

# Design and Mining Suspicious Tax

<sup>1</sup>Dipalee Dedge, <sup>2</sup>Nikita Dedge, <sup>3</sup>Komal Hargude, <sup>4</sup>Pratiksha Kolhe, <sup>5</sup>Prof. Mrs. S.R.Hiray

Department of Computer Engineering, Sinhgad's College of Engineering, Pune-41.

**Abstract-** Now a day, managing taxation information is more challenging. Many peoples and group of companies find a legal way to avoid the tax. So it is important to find a way to identify these suspicious individuals and companies. For that, Graph-based method is used to characterize their property, which detects and describe suspicious relationships behind an Interest Affiliated Transaction (IAT). To identify economic behaviors, social relationships and the IAT's between the taxpayers and generating Tax-payers Interest Interacted Network (TPIIN) and also using the Colored Network Based Model (CNBM).

**Keywords:** Tax Evasion, Interest Affiliated Transaction, Big Data, Homogeneous and Heterogeneous Network

## 1. INTRODUCTION

Tax is imposition financial charge or other levy upon a taxpayer by a state or other the functional equivalent of the state. Tax revenue collection is considered a top priority in every national jurisdiction [1]. Tax evasion occurs when individuals deliberately fail to comply with their tax obligation. The resulting tax revenue loss may cause serious damage to the proper functioning of the public sector.

In India, most of the persons do not pay their taxes. They try to avoid this by taking the benefit of some loopholes in the Indian tax system. Thus the amount, which would have been used for economic and social development, is used for anti-social activities. All this creates black money and social evils in the society. The level of Evasion Tax also depends on the chartered accountants and tax lawyers who help companies, firms, and individuals evade paying taxes. Tax Evasion is a crime in all major countries.

### 1.1 Survey of Tax-payers in India

According to the report released by Indian Finance Ministry, estimated number of taxpayers for financial year 2011-12 stands at just 3.24 crore people. That means, less than 3 people in 100 pay taxes. Out of these 3.25 crore people, 89 per cent pay taxes in the tax slab of 0 – 5 Lakh rupees, while on the other end of spectrum, only 1.3% of all tax payers have income about 20 Lakh [3].

**Table 1: Number of Tax Payers in India & their Slabs**

| Slab              | Number in lakhs | Percentage of tax payers |
|-------------------|-----------------|--------------------------|
| 0-5 lakh          | 288.40          | 89%                      |
| 5-10 lakh         | 17.88           | 5.5%                     |
| 10-20 lakh        | 13.78           | 4.3                      |
| More than 20 lakh | 4.06            | 1.3                      |

**Source:** Report released by Ministry of Finance, April, 2012.

**Table 2: Income Tax collection**

| Slab              | In Crore | % of tax collected |
|-------------------|----------|--------------------|
| 0-5 lakh          | 15010    | 10.10%             |
| 5-10 lakh         | 21976    | 14.80%             |
| 10-20 lakh        | 17858    | 12.10%             |
| More than 20 lakh | 93229    | 63%                |

**Source:** Report released by Indian Finance Ministry, April, 2012

## 1.2 Detailed Design of the System

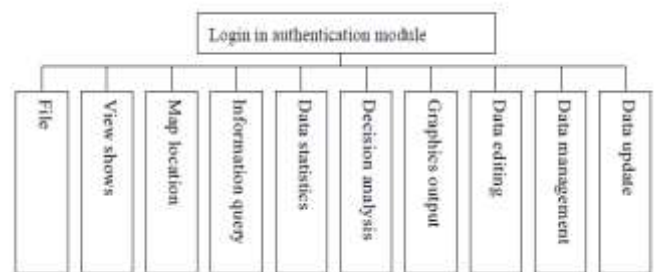


Fig 1: System Function Diagram

- A. Design of login module: This is a common module for ordinary users and administrator of the system. It provides username and password as ordinary module. For login valid username and password is necessary for performing different operations on data. It also keeps the record of logging of users, whether the login is success or fails [2]. It contains following functions:
  - B. Data Update
  - C. Data Management
  - D. Data Editing
  - E. Graphics Output
  - F. Decision Analysis
  - G. Data Statistics
  - H. Information Query
  - I. File

## 2. RELATED WORK

**Task Force on Direct Taxes** was introduced in 2002. It recommended the income tax department must increase expenditure on tax payers services, the Permanent Account Number should be extended to cover all citizens and therefore serve as a Citizen Identification Number, the

department should set up a structure for Electronic Data Interchange (EDI). **Torgler (2006)** is the study observed the impact of non-economic factors on three tax compliance variables namely justification of tax evasion, corruption and claiming government benefits without justification. **Arora R.S. and Rani Vanita (2010)** Studied and investigated the opinions of tax professionals from the different districts of Punjab regarding tax evasion and corruption in Indian Tax system. They pointed out the high tax rates, corruption, social acceptance of tax evasion, low probability of detection and low tax morality are the main causes of tax evasion [3].

The “Design and Implementation of property tax grid management system” states that the real estate management-It start from the viewpoint of the combination of collection resources; continuously strengthen department co-ordination to carefully push addition of real estate tax management [2].

The “Taxation on land development with the density control and environment externality” shows the social planner resp. implement the taxation on property development to improve the size of negative environmental externality and an increase of the tax rate on the free land.

The “ An Empirical Analysis of Factors Influencing Effective Tax Rates Of High-tech Enterprises in China” state the factors that disturb the effective tax rates (ETR) listed companies can be divided into external factors and internal factors, including the level of economic growth, tax policy, capital structure, profitability, and equity structure [4].

Objective of the work:

- 1) To find out the causes of income tax evasion.
- 2) To find out the impact of tax evasion in India

### 2.1 IAT Based Case Study

**Case 1:** A company C1 produces different products and its annual net profit was negative. The another company C2 held the shares of C1. The products of C1 are sold to the company C3. The different legal persons L1, L2, L3 controlling the companies C1, C2, C3 respectively. The Tax Administration Office (TAO) verified that C1 server as a producer, C2 for investment management and C3 for marketing analysis, delivery of products, sale. The TAO found that the same products produced by the similar scale enterprises in the same industry [1].

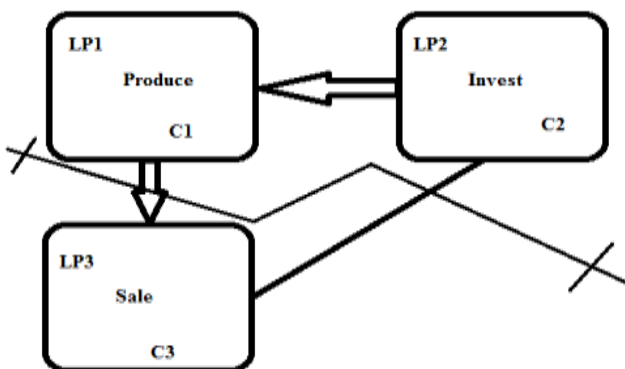


Fig 2: Case 1

**Case 2:** A company C2 in Pune sold 1000 number of products at 100 Rs to C3. This price offered by C2 was very cheaper, though the rough price was 170 Rs. which they offered to other companies. After the verification, found that companies C2 and C3 also had relationship with C1, C1 had the shares in C2 and C3 companies. Therefore, TAO believed that the companies C2 and C3 made tax adjustment.

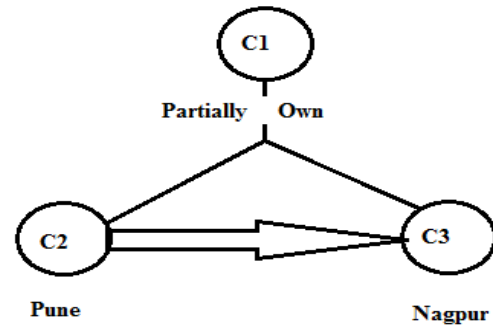


Fig 3: Case 2

### 2.2 Needs and Origin

From the case studies the summary can be as follows: Number of complex, hidden interactive relationships are exist between so many companies, A) the relationship can be in between companies executive or managers or may be in between legal persons of different companies. B) They share secure and hidden relationships between stakeholders. C) Or they may have same parent as a controller for two or more companies, who is involved in transaction. These relationships are heterogeneous and also of various types.

These observations encouraged to form a network which takes information from different sources like financial reports such as shares of companies or individual, profit of company, relationship with other company, an amount of tax payed by the company, individual's property.

The Fig 3 gives the relationships between the legal persons of companies and the tax-payers information and forms a homogeneous network. These homogeneous networks are group into heterogeneous network, which is represented using Colored Network Based Model (CNBM) [1].

The above Fig 3 is used to finding out relationships, which are stated above. For this data of each individual is collected. Then all the homogeneous information is collected from the companies and tax-payers. These all-homogeneous networks are grouped together to form the heterogeneous network and detect the companies and individuals who are avoiding tax.

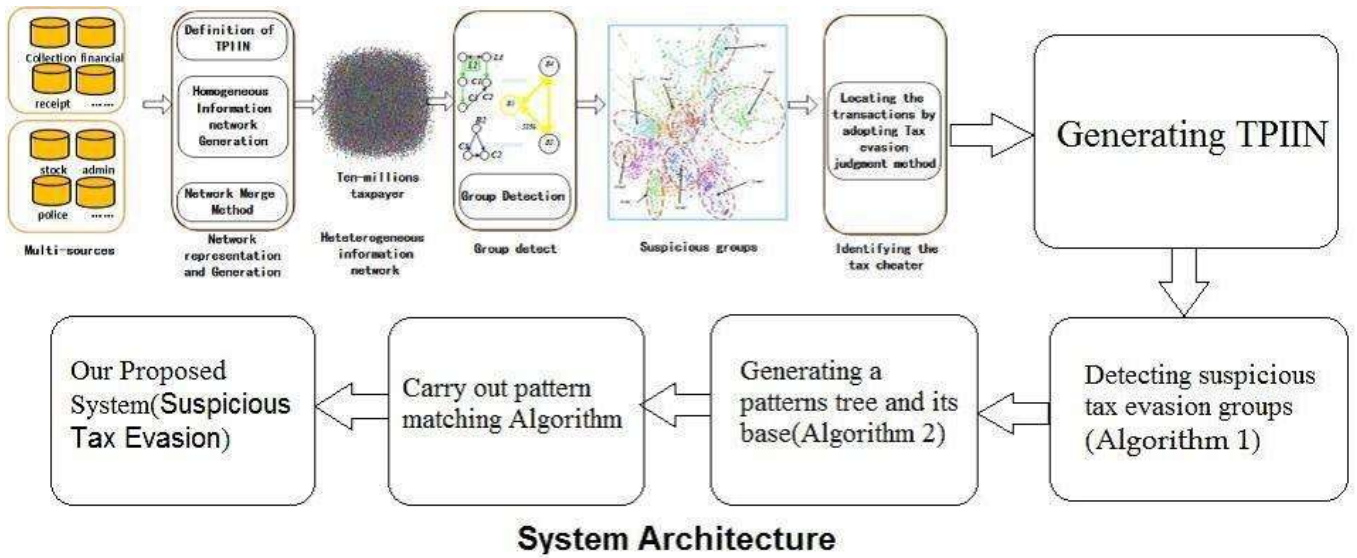


Fig 4: System Architecture

### 3. GENERATION OF TPIIN

TPIIN is a Tax-Payers Interest Interacted Network. TPIIN is generated after Multi-Network Fusion Method has been adopted to abstract various relationships between individuals from different information sources. TPIIN is a large-scale graph and used to identify the suspicious tax.

For TPIIN generation two algorithms are used:

**Algorithm 1:** Detecting suspicious tax evasion groups.

Input: Information about all companies, individuals and their relationships.

Output: Suspicious groups are detected.

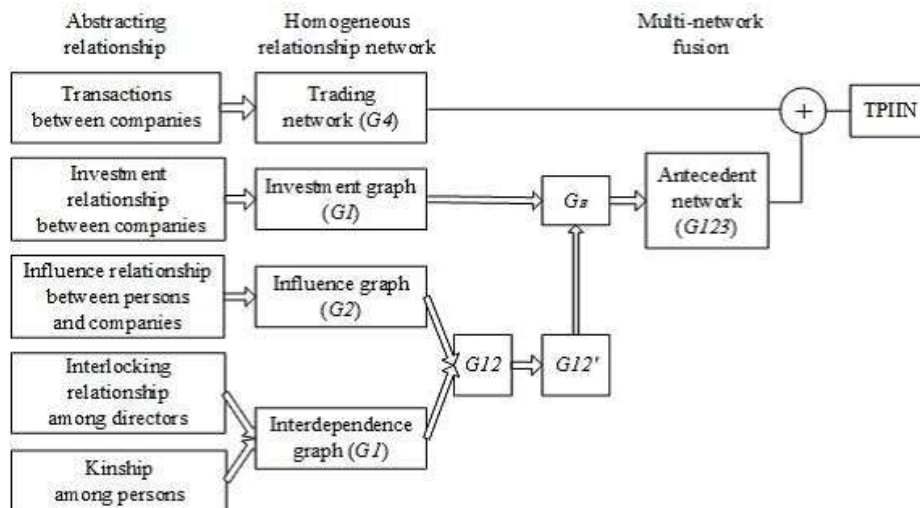


Fig 5: Multi-Network Fusion Method [1]

Abstracting relationships take the all information about different companies, their relationships, which can be kinship and interlocking. The homogeneous relationship network collects all data about the companies and generates the graph like investment graph, influence graph, trading network. These are graphs are combined together to form the  $G_{123}$  which is antecedent network and forms Directed Acyclic Graph (DAG).

Begin

1. Find all companies involved in the influence graph and generate antecedent network.
2. Abstract all the trading relationships to form a trading network.
3. Find each Maximal Weakly Connected Sub-graph (MWCS) and generate the set of information about all relationships.

4. After this pattern matching is applied.

Stop

**Algorithm 2:** Generating patterns tree and its base.

Input: Output of algorithm 1.

Output: Find all patterns successfully.

Start

1. For each node find out the in-degree and out-degree and store them into two different arrays.
2. Then sort all these elements and store them in a sorted order depending upon in-degree and out-degree.
3. Store all the node data according to the Edge and Color.
4. From the generated TPIIN the antecedent and trading network is formed.
5. Pattern tree is generated.

Stop

#### **4. CONCLUSION**

The new system should offer greater flexibility for the tax management. Using this way it can find the suspicious taxpayers who are not giving proper amount of tax. The proposed system uses various homogeneous relationships in big data to form the heterogeneous information network.

#### **REFERENCES**

- [1] Feng Tian, Tian Lan, Kuo-Ming Chao, Nick Godwin, Qinghua Zheng, Nazaraf Shah, Fan Zhang, "Mining Suspicious Tax Evasion Groups in Big Data", IEEE 2016
- [2] WANG Jue1 ,"Design and Implementation of Property Tax Grid Management System", IEEE 2014
- [3] Lalit wadhwa, Dr. Virender Pal "TAX EVASION IN INDIA: CAUSES AND REMEDIES" , journal 2012
- [4] ZHAN Di, LI Xiu-hua "An Empirical Analysis of Factors Influencing Effective Tax Rates of High-tech Enterprises in China" , 2013 International Conference