South Asian Journal of Research in Engineering Science and Technology

Volume 3, Issue **10**, **October 2018**

www.sajrest.com
SAJREST©2018 e-ISSN: 2455-9261

pp. **29—35**

GRAPH MINING ON TAX EVASION GROUPS USING BIG DATA

Arun D[1], Nazrin Banu A[2], Soundarya P[3], Bismina H[4],
[1]Assistant Professor, [2,3,4] UG Scholar,
Arasu Engineering College,
Kumbakonam, India.
arungetu@gmail.com

Abstract*—*Tax fraud has been a big concern for many organizations across industries billions of moneys are lost yearly because of this fraud. This paper aims to detect tax fraud using data mining tools. Currently, tax authorities face the challenge of identifying and collecting from businesses that have successfully evaded paying the proper taxes. In solving the problem of tax evaders, tax authorities are equipped with limited resources and traditional tax auditing strategies that are time-consuming and tedious. These continued practices have resulted in the loss of a substantial amount of tax revenue for the government. The objective of the current study is to apply a data mining technique to enhance tax evasion detection performance. Using a data mining technique, a screening framework is developed to filter possible non-compliant value-added tax (VAT) reports that may be subject to further auditing. The results show that the proposed data mining technique truly enhances the detection of tax evasion, and therefore can be employed to effectively reduce or minimize losses from VAT evasion. To propose a Colored Network-Based Model (CNBM) for characterizing economic behaviors, social relationships, and the IATs between taxpayers, and generating a Taxpayer Interest Interacted Network (TPIIN). To accomplish the tax evasion detection task by discovering suspicious groups in a TPIIN, methods for building a patterns tree. To this end, this project greatly improves the efficiency of tax evasion detection, as well as provides a clear explanation of the tax evasion behaviors.

I. INTRODUCTION

Big data is a term for data sets that are so large or complex that traditional data processing application software's are inadequate to deal with them. Challenges

include capture, storage, analysis, data curation, search, sharing, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem.

Data mining is an important part of knowledge discovery process that we can analyze an enormous set of data and get hidden and useful knowledge. Data mining is applied effectively not only in the business environment but also in other fields such as weather forecast, medicine, transportation, healthcare, insurance, government…etc. Data mining gives financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank, and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner.

Tax Information System is a professional financial services company established by experienced CPA's and chartered accounts to provide cost effective and efficient financial and tax planning solutions to a wide range of clients across the globe. Tax information system is providing single window tax service to the consultants having income minimizing the overall tax burden claiming foreign tax credits and other tax planning strategies according to the income tax. The training data need to be manually labeled. The taxpayer covert relationships are not recorded in the National Tax Information Collection System (NTICS). It is too expensive to overlook the deductions. Let out professionals plan and prepare your tax returns to make sure you get the maximum benefit you deserve. Their efficiency is low as they need to identify the transactions (including their detail information) one by one. There exists a complex, covert interactive relationship between the transaction parties.

In this paper, we propose a colored network-based model (CNBM) for characterizing economic behavior, social relationship and the IATs between tax payers. In the first phase is to discover the suspicious groups from the heterogeneous information network built based on the CNBM. In the second phase, traditional methods can be used on all transactions related to the suspicious trading relationships to detect tax evasion within the set of suspicious groups. The proposed is more effective and efficient than existing it aims to select the suspicious relations first via other related data source and identify those transactions. This can overcome the difficulties in the existing system. It provides well organized and detailed literature review on detecting tax fraud. The proposed method, which identify valuable information by hidden patterns found in a large database.

II. RELATED WORKS

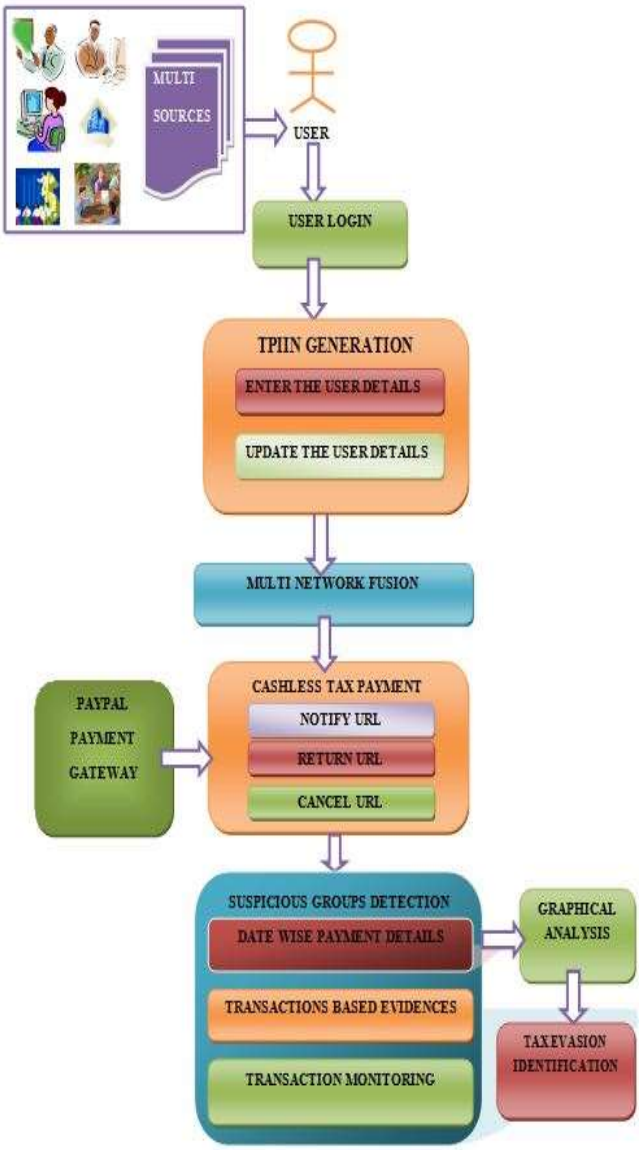
Xiaoqing Liu, Ding Pan and Shihong Chen, "Application of Hierarchical Clustering in Tax Inspection Case-Selecting" [13] illustrates that some enterprises have multiplicative ways of going about tax evasion, which becomes one puzzle in tax inspection. Tax inspectors to carry out rapid and accurate work have become extremely important. The traditional inspection case-selecting is mainly based on reported information. This method to judge the delineation of the characteristics of those unscrupulous taxpayers largely depends on the past experience and some intuition of the professional inspectors. This paper uses the hierarchical clustering in the tax inspection case-selecting. First, this paper describes the theory of clustering. Second, it analyses the index data of 30 enterprises using the hierarchical clustering and gets the analyzing result. Finally, the result is compared with the known taxation case. Then we get the conclusion that the hierarchical clustering method can assist the case-selecting and can improve the efficiency and effect of the tax inspection. Pamela Castellon Gonzalez, Juan D. Velasquez, "Characterization and detection of taxpayers with false invoices using data mining techniques" [7] illustrates that it is possible to characterize and detect those potential users of false invoices in a given year, depending on the information in their tax payment, their historical performance and characteristics, using different types of data mining techniques. First, clustering algorithms like SOM and neural gas are used to identify groups of similar behavior in the universe of taxpayers. Then decision trees, neural networks and Bayesian networks are used to identify those variables that are related to conduct of fraud and/or no fraud, detect patterns of associated behavior and establishing to what extent cases of fraud and/or no fraud can be detected with the available information. This will help identify patterns of fraud and generate knowledge that can be used in the audit work performed by the Tax Administration of Chile (in Spanish Service de Impuestos Internos (SII)) to detect this type of tax crime. J. Hasseldinea and G. Morris, "Corporate social responsibility and tax avoidance: A comment and reflection," [10] introduces that 'Smoke and Mirrors' (hereafter S&M) identifies an area of considerable importance but that it is misleading and problematic for several reasons. First, it glosses over the important distinction between tax avoidance

and tax evasion. Despite using the term 'tax avoidance' in the title, to establish its conclusion, the paper relies predominantly on a handful of examples involving fraud, deceit and corruption, which are behaviors usually associated with 'tax evasion'. In the context of corporate social responsibility, we explain why this distinction is crucial and offer directions for future research in this area. Second, Sikka's paper ignores voluminous extant research on tax compliance, corporate tax avoidance and its relationship with CSR. Third, the paper mis-reports key statistics on the tax gap in the UK and US, and finally, it omits a robust discussion of the considerable policy response to corporate tax avoidance, which has been promoted by numerous tax agencies and international organizations such as the OECD. In the current paper, while recognizing the merits of S&M, we highlight the problems listed above, seek to remedy them, identify additional areas of concern and encourage further research attention in this area. W. F. Fox, L. Lunab, and G. Schau, "Destination taxation and evasion: Evidence from US inter-state commodity flows," [5] introduces a structural model where commodity shipments are differentiated by origin state and examine how commodity flows respond to destination taxes, allowing for tax evasion as a function of distance between trade partners. After accounting for transportation costs, we find that the effect of taxes decreases as distance increases. This is consistent with the notion that longer distances between trade partners hinder government oversight and increase the likelihood of successful tax evasion. This is important to policymakers because it evidences the difficulty of enforcing destination taxation in open economies such as U.S. states and the European Union. L. Antunes, J. Balsa, and H. Coelho, "Agents that collude to evade taxes," [2] explores the link between micro-level motivations leading to and being influenced by macro-level outcomes to study the complex issue of tax evasion. If it is obvious why there is a benefit for people who evade taxes, it is less obvious why people would pay any taxes at all, given the small probability of being caught, and the small penalties involved. We use exploratory simulation and progressively deepening models of agents and of simulations to study the reasons behind tax evasion. We have unveiled some relatively simple social mechanisms that can explain the compliance numbers observed in real economies. We claim that simulation with multiple agents provides a strong methodological tool with which to support the design of public policies.

III. ARCHITECTURE DESIGN

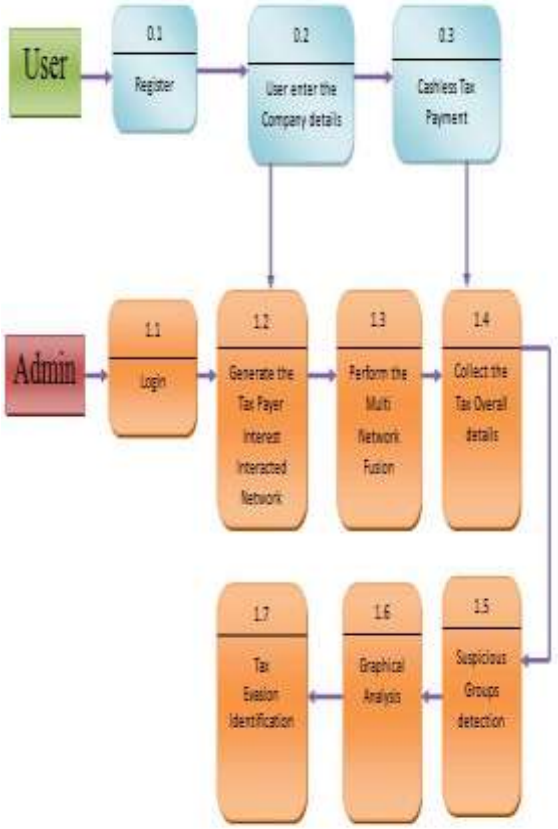
System architecture is the conceptual model that defines the structural behavior and more views of the system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about structures and behaviors of system.

Below Architecture depicts the entire system flow on tax payers from various sources. First of all user can register and login into the website. The registered users give the company details. From the User details, based on that admin can generate the taxpayer interest interacted network. Next the admin can perform the multi network fusion.



The user can pay the income tax by using the PayPal payment gateway. The user payment informations are collected and detect the suspicious groups based on date wise payment details called as Suspicious groups detection. Based on the evidence admin can identify the tax evaders and generate the report in the graph format called as Graphical Analysis.

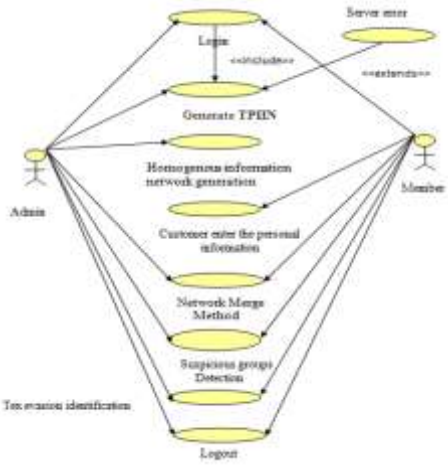
IV. DATA FLOW DIAGRAM



A Data flow diagram is a graphical representation of the "flow" of data through an information system, modeling its process aspects.

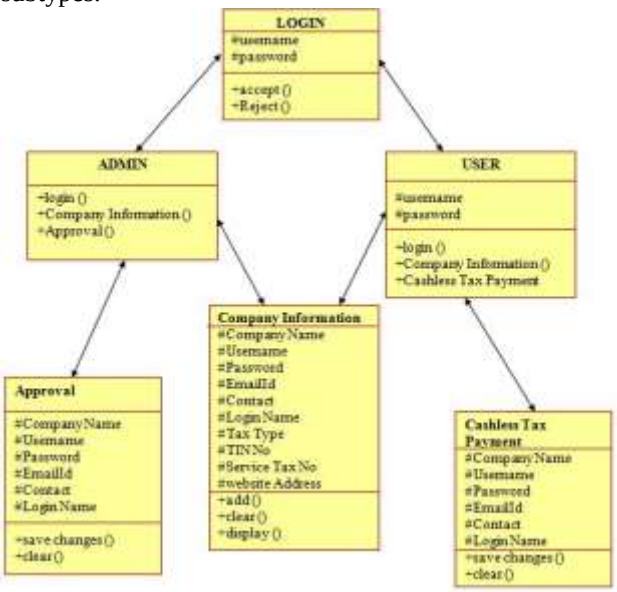
V. USE CASE DIAGRAM

Use Case represents specific flow of events in the system. Use case defines the outside (actor) and inside (use case) of the system's behavior. A Use case diagram is graph of actor, a set of use case enclosed by a system boundary, association between the actors and the use cases and generalization among the use case.



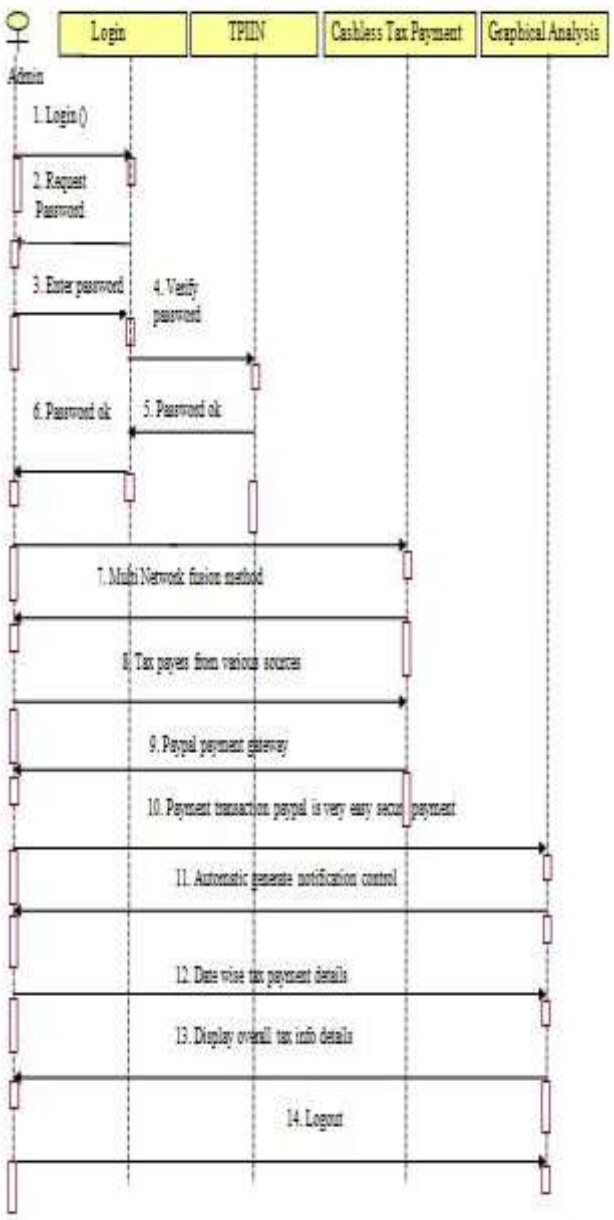
VI. CLASS DIAGRAM

Class diagram describes file types of objects in the system and various kinds of static relationship that exist among them. There are two kinds of relationships namely association and subtypes.



The Class diagram shows the attributes and operation.

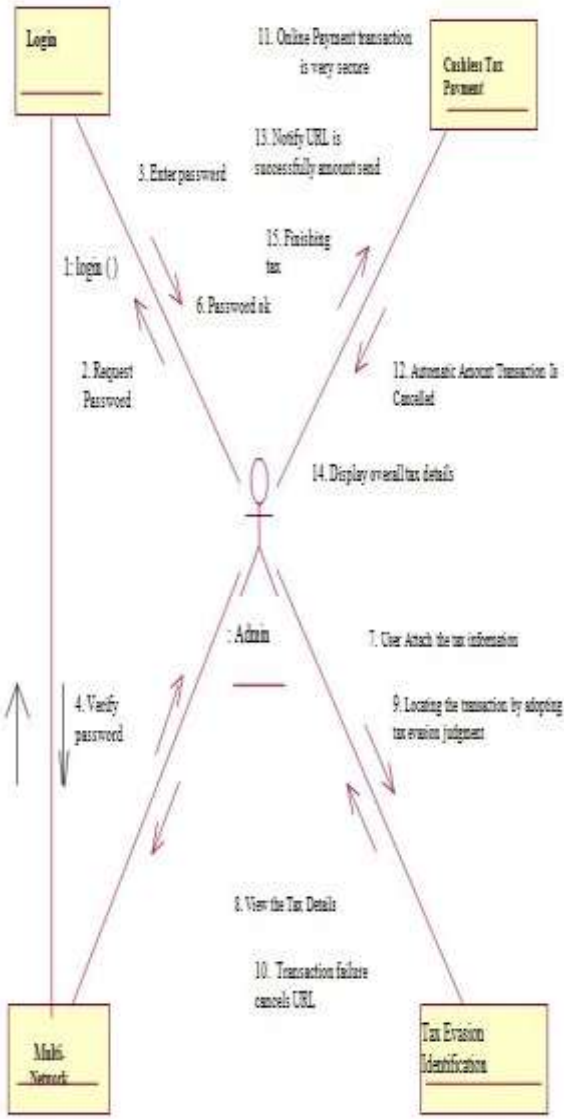
VII. SEQUENCE DIAGRAM



Sequence diagram is an interaction in time sequence. It shows the objects participating in interaction by then lifelines and the message they exchange, arrange in a time sequence. It has two dimension namely vertical dimensions represent time, horizontal dimension represent different objects.

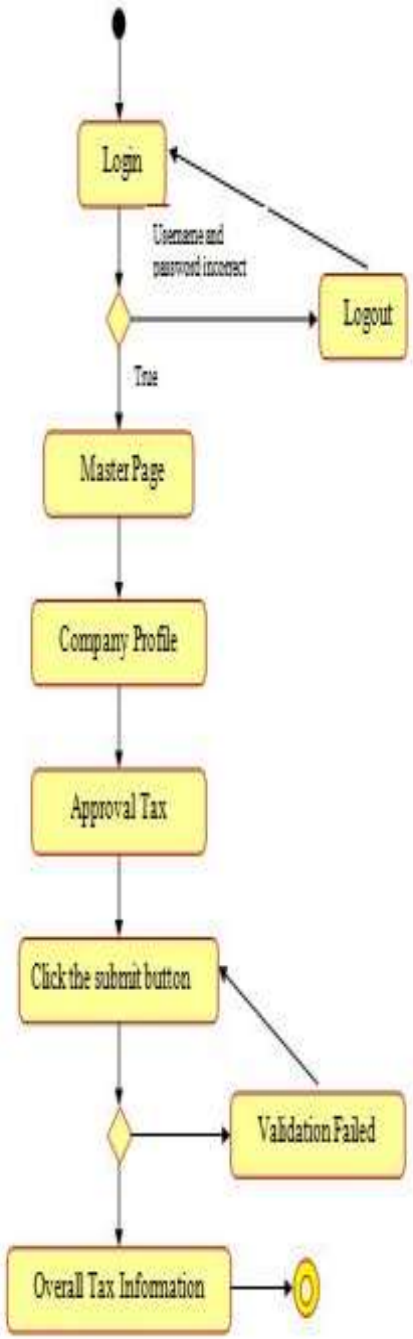
VIII. COLLABORATION DIAGRAM

Collaboration diagram represents a collaboration which is a set of objects related to achieve a desired outcome. In collaboration, the sequence is indicated by numbering the message several numbering schemes are available.



IX. ACTIVITY DIAGRAM

Activity diagram is useful in connection with overflow and in describing behavior that has a lot parallel processing. It describes sequencing of activity with for conditional and parallel behavior. Activity diagram is a variant of state diagram in which states are activity states.



X. SYSTEM IMPLEMENTATION

a. Generation of TPIIN

TPIIN is generated after a Customer enter the details from various information sources. The generated TPIIN is a large scale graph, our task of identifying the suspicious tax evasion groups is a three-step approach as follows. The first step is to segment a large scale TPIIN into small weakly connected sub graphs by applying divide and conquer strategy. The second step is to propose an algorithm for constructing a patterns tree. The third step executes the task of detecting the suspicious groups of tax evaders. The task finds any two matched component patterns.

b. Multi network fusion

Gathering corresponding data from various information sources, different homogeneous relationship graphs are formed. Then, after carrying out a procedure of multi network fusion on these homogeneous relationship graphs.

c. Cashless tax payment

PayPal is an India, international digital wallet based e-commerce business allowing payments and money transfers to be made through the Internet. Online money transfers serve as electronic alternatives to paying with traditional paper methods, such as checks and money orders. The Payment Gateway module is used to transfer the amount through the PayPal payment account. The amount transfer in PayPal gateway is more secured. This is a facility provided to the taxpayers to make income tax payments through internet using PayPal payment gateway facility. Three sub Modules:

i) Notify URL

The URL of the scripts that checks the IPN notification from PayPal and send back a confirmation to PayPal.

ii) Return URL

The URL where users want to send the customer to after the payment is complete.

iii) Cancel URL

Cancel URL when the payment fails the customer is redirected.

XI. GRAPHICAL ANALYSIS

Working with your data is just as important as storing it correctly, through the Graphical Reporting component, you can identify trends, view unfinished tax, identify project bottlenecks, visually search through interactive data giving you the insights you need. Graphical reports, also referred to as visual reports, implement graphical elements to make the

report data more visually appealing and to enhance usability by displaying data graphically in chart or graph formats.

XII. CONCLUSION

In this paper, the identification of the tax evaders can be deployed by using tax payer interest interacted network. Moreover, we proposed a multi network fusion method to integrate all the transactions based evidences for detecting the suspicious groups. Through the Graphical Reporting component, we can identify trends, view unfinished tax, identify project bottlenecks, visually search through interactive data giving you the insights you need. We implement graphical elements to make the report data more visually appealing and to enhance usability by displaying data graphically in chart or graph formats. Thus the tax evaders behavior can be displayed.

REFERENCES

[1]Liu, Xiaoqing, Ding Pan, and Shihong Chen. "Application of hierarchical clustering in tax inspection case-selecting." Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on. IEEE, 2010.

[2]Gonzalez, Pamela Castellon, and Juan D. Velasquez. "Characterization and detection of taxpayers with false invoices using data mining techniques."Expert Systems with Applications 40.5 (2013): 1427-1436.

[3]Hasseldine, John, and Gregory Morris. "Corporate social responsibility and tax avoidance: A comment and reflection." Accounting Forum. Vol. 37. No. 1. Elsevier, 2013.

[4]Fox, William F., LeAnn Luna, and Georg Schaur. "Destination taxation and evasion: Evidence from US inter-state commodity flows." Journal of Accounting and Economics 57.1 (2014): 43-57.

[5]Antunes, Luis, Joao Balsa, and Helder Coelho. "Agents that collude to evade taxes." Proceedings of the 6th international joint conference on Autonomous agents and Multiagent systems. ACM, 2007.

.