Implementing Peer Group Analysis within a Track and Trace System to Detect Potential Fraud(s)

Foo Chi Hui[1], Venkaiah Chowdary Koneru[2], Norazman Mat Ali[3], Safurah Harun[4]
Information Communication and Technology, MIMOS Berhad
Malaysia
[1]ch.foo@mimos.my
[2]venkaiah.chowdary@mimos.my
[3]norazman.matali@mimos.my
[4]safurah.harun@mimos.my

***Abstract*— Tracking and tracing of goods movement is a key requirement for supply chain management and analysis. Data collection can be broad and large in volumes. Goods can moves in complex supply chain distributions, where disputes, frauds and thefts can happens. This paper aimed to develop a practical method to analyze the incoming data and employ unsupervised potential fraud detection in near real-time. The method is designed and discussed around peer group analysis (PGA) approach which is commonly used in financial market. The paper shall focus on two steps. First, monitor and groups good movements and categorize vendors or suppliers with similar trend / behaviours into dedicated peers. Second build a tool / services that detect anomalies in event transactions. The monitoring service shall detect the outlier or individual objects that distinct from peers which potentially fraud / alerts.**

***Keywords*—** *supply chain, tracking and tracing, fraud detection, peer group analysis, PGA tools*

1. Introduction

Tracking and tracing is an important strategic process within a supply chain. It is the basic of knowing and recording the movement of goods. The purposes of tracking and tracing can be safety control, quality assurance control, logistical supply management, prevent counterfeiting, marketing and fraud detection. Any tracking and tracing software provide a systematic way to collect movement and transaction data from full supply value chain. With good architecture design, the traceability platform traces raw materials / ingredients of products too. Drilling down into each raw material reveals raw materials product lifecycle. Our study focuses on using PGA to detect potential fraud(s) and prevent hazards / contaminated products to reach retailers or consumers by providing early warning alerts and escalate potentially catastrophic consequences.

International Journal of Supply Chain Management
IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

PGA is an outlier detection method. It is fundamental in data mining. It is a method for monitoring local abnormality over time. PGA already has number of successful application in fraud detection, such as applying fraud detection in stock analysis [1] and credit card fraud detection [2]. Multiple fraud detection methods are available for fields in credit card, telecommunications, and network system intrusions. But supply chain product movement fraud detection area still lacking.

Fraud in supply chain can be broadly categorized into three groups.

- Behaviours fraud
- Application fraud
- Transactional fraud

In this paper, we are focusing on transactional fraud within supply chain. Based on interviews, data collections, fraud takes place when collectors / brokers / exporters try to manipulate their product origin, product volumes without regards for the consumer’s safety.

In section 2 of this paper, we will outline tracking and tracing platform.

In section 3, we discuss on how we use peer group analysis within tracking and tracing platform. Then in section 4 we illustrate PGA implementation with real data set consisting of observation from the prototype system.

2. Methodology

For statistical fraud detection methods we can broadly categories into “supervised” and “unsupervised” methods. Supervised method of fraud detection consists of models that are trained base on known fraud cases from large data source. Limitation of a supervised fraud detection method is it might suffer from unbalance class size [3] and has limitation on detecting only known patterns of frauds. On the other hand, unsupervised fraud detection method aims to identify subject with same behaviour/trend into groups automatically. Subject within a group are call peers. Peers with similar behaviours/trend are monitored for outliers. Once outliers detected, the subject will be mark as potential fraud.

Transactional fraud detection has been implemented using difference, such as data mining, clustering, statistics, and artificial intelligence. Two leading fields of transactional fraud detection are stock market fraud detections and credit card fraud detections.

Fraud detection using peer group analysis was first introduced by Bolton & Hand [5] for credit card fraud detection, where the only consideration of the method was the card's spending amount by a period of time. This method was proven insufficient for us to follow suits using only transactional values running our PGA tool in a supply chain. This is due to the complexity of supply chain. Measurement and units may change when product change hand. Take a common trade items for example, raw materials can come in as "Tons" and "Types" may varies. Hence we added support of multiple attributes within our tools. For example: location [location type, land size, output, and units]. Each attribute can be associated with "weight". A "weight" system allows domain expert to grant "weight points" for a specific attributes and enhance / fine tune our PGA fraud detection tools sensitivity.

To narrow down our scope of research, we have set the following goals:

- To identify trade item from which location of declared capacity are differences with actual delivered capacity.
- To identify farms / collectors (business location) which volumes rise and fall quickly
- To identify from which point counterfeit processed products.

To achieve the research goal, first data collected from differences sources are consolidated into single database. Data cleansing are done prior running PGA tools on it.

The tools are developed to speed up and simplify our test and simulation. We simulate the fraud cases data in database using the said tools and illustrated into various set of graft. The graft is used by human expert to verify against the accuracy of the alerts from our tools.

It is important to highlight that our method unable to predetermine actual fraudulent or contamination but determine a potential fraud base on outlier detection. Verification will still require expert or regulators to perform actual investigation. However we recommend our approaches for early determination of potential fraud being detected.

3. System Overview

PGA fraud detection tools are implemented on an in house track and trace platform. The track and trace platform is build according GS1 standard [4] for ease of integration with external system.

First module, our prototype "peer grouping mechanism" is design to group business location that having the same behaviours. Same behaviours are determined by number of transaction (events) at the location that tied to a specific product by a fixed time *T*. *T* is set by our tool configuration together with other attributes such as [business location type],[trade item type]. These build in parameter filtered irrelevant data and reduce processing time. Such location profiling essentially eliminates observes group outside the intended scope. A business location can be selected as *kpeer* with a set of pre-selected *criteria*. Details will be discussed in section 4. Peer grouping mechanism also function as a monitoring tool processing real time data and update peer category.

Second module is "peers monitoring and analysis mechanism". It provides a scheduler on scanning and monitoring real time transactions / events of business locations. It calculates 1Q, mean and 3Q of each group from beers by T. Different parameters can be included to increase the accuracy. Data are normalized and include into a scoring table. In experiment section we will provide more details on normalized scoring system.

Third module is "flagging mechanism". Flagging mechanism place and alerts on screen base on result detected by peers monitoring and analysis mechanism. Type of alerts depends on setup and configuration. For example, alerts can be flag by a company which means flagged company are detected with outline behaviours. On the other hand, alerts can be flag base on trade items. The system can't pin point where is the issue, it highlighted that with outlined behaviours will likely be a fraud. Human intervention is requires to investigate into alerts.

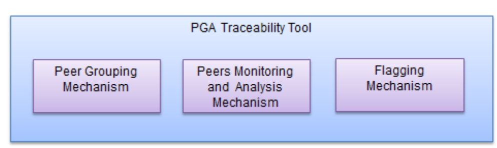

Figure 1. System overview

4. Experiment & Results

Our experiment runs on the following process:

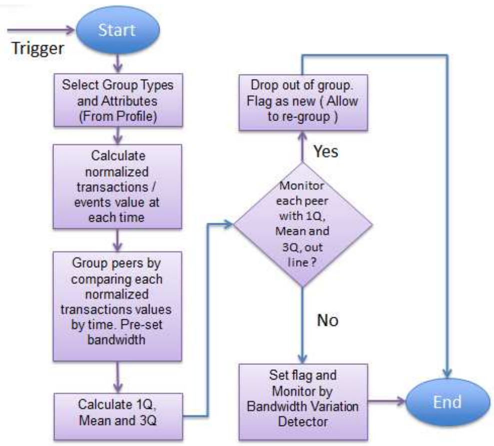

Figure 2. Process

We simulate data from October 2012, to December 2012 for the daily events of durian collections, processing and export from each of 33 exporters. The data are port over from manual processes. Data accuracy is not verified due to the difficulty of validating such items. However, we believe with traceability platform integrated, more streamlined and accurate data can be collected.

We set the categories as business location [exporters], the *T* is from 14 days to 2 months. The *npeer* = [15- 33]. Number of exporters = 203. Exporters consists of differences industries. Capacities (transactions) are normalized to metric tons.

A sample of data shown table below:

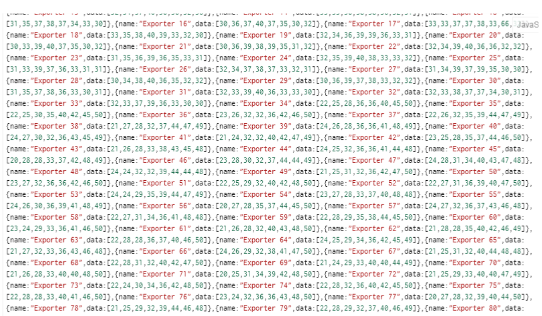

Figure 3. Data

We simulate our tools with different settings and variables many times, changing *T*, *npeer* and *groups*. The following graft is generated for comparison purpose. We only show the graft that is more interesting here.

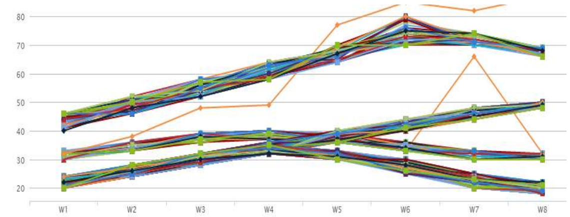

Figure 4. Indicated normalized capacity output versus T of all exporters.

Interestingly if we do not filter data by industries, PGA tools are able to pick up 28 of the exporters from our data as a group. This is partially because of the data taken from Durian Season and selective 33 exporters export durian.

Outlier is outline behaviours of one or multiple exporters. Our PGA tools outlier detection operates at group levels. We provide calculation of 1Q, Mean and 3Q (Standard deviation) for transaction of each days. Then we employ statistic test by providing another months of transaction into our data. We monitor if each of observed exporters within the group behave the same. Exporters that behave "gradually" different transactions over time are flag as possible re-grouping peers. It is possible that the exporter belongs to another group. The term "gradually" is defined by the outline behaviours that falls within 1Q and 3Q. However exporters with transaction over 1Q and 3Q are flag as outlier. Furthermore, we can reduce the sensitivity of outlier's detections by averaging number days over the 1Q and 3Q.

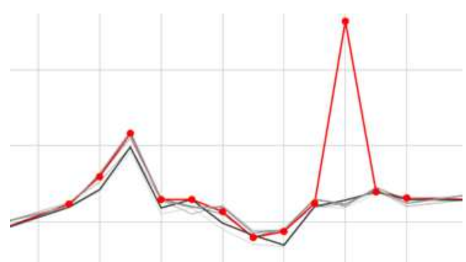

Figure 5. Sample data (a)

Figure 7 indicates Exporter 87 which transactional behaviours depart from Group [A]. The departure of Exporter 87 is smaller than 1Q and 3Q hence it is large enough to flag as an alert.

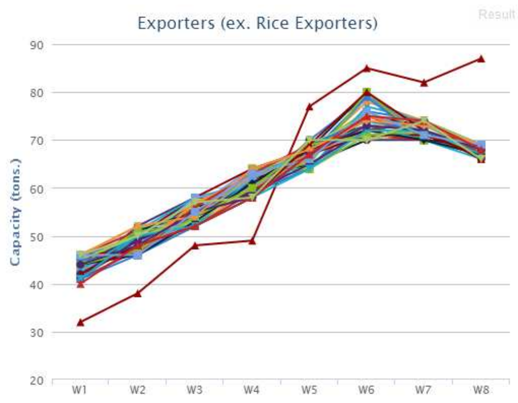

Figure 6. Sample data (b)

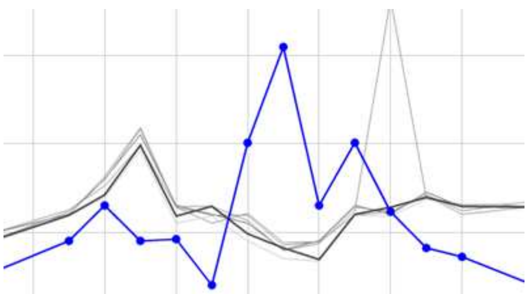

Figure 7. Sample data (c)

5. Discussion

From above Figure 4, 5 and 6 we can do a comparative analysis where Exporter 17 have sudden rise of capacity on 3rd week of December. The behaviours of this exporter were similar with its peers for a month. From here, our PGA tool flag it as potential fraud that worth investigating.

*Real world investigation carried out within related department, found out the exporter 17 is taking in unregistered farm's durian to export. However, the durian was being sampling for bacteria count and dim qualified.

For Figure 7, it shows the gradually departure of transactional behaviour from Exporter 87.

Our test is run based on passed consolidated records. For real world implementation, it is able to run on real time just like credit fraud detection. Thus PGA tool can help prevent expensive rejection on foreign country and damage Malaysia export fruits brand.

6. Conclusion

Our approach in this article describes early stage of research to produce a platform / frameworks for unsupervised fraud detection for supply chain.

In this paper we also demonstrated implementation of peer group analysis in an unsupervised supply chain transactional data. Peer group analysis groups' behaviour or character of business location from a sequence of set parameters and calculated / normalized transactions volumes. The results of running our tools are varying from data type. Data accuracy, structured and cleanliness made will take key criteria for accuracy of analysis. For example, an empty node of transaction can make a sudden drop of a node down which might trigger an alert.

Due to limited real world data collected to date (Number of transactions) performing unsupervised analysis accuracy hard to be validated. Further testing on our PGA tools is required especially on profiling on new cluster. However we understand that with domain expert we can narrow down the scope by applying more filtering measure and threshold.

However we have shown PGA have essentially able to identify spike or change of behaviours in cost effective way compare to generic audit. With tools available, PGA can help alert / identify potential fraudulent at early stage. We have illustrated such capability using visual chart.

The data set is chosen for these experiences. Such method can be expending to others transactional movement of products within supply chain. Especially for those high values products such as bird nest. PGA requires large group of data to formulate accurate alert.

7. Future Work

As an early stage of our research, there are a lot more to refine within our PGA tools and models. We plan to develop a dynamic risk profiling system aims to increase accuracy of clustering. Furthermore, a scoring system with weight age should be included into the equation. The scoring method can be based on additional parameters.

Most importantly, experiment with real world complete and accurate real world data. Implementation on other products type is within our plan too.

References

[1] Kim, Y. "Stock fraud detection using peer group analysis - ScienceDirect.com." 2012. <http://www.sciencedirect.com/science/article/pii/S0957417412002692>

[2] "Plastic Card Fraud Detection using Peer Group Analysis ..." 2007. 15 Mar. 2013 <http://videolectures.net/mmdss07_weston_pcf/>

[3] Bolton, Richard J, and David J Hand. "Unsupervised profiling methods for fraud detection." *Credit Scoring and Credit Control VII* (2001): 235-255.

[4] "Home | GS1 - The global language of business." 2003. 29 Mar. 2013 <http://www.gs1.org/>

[5] Bolton, Richard J, and David J Hand. "Unsupervised profiling methods for fraud detection." *Credit Scoring and Credit Control VII* (2001): 235-255.

[6] "Peer Group Analysis and Descriptive Statistics - IMF." 2006. 10 Jan. 2013 <http://www.imf.org/external/pubs/ft/fsi/guide/2006/pdf/chp15.pdf>

[7] "EPCIS (Electronic Product Code Information Service) - GS1." 2011. 11 Jan. 2013 <http://www.gs1.org/gsmp/kc/epcglobal/epcis/epcis_1_0-faq-20070427.pdf>