted agents. However, as the proposed systems was not dealing with allocation problems but moderately are select objects for monitoring after the allocation has been done, the request category of each agent (explicit type or sample type) does not incense propose solution and therefore is not relevant.

3.1.3Leakage Detection Module

In this agents are divided into trusted and untrusted agents. For a particular group of agents, trusted agents are frequently the group‟s members, and untrusted agents are those who do not be in the right place to the group. The description as trusted or untrusted agent depends on the correlation to the

data objects right to use. For example, internal staff members of the society or group may be considered as trusted while delegates of third party and company associates will be considered as untrusted, or in the marketing field, the employees of one department may be considered as untrusted when referring to the information or data that is under the responsibility of another department in the same group or society e.g., information technology. A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some of the confidential data or information is leak and originate in an unauthorized place e.g., on the website or somebody‟s workstation. The distributor must evaluate the likelihood that the leak confidential data or information came from one or more agents, as opposed to having been separately gathered by other means.

3.1.4 Probability Distribution and find the Guilt Agents

In this module issue or factors interact and to check if the interactions are match with our intuition. Also study two simple scenarios as impact of probability p and impact of go beyond between Ri and S. In each scenario there must have a target that has obtained all the distributors objects, i.e., T = S.

- **Mathematical Model**

1) Identify the set of Distributors. D = {d1; d2.dn} Where, D is the main set of Distributors.
2) Identify the set of agents A = {a1b1;a2b2.. : anbn} where, A is the main set of Trusted and Untrusted Agents
3) Divide Trusted and Untrusted Agents from A Trusted agents the group‟s members. Untrusted Agents do not belong to the group. Both are related to relation with data objects.
4) Identify the set of Leakage Detection L = {D; A; F; Pr} where, L Leakage Detection A →Agent F Set of fake objects Pr Probability functions.
5) Identify the set of Data Allocation Strategy Sample Request = SAMPLE (T; mi): Any subset records from T can be given to Ui. Explicit Request = EXPLICIT (T; Condi): Agent Ui receives all the T objects that satisfy Condi.
6) Set of Pertubated records P = {d1f1} Where, p is the set of Pertubated records.
7) Set of Guilty agents G = {g1d1} Where, G is the set of Guilty agents. (Pr + trusted agents).

To resolve optimizing problem of data distributors for distribution of data objects, initially to select data objects for monitoring that are being accessed by a minimum number of trusted agents while ensuring that each untrusted agents list contains at least k monitored objects ($k \geq 1$) . As such, ks value should be determined according to the monitoring effort required for each data object and according to the organizations monitoring ability. To achieve the target, minimize the number of objects that are included in the lists of trusted agents and are being monitored that is, minimize the number of cases where xi = 1and IL (ti) (oi) = 1. This is subject to the following two types of constraints:

1) In each untrusted agents list, at least k data objects should be monitored (total of |UT| = m constraints).
2) Each monitored data object can be accessed by c trusted agents at the most (total of |L| constraints one constraint for each data object in L).

$$Min \sum_{(i=1)}^{n} \sum_{(\forall O_{(j \in l(t_i))})} x_j[I_l(t_i)(O_j)]$$

subject to

$$\forall ut_i \in UT : \sum_{(Ojl(uti))} x_j[I_l(ut_i)(O_j)] \geq k(1)$$

$$\forall O_i \in L : \sum_{(\forall t_j \in T)} x_i[I_l(t_j)(O_i)] \leq c$$

$$\sum_{(i=1)}^{n} x_i \leq b$$

$$\forall i, x_i \in \{0, 1\}$$

Optimization problem of data distributors is assumed that the organization has assigned a budget for the monitoring task, which is represented in the last constraint [1].

- **Algorithm**

General algorithm for Guilt detection Models: Input: Sensitive data with fake records. Output: Guilty agent‟s record.
Step1: Initializes the list which contains agent who has received the input data.
Step2: Divide agents trusted agents and untrusted agents.
Step3: Get the list of fake records and agents from the database and store it in different lists.
Step4: Find the guilty agent from untrusted agents.
Step5: If fake record is found in the data we are comparing then respective agent is guilty agent
Step6: If no guilty found try and find a leaker from trusted agents.
Step7: If any agent from trusted agent is found guilty convert that agent to untested agent and repeat from step4

This algorithm achieved either optimal or near optimal solutions for the optimization problems. It will illustrate the effect of the ratio of the trusted and untrusted agents, and the number of monitored objects on the expected detection rate. It possibly implemented guilt detection method by integrating honeytokens.

4. Result and Experiments

First assumes and verify that all the hardware would function properly and all the information provided by the user should be correct to avoid any sort of problem regarding relationship with the customer.

- The user should be login to the system Access is dependent on basis of level.
- Username and password is availability to the administrator.
- The user is an authorized one

In experiment, a random sample is selected without replacement from a population of objects, where some objects are classified as successful and the rest are classified as failure. The manual numerical experiment estimates the probability that a random sample will contain a certain amount of success, where the probability for success changes during the experiment. In the context of guilt detection,

where a malicious agent exploits several accessible objects inappropriately, we would like to estimate the probability that at least one of those misused objects will be monitor.

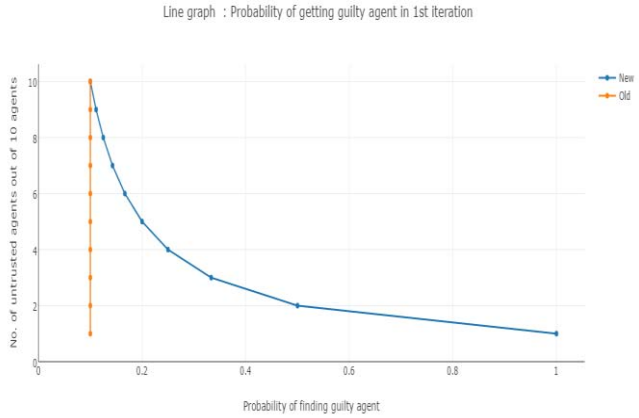


Chart -1: Probability of finding guilty agent in 1st iteration

The effectiveness of a system is described with its "no of untrusted agents "and "Probability of finding agent guilty in 1st iteration".

In above figure we compared 10 agents randomly containing trusted and untrusted agents, considering there is a guilty agent is present in the set.

So as a result we found that in old scenario where trusted agents and untrusted are not divide probability of finding guilty remains 0.1 as probability =1/10 , where as in our scenario probability of finding guilty is rapidly increasing in case of decrement of untrusted agents in the set below table is tabular representation of above diagram.

This will reduce time required to find guilty agents as well as efforts required to find out guilty agent.

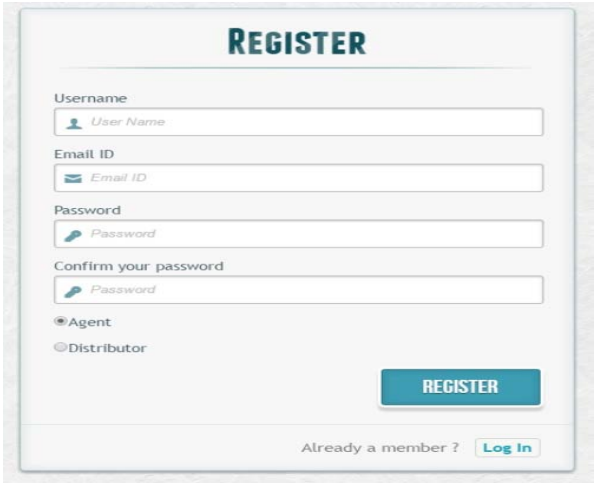
Table 1: Probability of finding guilty agent

No. of trusted agents	No. of untrusted agents	Probability of finding guilty agent with new scenario	Probability of finding guilty agent with old scenario
0	10	0.1	0.1
1	9	0.111111	0.1
2	8	0.125	0.1
3	7	0.142857	0.1
4	6	0.166667	0.1
5	5	0.2	0.1
6	4	0.25	0.1
7	3	0.333333	0.1
8	2	0.5	0.1
9	1	1	0.1

- **Output Screens:**

Following requirements are included in this project

5. Registration Module



- Registration Module will contain functionality to register user for Guilty Agent Tracker.
- Any distributor or agent can register via this screen.
- User has to fill in their correct details and select respective profile for successful registration.
- User will use their registered Username and Password for Login.
- After Distributor/Admin approves agents/distributors registration, Agents/Distributors will be able to login the system.

6. Login Module

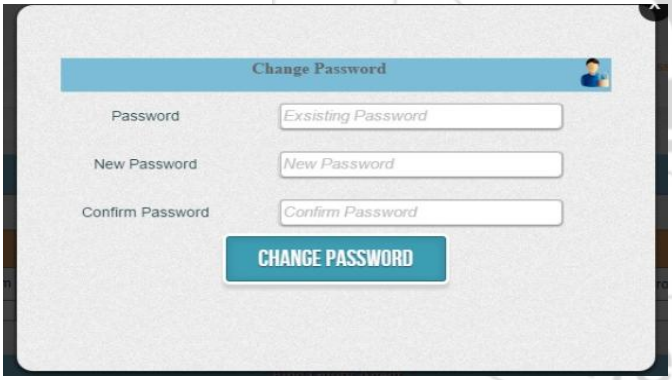
- Login Module will contain functionality of authorizing user.
- Profile includes Distributor, Agent and Admin; wherein user has to enter their respective credentials for Login.
- For successful login, User has to enter their Correct Username and Password provided at the time of registration.
- If incorrect user input is provided it will prompt validation for incorrect details and user will not be able to login the application.
- For Instance, if Agent wants to use this application, then respective agent will have to enter Username, Password and select Agent control for successful login.

- After Distributor/Admin approves agents/distributors registration, Agents/Distributors will be able to login the system.

7. Forgot Password Module

- Forgot Password Module will contain functionality wherein user can get their respective registered password.
- User will have to enter their Username provided while registration.
- Once user enters correct Username, email will be send to his registered mail id; which will contain his password.

8. Change Password Module



- Change Password Module will contain functionality wherein user can change their password.
- Authorized User will enter their old password and add/confirm new password.
- Once all the conditions are satisfied, user will be able to change their password.
- Only agent/distributor that is scrutinized by distributor/admin will be able to change password.

9. Admin Module

- Admin Page will be accessible only to Admin users.
- Admin user has two major role under Guilty Agent Tracker; Approval of Distributor and Deletion of Files.
  a. Distributor Approval: Once a new distributor registers to the system, Admin profile will scrutinize the details and approve/reject the distributor based on details provided.
  b. File Deletion: Admin will be able to delete files which are uploaded by distributor. If any deleted file is mapped under agent; the agent will not be download the file.

10. Agent Module

- All approved agents will be able to login Guilty Agent Tracker.
- Once successful login, agent will be able to download file which are tagged under him.
- On download of file, honey token will be appended to the file which was attached to agent by distributor while approval,
- Multiple files can be tagged under single agent.
- User can change password using Change Password link.

11. Distributor Module

Only authorized distributor will be able to login Guilty Agent Tracker. Once successful login, distributor will be able to perform three major roles.

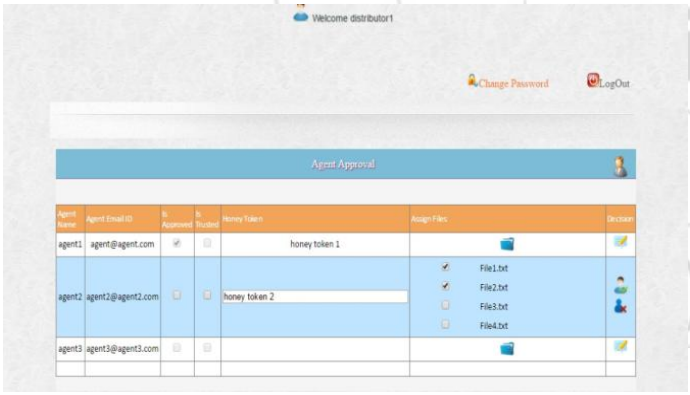
- Uploading Files and Documents
- Approval of Agent.
- Finding Guilty Agent.

Distributor Upload Module

Here approved and authorized distributor will be able to upload all file and documents which will be mapped to approve agent.

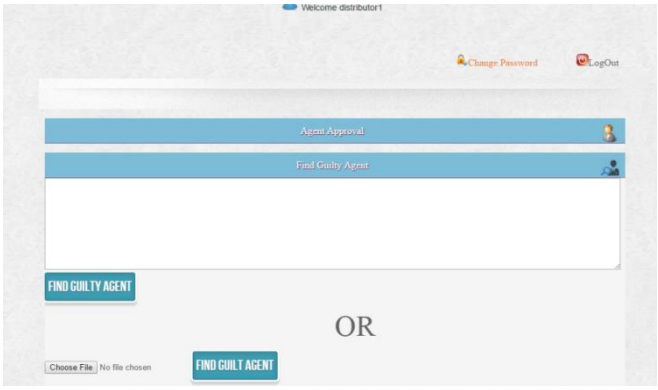
Files can be uploaded in text, word and excel format.etc.

Distributor Agent Approval Module



- Under Agent Approval Module, Distributor will approve new Agent based on registration details.
- After Agent approval, distributor will assign honey token to the agent which will be in useful when agent will download the assigned file, in order to indentify whether agent is guilty. Distributor can assign more than file to the agent.
- Distributor can also edit whether agent is trusted or guilty agent.
- If distributor does not approve, then agent will not be able to login the system.
- Only approved and trusted agent will be able to login successfully.

Finding Guilty Agent



- In finding guilty agent module , there are two options are present 1) Text copy paste 2) Upload document
- Any distributor can use any of above option to find a guilty agent.
- After pressing „Find guilty agent" internally all the permutation and combinations has been checked.
- List of fake records and agents from the database were fetched.
- Initially just untrusted agents were checked.
- If fake record is found in the data we are comparing then respective agent is guilty agent and that agent is blocked forever.
- If no guilty found try and find a leaker from trusted agents.
- If any agent from trusted agent is found guilty convert that agent to untrusted agent.

12. Conclusion

In competitive world, technologies based on many companies outsource certain business processes and associated activities to a third party. There is always need to hand over sensitive data to agents, who may maliciously leak it. To overcome drawback of the existing systems, present a method for selecting specific data objects to efficiently monitor and detect data mishandling incident perform by insiders. In the addressed scenario, trusted and untrusted agents are authorized to access a predefined list of data objects out of a shared data object collection. We also present allocation strategies for distributing objects with addition of fake objects to agents, in a way that improve probability of identifying guilt agents who leaks the confidential information. The proposed algorithm presented implements a variety of data distribution strategies. It is shown that distributing objects thoughtfully can make a significant difference in identifying guilty agents, especially in cases where there is huge overlap in the data or information that agents must receive.

13. Acknowledgment

We thank all the anonymous reviewers and editors for their valuable comments and suggestions to improve the quality of this manuscript.

References

[1] Asaf Shabtai, Maya Bercovitch, Lior Rokach, And Yuval Elovici ”Optimizing Data Misuse Detection”, ACM DOI: http: //dx.doi.org/10.1145/2611520 May-2014.
[2] Panagiotis Papadimitriou and Hector Garcia-Molina, Member IEEE. Data Leak-age Detection,IEEE Transaction on Knowledge and data Engineering, 2010.
[3] Papadimitriou, Panagiotis, ”A Model for Data Leakage Detection”, Data Engineering, 2009. ICDE ‘09. IEEE 25th International Conference, March 29 2009-April 2009.
[4] Pierangela Samarati and Sabrina C. di Vimercati. 2010 ”Data protection in outsourcing scenarios: Issues and di-rections”. In Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security (ASIA CCS10). 114.
[5] Jonathan White and Brajendra Panda. 2010. ”Insider threat discovery using automatic detection of mission critical data based on content”. In Proceedings of the 6th Interna-tional Conference on Information Assurance and Security (IAS10). 5661.
[6] Maya Bercovitch, Meir Renford, Lior Hasson, Asaf Shabtai, Lior Rokach, and Yuval Elovici. 2011. ”HoneyGen: An automated honeytokens generator”. In Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI11) 131136.
[7] Malek Ben Salem and Salvatore J. Stolfo. 2011. ”Decoy document deployment for effective masquerade attack de-tection”. In Proceedings of the 8th Conference on Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA11) 3554.
[8] Wu, Jiangjiang, ”An Active Data Leakage Prevention Model for Insider Threat Intelligence Information Process-ing and Trusted Computing”(IPTC), 2011 2nd International Symposium on, 22-23 Oct. 2011.
[9] Xiaosong Zhang , ”Research and Application of the Trans-parent Data Encpryption in Intranet Data Leakage Preven-tion,”Computational Intelligence and Security, 2009. CIS ‘09. International Conference. 11-14 Dec. 2009. [
[10] Kevin Borders, Xin Zhao, and Atul Prakash. 2006. ”Siren: Catching evasive malware”. In Proceedings of the IEEE Symposium on Security and Privacy. 7885.
[11] Li, Xiao-Bai , ”A Tree-Based Data Perturbation Approach for Privacy-Preserving Data Mining”, Knowledge and Data Engineering, IEEE Transactions, Volume: 18,Issue: 9, 2006.
[12] K. Deb, S. Agrawal, A. Pratab, T. Meyarivan, “A Fast Elitist Non-dominated Sorting Genetic Algorithms for Multiobjective Optimization: NSGA II,” KanGAL report 200001, Indian Institute of Technology, Kanpur, India, 2000. (technical report style)
[13] J. Geralds, "Sega Ends Production of Dreamcast," vnunet.com, para. 2, Jan. 31, 2001. [Online]. Available: http://nl1.vnunet.com/news/1116995. [Accessed: Sept. 12, 2004]. (General Internet site)