iCBST 2014 International Conference on Business, Science and Technology which will be held at Hatyai, Thailand on the 25th and 26th of April 2014.

Australian Journal of Basic and Applied Sciences, 8(5) Special 2014, Pages: 498-507

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

RFID-enabled Supply Chain Simulator for Counterfeiting Attack Detection

Manmeet Mahinderjit Singh and Lim Mei Teen

Distributed System & Security School of Computer Sciences Universiti Sains Malaysia

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form 12 March 2014
Accepted 14 April 2014
Available online 25 April 2014

Keywords:
Radio Frequency IDentification (RFID), Supply Chain Management (SCM), Monte Carlo (MC), Electronic Product Code (EPC).

ABSTRACT

Radio Frequency Identification (RFID) technology has been used in many application areas because of its efficiency and benefits. RFID technology supports many different types of application namely supply chains, access control application, healthcare and human tracking. Since the RFID data is application dependent and are owned by various products manufacturers and industries, sharing of data presumed to be confidential and sensitive are never an option. As a result, there is lack of RFID data for researchers to do testing in especially RFID supply chains. Due to security, privacy or cost issues, the data in the supply chain is vulnerable to both active and passive attack that lead to counterfeiting attacks. In this project, we will design an RFID-enabled supply chain simulator to generate data that able to track and trace the movements of RFID tagged items in the supply chain from manufacturer to distributor and then to the retailer. We will use some techniques such as Monte Carlo (MC) method to generate the dataset. In addition, injection of counterfeit RFID tags such as cloned and fraud tags will also be done. The prototype software will be used by the researchers who are interested in using the dataset for counterfeiting attacks such as cloned tags and fraud tags detection, Intrusion Detection System or any data mining techniques. This simulator stands as a complete package in generating single supply chain partner datasets, simulate overall single and multi-link supply chain datasets, generating cloned and fraud dataset and tracking and tracing the whereabouts of cloned and fraud RFID-tag. A cost-sensitive experiment is also executed by using dataset generated from the Monte Carlo Simulator and Weka tool.

To Cite This Article: Manmeet Mahinderjit Singh and Lim Mei Teen., RFID-enabled Supply Chain Simulator for Counterfeiting Attack Detection. ***Aust. J. Basic & Appl. Sci.,*** *8(5): 498-507, 2014*

INTRODUCTION

RFID is a technology which functions with radio-frequency waves to identify animals, people and objects automatically without light of sight (Ilie-Zudor *et al*, 2006; Derakhshan *et al*, 2007). It consists of three components: an antenna or coil, a reader and a radio-frequency tag where the reader will emit radio waves to activate the tag to retrieve the information stored in the (Derakhshan *et al*, 2007). Due to RFID tags can provide higher read rate, durability, security, and automation, it has been used in many applications like manufacturing and SCM, monitoring and tracking, human identification, healthcare, baggage, and toll road (Ilie-Zudor *et al*, 2006). However, in this project, we will only focus on RFID item tracking in SCM which involves only the manufacturers, distributors and retailers. There are a few challenges of designing a supply chain simulator as well. For instance, there are billions of tags that the supply chain has to track and trace every day (Derakhshan *et al*, 2007). It is not an easy job to track the timestamp and location of a tag, and trace back all the information as the tags may belong to multiple owners sight (Ilie-Zudor *et al*, 2006; Derakhshan *et al*, 2007). Moreover, vast volume of data that generated in once may cause data lost and get unreliable readings. Multiple readings of the readers at one go can also cause inaccuracy of data. Although there are some simulators in hand, it is expensive and not in web-based. The project problem that we are facing now is lack of RFID data (simulator) in Malaysia for researchers to do testing in RFID supply chain. In Malaysia, some of the manufacturers, distributors, and retailers have already implemented RFID system in their daily operations and management system. All the daily transactions and RFID tags information will be recorded and saved in their own databases. Note that the tags information of each manufacturer, distributor, and retailer is confidential and private. Therefore, there is lack of experiment and testing has been conducted by the researchers in RFID enabled supply chain in Malaysia due to lack of data sharing among the supply chain.

Corresponding Author: Manmeet Mahinderjit Singh, School of Computer Science, Universiti Sains Malaysia.
E-mail: manmeet@cs.usm.my

In addition, RFID cloning and fraud attack is only able to be overcome with countermeasures beyond static preventive mechanism. Since many previous researches focuses on tags preventive models, we argue that detection of cloning and fraud attack is the first defense in eliminating these security attacks. However the issue we tackle here is beyond the effort to minimize the error rate: the percentage of the incorrect prediction of class labels and higher detection accuracy. In real world application, cost is treated unequally and the difference of misclassification cost can be significant. We argue that cost sensitive approach is essential in reducing risk of counterfeiting in a SCM. For example, in medical diagnosis of a certain cancer disease, if the cancer is regarded as the positive class, and non-cancer (healthy) as negative, then missing a cancer (the patient is actually positive but is classified as negative; thus it is also called "false negative" is much more serious (thus expensive) than the false-positive error. The patient could lose his/her life because of the delay in the correct diagnosis and treatment. Similarly, in RFID cloned and fraud detection, false negative or failure to detect fraud tags could be very expensive with counterfeiting issue hitting the market with million dollars of loss. Thus the next motivating factor in this research is none of the previous approaches deals with analysing the cost effects in a system. Eliminating false negative has never been discussed in the RFID based security system before. Consequently, in this paper, we design an RFID-enabled supply chain simulator by applying the concept of Monte Carlo. There are three objectives in the project which are; 1) to generate real-time dataset by simulation process as part of the prototype. This simulation process will auto-generate the dataset in one stage (which include manufacturer, distributor and retailer) or multiple stages from manufacturer tagging process to retailer shelving process for the researchers; 2) to populate real-time dataset by simulation process as part of the prototype. This simulation process can be used by the researchers to generate the dataset separately either in manufacturer, distributor or retailer site and 3) to generate a friction of cloned tags for experimental purposes in detecting counterfeit attack. As a case study, the dataset simulated will be used to detect cloned and fraud RFID tag in a cost-sensitive environment. The RFID tag cloned and fraud process will employ RFID SCM tracking and tracing functions such as tags history attributes event timestamp and time to live (TTL) (Xue *et al*, 2009) as important factors. The outcome of the project is randomly generated and populated real-time dataset. The data set can be used by the researchers to do further testing for cloned tag detection or any data mining. In addition, the simulator can be used by the manufacturers, distributors, and retailers to track and trace the information of the EPC tagged items (when, where and what) in RFID supply chain. Some experimental results of counterfeiting detection using cost-sensitive algorithm such as Metacost bagged with some classification algorithm will also be shown. The remainder of the paper is structured as follows. Section 2 discusses the literature review. Section 3 describes the proposed system – MC Simulator of SCM. Section 4 shows the system evaluation with some test cases. Section 5 presents an example of MC generated RFID dataset used for misclassification counterfeiting detection. Finally, section 6 provides the conclusion.

2 Backgrounds:

In this section, we will discuss the background and some relevant past researches such as the challenges and issues faced in RFID SCM, EPCglobal network, some techniques used to generate RFID dataset, existing RFID simulators, and counterfeiting issues in RFID SCM.

2.1 RFID SCM Data Structure:

From the existing researches, there are lots of benefits brought by the RFID in SCM. For example, it can improve the automation of logistic processes, prevent inventory reduction, and allow inventory renewal (Derakhshan *et al*, 2007; atlas RFID Solutions, 2010). It also saves time and increases the efficiency of workforces (Derakhshan *et al*, 2007; atlas RFID Solutions, 2010). However, there are a few challenges and issues faced in the RFID SCM. Derakhshan *et al* (2007) indicates that one of the challenges in RFID supply chain is the difficulty to track the spatial and temporal of the items as the tagged items will move from one location to another. Besides, there are a large numbers of data will be generated and updated continuously per day (Ahsan *et al*, 2010). Some of the data may be missing or are unreliable readings (Derakhshan *et al*, 2007, Ahmed *et al*, 2009). Multiple readings of the readers at the same time will cause data redundancy during data accumulation as well (Ahsan *et al*, 2010). Next, we will have a discussion on the EPCglobal network in the following subsection.

2.2 EPCglobal Network:

EPCglobal network is a global standard for real-time, automatic identification, and information sharing on the items in the supply chain (Zebra Technologies, 2010; EPCglobal, 2007). Principally it comprises identification system (EPC tags and readers), EPC Middleware, EPC Information Services (EPC IS), and Discovery Services (Zebra Technologies, 2010). The EPC IS lets the partners to share information via a set of service operations and associated EPC-related data standards which are safeguarded (Zebra Technologies, 2010; EPCglobal, 2007). While the Discovery Services enables users to search data and history about an EPC (Zebra Technologies, 2010). Note that the EPC of the item is in cooperation with Object Naming Services (ONS) where

the ONS ties the EPC identification to a location on the Internet, EPC IS (Sikander, 2005). We can know more information of the EPC from the EPC IS as well (Sikander, 2005).

2.3 Simulation Methods:

There are a few methods that can be used to simulate RFID dataset in SCM. For instance, we may generate the data either manually or using some techniques such as SYSRFID tool or Monte Carlo method. Table 1 below shows the comparison of the simulation methods that can be used to simulate RFID dataset in SCM. In SCM, there are a large numbers of data will be generated and updated continuously per day. Although we can generate the data manually, it takes us much time to generate the huge and different datasets. On the other hand, SYSRFID tool uses the client-server architecture to generate a large number of realistic synthetic RFID dataset efficiently (Virgilio, 2011). Users just have to fill in relevant information into the column provided on the web interface of SYSRFID (Virgilio, 2011). Then the system will automatically generate the datasets in a structured file and send it back to the users through an email (Virgilio, 2011). However, it is costly to process all the datasets even though it saves time (Virgilio, 2011). Yet, MC method examines the complete range of risk with each possible risky input (Raychaudhuri, 2008; RiskAMP, 2009). It can be generated by using add-ins to Microsoft Excel has saved the time and cost to simulate RFID dataset (Raychaudhuri, 2008; RiskAMP, 2009).

Table 1: Comparison of simulation methods.

	Manually	SYSRFID	Monte Carlo
Time Consumed	High	Low	Low
Cost	Low	High	Low

Table 2 below demonstrates some of the existing simulators.

Table 2: Existing RFID Simulators.

Name	How it designed	Purpose	Benefits/ Disadvantages
SCM web-based simulator (Jeong et.al, 2009)	- JSP and MS-SQL, with Tomcat 6.0 and Apache server. - Architecture: Supply Chain Model Generator, Simulation Engine and Results DB	Simulate supply chain network for effective decision making	✓ Conduct in any computer connected to internet. ✓ Support multi-layer distribution center design . × No multi-layer factories and retailers. × No track and trace function.
RFID object tracking system	- Eclipse RCP technology and GEF - Simulation	Simulate huge RFID object tracking systems	✓ Assist researching and testing of RFID
Name	**How it designed**	**Purpose**	**Benefits/ Disadvantages**
simulation platform (Bai et.al,2011)	strategy: discrete event scheduling and activity scanning		uncertain data management and space-time information query × No multi-layer processes in SCM simulation.
Test based RFID deployment simulator (Zeng et.al,2008)	Integrate modeling and simulation with test in RFID deployment.	Model virtual devices based on test data such as tags, readers and environmental factors	✓ Improve RFID deployment efficiency × No multi-layer processes in SCM simulator. × No track and trace function in the simulator.
IBM Supply Chain Simulator (Bagchi et.al, 1998)	Combination of graphical process modeling, discrete event simulation, animation, activity based costing and optimization	Help corporations to make business decisions about the design and operation of its supply chain	✓ Save money ✓ Make operational decisions × Not a web-based simulator.

2.4 Counterfeiting Issues:

There is counterfeiting problem happens to RFID SCM due to the lack of security measures and trustworthy among the partners since all the processes and transactions involved are automated (Mahinderjit-Singh and Li, 2010). For instance, RFID tag cloning occur when two products, A and B both are found that they have the same

EPC in a database (Mahinderjit-Singh *et al*,2011). One of them is genuine while the other one must be a cloned tag. Besides, RFID SCM has various network architectures, insecure communication channels of goods movement, different standards among partners, and security key management for tags and readers have increased the risks of RFID tag cloning (Mahinderjit-Singh and Li, 2010 and Mahinderjit-Singh *et al*,2011) . Table 3 shows RFID cloning and fraud attacks (Mahinderjit-Singh *et al*, 2011). There are four attack types: skim, eavesdrop, man in the middle, and physical attack which contribute to the counterfeit issues in RFID SCM. TTL stands for time to live while R/W is read and write value (Xue *et al*,2009). Next , we will discuss our proposed system.

Table 3: RFID cloning and fraud attacks (Mahinderjit-Singh *et al*, 2011).

Attack Types	Attack Pattern	Attack Levels	Model features
Skim	Cloned	Low(Tag, Reader)	Content Timestamp/TTL R/W on Tag & Reader
Eavesdrop	Cloned	Low(Tag, Reader, DB)	Content Timestamp/TTL R/W on Tag & Reader Location
Man-In-The Middle	Cloned Fraud	High (Tag,Reader, DB)	Content Timestamp/TTL R/W on Tag & Reader Location
Physical	Cloned Fraud	High (Tag,Reader, DB)	Content Timestamp/TTL R/W on Tag & Reader Location

3.0 Proposed System - MC Simulator of SCM:

A supply chain may involve a large numbers of manufacturers, distributors and retailers. As an item moves from a manufacturer to a retailer in the supply chain, location, timestamp and other characteristics of the item may be changed. Thus, the item tracking process in the supply chain is complex. In this section, we will discuss on the RFID data structure, physical and logical frameworks of the system, and make a few assumptions before starting to design the MC simulator.

3.1 RFID Data Structure

Based on the Table 1, the Monte Carlo (RiskAMP, 2009) method can be generated by using add-ins to Microsoft Excel has provided a more suitable, less cost and less time consumed method for us to simulate RFID dataset. Therefore, we would like to use the MC method to generate real-time dataset for the MC simulator. Basically the MC algorithm execution is shown in Figure 1 below. The simulation will be run for 1000 times randomly (RiskAMP, 2009).

```
int N = 1000; // repetition for 1000 times

Random rand = new Random(); // random function
```

Fig. 1: Monte Carlo algorithm.

After the dataset has been generated, it can also be preprocessed so that the preprocessed dataset can be used by the researchers to do further testing for cloned tag detection and data mining testing. The preprocessed dataset is calculated based on the mean, standard deviation, and read and write (R/W) value from TTLi to TTLS for each site. The notion TTL will be discussed further in section 5.0. Next, we will look into the project assumptions.

3.2 Project Assumptions:

There are a few assumptions have been made in the project before starting to design the MC simulator, which are:

- The project will only focus on RFID item tracking in SCM.
- All the RFID tags that we use in SCM are passive tags.
- There are multiple links in the supply chain.
- Each location has only one administrator and two workers.
- The product that we are going to use in the supply chain is under beverages industry.
- The product may be packed in cases, boxes or pallets.

- EPC of an item is unique.
- Manufacturer, distributors and retailers are operated 24/7 while truck company operates from 8am to 10pm.
- The entire tracking process of the product will be started from the tagging process in the manufacturer factory until it was put on the shelf in the retailer shop.
- The readers are mounted on the wall in manufacturer factory, distributor warehouses, retailer shops, and trucks. EPC of the items will be read by the readers randomly.
- The trucks which are used for shipping purpose are belonged to a third party.

3.3 Framework for Web-enabled Supply Chain Simulator:

Supply chain multi-link involves three partners: manufacturer (M), distributor (D) and retailer (R). In the manufacturer factory, there are four processes: tagging, packing, loading, and shipping. In the distributor warehouse, there are six processes: unloading, unpacking, in storage, packing, loading, and shipping. In the retailer shop, there are four processes: unloading, unpacking, in storage, and shelving. An item, Coca-Cola will be shipped by a truck from the manufacturer factory to a distributor warehouse for storage. Then, it will be sent by a truck to a retailer shop for shelving. The entire tracking process of the Coca-Cola will be started from the tagging process in the manufacturer factory until it was put on the shelf in the retailer shop.

4.0 Evaluation Of MC System:

The system is developed by using Java Server Pages (JSP) and Oracle database, with Apache Tomcat 7.0.27 server. There is also an existing data mining tool, Weka has been integrated into the system as well. The user can input the preprocessed dataset into the Weka for data classification and cloned tag detection.

4.4.1 Scenario 1: Simulation of raw data with simulation model:

A user wishes to simulate data for multiple links in supply chain with simulation model. Figure 4 below shows the simulation model of the system. User has to fill in the relevant information into the columns provided on the web system and click the “Start Simulation” button. Then the system will automatically generate the dataset and output it as csv file. Note that there is an option which allows the user to inject the cloned tags into the dataset. User can define the percentage and starting process of the cloned tags as well. After the dataset has been generated, it can be sent for preprocessing. The preprocessed dataset can then be feed into the WEKA (http://www.cs.waikato.ac.nz/~ml/weka) for data classification.

4.4.2 Scenario 2: Populate data in manufacturer, distributor or retailer side:

The SCM dataset population process can be done separately in manufacturer, distributor or retailer entity. Assume a user wishes to populate data in manufacturer site. The user can select the processes to be populated in the manufacturer side and destination for the shipping items in this page. User can decide on to populate the dataset from tagging process up to tagging, packing, loading or shipping process. After user has filled in all the relevant information, he clicks on the “Run” button. The system will then automatically populate the dataset for the manufacturer site only.

4.4.3 Scenario 3: Tracking of spatial and temporal of an EPC tagged item:

A user wants to track the information of an EPC tagged item, 10.10.1.3191 (when, where, what). Figure 5 below shows the track module in the system. By entering the EPC of the item, the system is able to track the exact location, current process, and attack status of that particular item.

5.0 Rfid Supply Chain Simulation and Evaluation:

The strength of any RFID application is fully capitalised when the temporal and location information are correctly utilised in eliminating data security issue in RFID. Real time monitoring of events such as fraud and cloning attacks in RFID application are still rare.

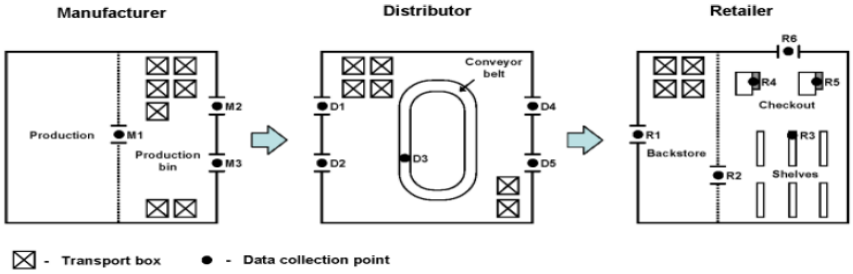

Fig. 2: Supply Chain Management (SCM).

Figure 2 shows a typical SCM environment with three different sites (Manufacturer, Distributor and Retailer). RFID tags are attached to the products for instance wine bottles. RFID based supply chain system involves the movement and flows of millions of data. The data generated consists of RFID tuples of the form of (EPC, location, time), where *EPC* is the unique identifier read by an RFID reader, *location* is the place where the RFID reader scanned the item, and *time* is the time when the reading took place. Tuples are usually stored according to a time sequence. Each sites will have their own database system and this distributed manner database system are combined with a centralized EPC global server; EPC- Information Server. (EPC-IS). The tracking and monitoring system can even play role as an intrusion detection system by using events rules and triggers function in database. Among the rules are as below:

- If, for instance, a product was identified at specific read points, e.g., „shelf‟ (R3) and then „exit‟ (R6),without having first been identified at the read point „checkout‟ (R4 or R5), then it could be a matter of cloned or fraud.
- If a pallet P, which is containing the objects O1, O2, and O3 when leaving the production facility (M2 or M3) was identified as having only the objects O1 and O3 at the distributors receiving dock (D1 or D2), then the object O2 could have been replaced with O4 during transportation. These mean counterfeit products are injected.

TTL indicates the time restriction that targets events should satisfy. Since most RFID application has a restriction time, we believe if carefully defined, we can use the notion of TTL to detect clones and fraud tags in a typical SCM. Based on TTL taxonomy (Xue *et al*, 2009), there are 4 different notions of TTL given based on the event types, both primitives and complex categorised based on events as Absolute TTL (TTL*a*), Relative TTL (TTL*r*), Periodic TTL (TTL*p*) and Sequential TTL (TTL*sE*). The detection process of cloned and fraud tags are able to manipulate all the above TTL notions. However, based on RFID applications, we determine that three relevant TTL notion for a SCM transactions and monitoring process is mainly *TTLa*, *TTLr* and *TTLs.* We also argue that the absolute *TTL (TTLa)* notion can be further categorised based on RFID applications. Some applications such as drugs and fast moving products for e.g. diary and foodstuff requires restriction in expiry date as the *TTLa* compare to product such as wine and jewellery. These expensive products emphasize more on manufacturing time. *TTLi* specifies the period of time a RFID tag is tagged on the product. By tracking, monitoring and storing the *TTLi* in the system; we are able to classify cloned RFID tags from genuine tags. Our simulated dataset employ all the above TTL notions and satify each condition mentioned. Next we will provide a survey on various costs sensitive methods.

5.1 Cost Sensitive Learning Methods:

Cost-Sensitive Learning is a type of learning in data mining that takes the misclassification costs (and possibly other types of cost) into consideration. The goal of this type of learning is to minimize the total costs (Turney,2000). Many works for dealing with different misclassification costs have been done, and they can be categorized into two groups. One is to design cost sensitive learning algorithms directly (Turney, 1999; Drummond and Holte, 2000). The other is to design a wrapper that converts existing cost-insensitive base learning algorithms into cost-sensitive ones. The wrapper methods are also called cost-sensitive meta-learning (Witten and Frank, 2005; Domingos, 2006) ,sampling (Zadrozny *et al*, 2003), and weighting (Ting,1998). Cost-sensitive meta-learning converts existing cost insensitive base learning algorithms into cost-sensitive ones without modifying them. Cost-sensitive meta-learning techniques can be classified into two main categories, *sampling* and *nonsampling*,in terms of whether the distribution of training data is modified or not according to the misclassification costs. This paper focuses on the nonsampling cost-sensitive meta-learning approaches. The non-sampling approaches can be further classified into three subcategories: relabeling, weighting, and threshold adjusting. The first is *relabeling* the classes of instances, by applying the minimum expected cost criterion (Witten and Frank, 2005). *Relabeling* can be further divided into two branches: relabeling the training instances (Witten and Frank, 2005) and relabeling the test instances (Domingos, 1999).

Table 4: An example of cost matrix for binary classification.

	Actual negative	Actual positive
Predict negative	C (0,0) , or TN	C(0,1), or FN
Predict positive	C(1,0), or FP	C(1,1) , or TP

Note that *C(i, i)* (*TP* and *TN*) is usually regarded as the “benefit” (i.e., negated cost) when an instance is predicted correctly. This is the minimum expected cost principle. In Relabeling approach such as Metacost (Witten and Frank, 2005) and Cost Sensitive Classifier (Domingos, 1999), cost C is known at the learning time.

5.2 Result of Experiments:

Applying the dataset from the simulated RFID supply chain from our proposed web-enabled RFID simulator, 3000 example of RFID traces are generated from manufacturer site up to retailer site. RFID traces is then pre-processed into audit dataset which includes attributes such as Tags ID, location ID, TTLs (mean), TTLt (mean) , TTLsE (mean and standard deviation) and Read/write (mean and standard deviation). The datasets are then feed into Weka engine by applying Metacost algorithm. The audit data will then be feed into a filtering system upfront for normalization purposes. CfsSubsetEval with Best First technique are used to determine the evaluation of attributes and search methods. The base classifiers used were Naive Bayes, Random Forest and Weka's implementation of a Support Vector Machine (SMO), JRIP and C4.5 (J48) decision tree. Default Weka options were used for the Naive Bayes , Random Forest and JRIP but for the SMO "build logistic models" was set to true and for the J48 tree "Pruning" was disabled. The standard cost-sensitive classifier was used for Naïve Bayes, SMO and Random Forest. There are two main goals of the classification experiments - to find the most robust and versatile classifier for imbalanced RFID dataset and to find out the optimal misclassification cost setting for a classifier. The engine is trained with a training dataset. Cloning attacks such as skimming, eavesdropping and man-in the middle are simulated. To train the models cross-validation was employed. Cross-validation is a standard statistical technique where the training and validation data set is split into several parts of equal size, for example 10% of the compounds for a 10 fold cross validation. An independent test dataset is simulated as well. However, for the differing classifiers they have used across-the-board costs of 20, 40, 60, 80,100, 200, 500, 1000, 2000 and 10000. Weka normalises (reweights) the cost matrix to ensure that the sum of the costs equals the total amount of instances. Receiver Operating Characteristic (ROC) curve is a plot of the probability of true positive (recall) as a function of the probability of false alarm across all threshold settings. An ROC curve provides an intuitive way to evaluate the classification performance of RFID detection system. Recall represents the probability of detection of cloned tags and precision is the proportion of the correctly predicted genuine tags in each prediction class. In this study, we will utilize ROC, recall and precision for models evaluation.

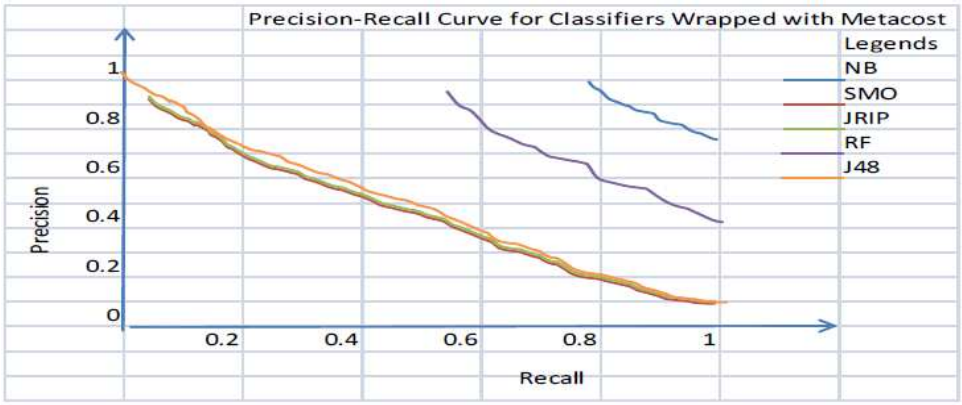

Fig. 4: Precision-Recall Curve for Classifiers wrapped with MetaCost.

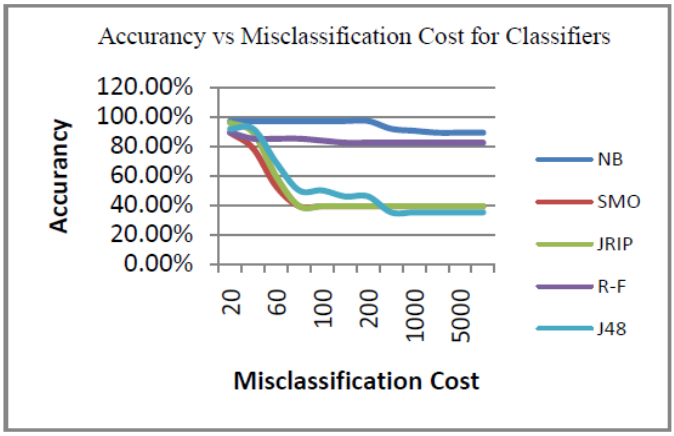

Fig. 5: Accuracy vs. Misclassification Cost for Classifiers Table 5: SMO (SVM) classifier TP, FP, Precision-Recall and Accuracy rate based on Cost ratios.

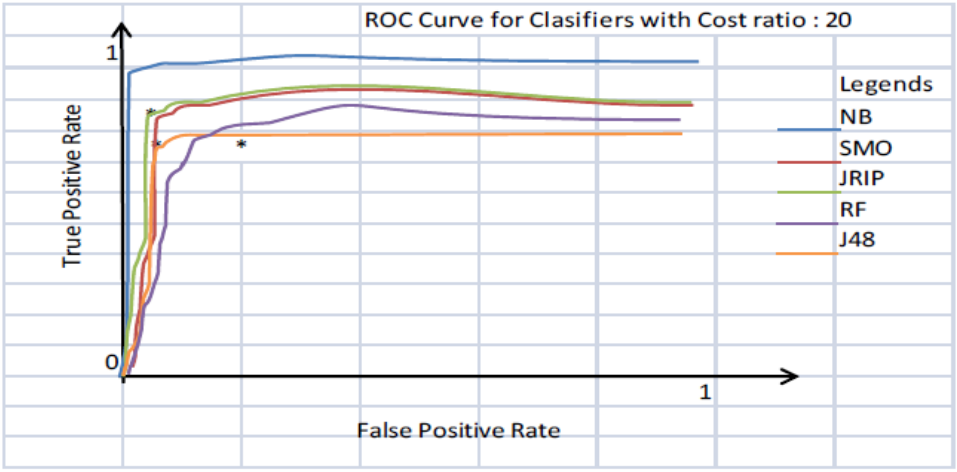

Fig. 6: ROC Curve for Classifiers wrapped with MetaCost (Cost ratio: 20).

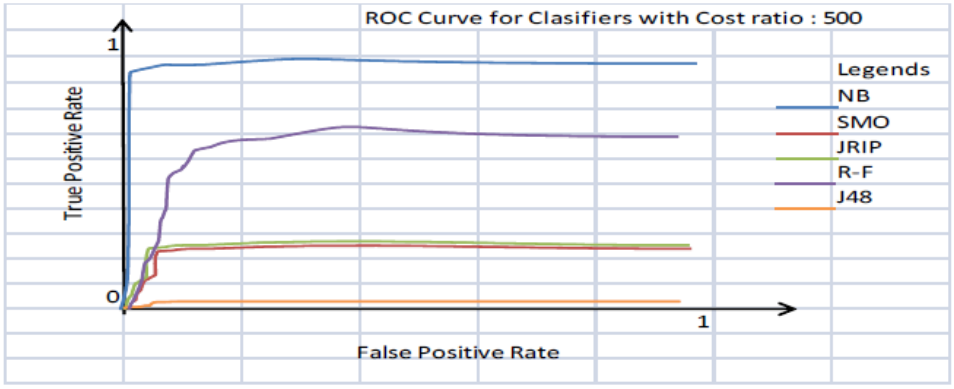

Fig. 7: ROC Curve for Classifiers wrapped with MetaCost (Cost ratio: 500).

5.3 Discussions:

In cloning detection of RFID enabled supply chain, misclassifying cloned tag as genuine is undesirable and cost is high. Our experimental result show that when increasing *cost-ratio* from 20 to 10,000, the recall rate would increase. Although not unexpected, is the decrease of *precision* which implies needless analysis of large number false positives (shown in Fig.4) SMO, JRIP and J48 algorithms consistently reach *Recall* rates close to 1 at high cost ratios, with precision slightly above 0.1. Based on Fig 5, we can conclude that as cost ratio increases, the accuracy of classifier decreases as well. Performance measurements are analysed for every cost ratios as shown in Table 5. Naïve Bayes is quite different from the other four classifiers. Its performance indices (*TP rate, FP rate*) are rather constant regardless of the cost (Fig 6 – Fig 7). An important implication from this study is that we can use cost to choose suitable operational threshold (based on different *cost-ratio*) to control a classifier"s performance. In this study, four classifiers except Naïve Bayes provide this flexibility. In practice, exact costs are rarely known and could change as we learn more about system requirements, its design, operational environment, etc. When considering a wide range of cost ratios the resulting models differ significantly. For instance from Fig 7, J48 classifier is made cost sensitive when the cost ratio was set to be 500 with accuracy of 35.1%. This means that FN needs to be 500 times more expensive than FP for J48 to transform to cost sensitive. Overall, J48 provides the most robust and versatile classifier for imbalanced RFID dataset compared to other classifiers. With respect to construct validity, cost ratios in our experiments, which vary from 20 to 10,000 might not include all meaningful cost differentials. Different intrusion detection systems may have their own cost ranges of interests. The selection of classifiers is another possible source of bias. We cannot exclude the possibility that a classifier not studied here could show significantly better performance. Nevertheless, based, we believe that the chance of such a classification algorithm being in existence is rather low.The results above could be implicated by the small datasets used in the training models. When small dataset are used, classifier cannot accurately estimate the class membership probabilities and the imbalanced in class distribution of the dataset. Any RFID cloned detection classifiers used must be correlated with cost since lower cost properties projects to lower or zero cloned tags in the system. This also impact positively in reducing the counterfeit attack which risks billions of dollars losses yearly in the market.

6.0 Conclusion:

In conclusion, this simulator can be used by the researchers who are interested in using the dataset for Intrusion Detection System or any data mining. The significances of the simulator are to enhance the automation of logistic operations, faster tracking and tracing the history of the items, and optimize process in RFID SCM. Lastly, the simulator are improvised to add more security attack, displayed cloned tag in image, and will be plugged into a real-life implementation in the RFID supply chain in the future. In this paper we have showed how RFID tag fraud and cloning attacks in a supply chain management can be detected using classification algorithms. We have shown that when relabeling approach is used, we are able to reduce the misclassification cost and eliminate the scenario of having cloned and fraud tags in the system.

ACKNOWLEDGEMENT

The authors are grateful to the School of Computer Sciences, USM for providing the facilities to carry out the research.

REFERENCES

Ahmed, N. and U. Ramachandran, 2012. Load Shedding Based Resource Management Techniques for RFID Data. Retrieved 17 September, 2012: http://www.cc.gatech.edu/projects/up/publications/RFID_CR_ Final.pdf

Ahsan, K., H. Shah and P. Kingston, RFID Applications: An Introductory and Exploratory Study. *IJCSI International Journal of Computer Science Issues*, 7(1): 1-7.

Bagchi, S., S.J. Buckley, M. Ettl and G.Y. Lin, 1998. Experience using the IBM supply chain simulator. In *Proceedings of the 1998 Winter Simulation Conference*, (Washington, DC, December 13 – 16, 1998), 1387-1394.

Bai, Q., T. Zhang, Y. Yin and G. Yu, 2011. Research and implementation of an RFID object tracking system simulation platform. In *2011 Chinese Control and Decision Conference (CCDC)*, (Mianyang, May 23-25, 2011), 4007-4012.

Derakhshan, R., M.E. Orlowska and X. Li, 2007. RFID Data Management: Challenges and Opportunities. In *2007 IEEE International Conference on RFID*, (Grapevine, TX, USA, March 26 – 28, 2007), 175-182.

Domingos, P., 1999. MetaCost: A general method for making classifiers cost-sensitive. In Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining, 155-164, ACM Press.

Drummond, C. and R. Holte, 2000. Exploiting the cost (in)sensitivity of decision tree splitting criteria. In Proceedings of the 17th International Conference on Machine Learning, 239-246.

Electronic Product Code (EPC) RFID Technology, 2010. Retrieved 22 September, 2012, from Zebra Technologies:http://www.zebra.com/us/en/solutions/getting-started/rfid-printing-encoding/epc-rfid-technology. html.

EPCglobal: Overview, 2007 Retrieved 22 September, 2012, from EPCglobal: http://www.gs1sy.org/ GS1System/EPCglobal/overview.htm.

Ilie-Zudor, E., Z. Kemeny, P. Egri and L. Monostori, 2006. The RFID Technology and its Current Applications. In *proceedings of The Modern Information Technology in the Innovation Processes of the Industrial Enterprises-MITIP* 2006, (Budapest, Hungary, January, 2006), 29-36.

Jeong, D., M. Seo and Y. Seo, 2009. Development of web-based simulator for supply chain management. In *Proceedings of the 2009 Winter Simulation Conference*, (Austin, TX, December 13 - 16, 2009), 2303-2309.

Mahinderjit-Singh, M. and X. Li, 2010. Trust in RFID-enabled supply-chain management. *Int. J. Security and Networks*. 5, 2/3 (March 2010), 96-105. DOI= http://dx.doi.org/10.1504/IJSN.2010.032208.

Mahinderjit-Singh, M., X. Li and Z. Li, 2011. A Cost-based Model for Risk Management in RFID-Enabled Supply Chain Applications. in Pengzhong, L. ed. *Supply Chain Management*, InTech, 201-225.

Raychaudhuri, S., 2008. Introduction to Monte Carlo Simulation. In *Proceedings of the 2008 Winter Simulation Conference*, (2008), 91-100. What is Monte Carlo Simulation, 2009. Retrieved 13 October, 2012, from RiskAMP: http://www.thumbstacks.com/files/RiskAMP%20-%20Monte%20Carlo%20Simulation.pdf.

RFID vs Barcode?, 2010. Retrieved 8 October, 2012, from atlas RFID Solutions: http://www.atlasrfid.com/Technology/RFIDvsBarcode.aspx

Sikander, J., RFID Enabled Retail Supply Chain. Retrieved 17 September, 2012, from MSDN: http://msdn.microsoft.com/en-us/library/ms954628.aspx.

Springer-Verlag. Turney, P.D., 1995. Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm. Journal of Artificial Intelligence Research, 2: 369-409.

Ting, K.M., 1998. Inducing Cost-Sensitive Trees via Instance Weighting. In Proceedings of the Second European Symposium on Principles of Data Mining and Knowledge Discovery, pp: 23-26.

Traub, K., *et al*., 2012. The EPCglobal Architecture Framework. Retrieved 22 September, 2012, from EPCglobal: http://www.gs1.org/gsmp/kc/epcglobal/architecture/architecture_1_4-framework-20101215.pdf

Turney, P.D., 2000. Types of cost in inductive concept learning. In Proceedings of the Workshopon Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning, Stanford University, California pp: 15-21.

Virgilio, R.D., 2012. SysRFID: generation of synthetic data in Supply Chains. Retrieved 13 October, 2012: http://www.cersi.it/itais2011/pdf/25.pdf

Weka Machine Learning Project, "Weka," URL :http://www.cs.waikato.ac.nz/~ml/weka

Witten, I.H. and E. Frank, 2005. Data Mining - Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers.

Xue Li, Jing Liu, Quan Z. Sheng, Sherali Zeadally and Weicai Zhong, 2009. TMS RFID: Temporal Management of Large-Scale RFID Applications, International Journal of Information Systems Frontiers, Springer, July. 2009 pp: 1-20.

Yonghui, W., X. Jingke and W. Shoujin, RFID Spatio-Temporal Data Management. *TELKOMNIKA*, *11*(3): 1348-1354.

Zadrozny, B., J. Langford and N. Abe, 2003. Cost-sensitive learning by Cost-Proportionate instance Weighting. In Proceedings of the 3th International Conference on Data Mining pp: 435.

Zeng, J., Y. Liu, C. Liu and D. Li, 2008. Research on Test Based RFID Deployment Simulator. In *Third 2008 International Conference on Convergence and Hybrid Information Technology*, (Busan, November 11-13, 2008), 1142-1146.