

- 1 -

**Minutes of the meeting (6) of the Project Review & Steering Group (PRSG) for the Project "Network Intrusion Detection Based on Offline and Online Data Mining Techniques"**

1.0 The meeting (6) of the Project Review & Steering Group (PRSG) for the project "Network Intrusion Detection Based on Offline and Online Data Mining Techniques" was held on 29<sup>th</sup> April, 2011 at IIT Guwahati.

The list of persons who attended the meeting is at *Annexure I*.

- 1.1 The Member-convenor of the PRSG welcomed the PRSG members and other participants. He briefly explained the deliberations and recommendations of the PRSG in meeting (5). He informed the Committee that the approved duration of the project was ending on April 30, 2011.
  - 1.2 Prof. Sukumar Nandi, Chairman of the PRSG welcomed the members of the PRSG and other participants. In his opening remarks, he mentioned that project efforts are expected to benefit the project team in capacity building in the area of application of data mining approaches for network intrusion detection. He suggested the team to explain the details of the achievements in the project vis-à-vis the project objectives.
  - 1.3 Minutes of the meeting (5) of the PRSG held on 12<sup>th</sup> November 2010 were circulated to the members. Since no comments were received from the members, the minutes of the meeting (5) were confirmed.
- 2.0 Prof. D.K. Bhattacharyya, the Chief Investigator of the project and his team made a presentation explaining the objectives of the project, achievements in the project, action taken on the recommendations of PRSG in meeting (5), details of capital equipment procured and funds released/utilized, etc

**Deliberations**

**2.1 Technical**

**2.1.1 Objectives and achievements**

- The team explained that the objectives of the project was to explore the potential usage of data mining approaches (clustering and association rule mining) for intrusion detection with a goal to detect known and unknown attacks and to reduce the false alarms.
- The team informed that as per the objectives of the project, they have carried out technical survey and developed (a) offline anomaly detection modules using (i) an incremental (supervised and unsupervised) subspace clustering and (ii) an outlier based semi-supervised method, capable of handling both known as well as unknown attacks with minimum false alarms, (b) feature extraction modules for both packet and flow level



network traffic data and (c) online (near real time with 2 minutes window) anomaly detection module using an incremental (supervised ) subspace clustering and feature extraction. They have explored application of association rule mining for misuse detection system. Validation has been done using standard datasets as well as dataset created utilising the test bed. A number of publications have been brought out based on the project efforts.

### 2.1.2 Details of project efforts

- The team presented the details of the achievements in the project. They first presented the details of technical survey carried out by them.

#### (A) Technical Survey

##### **Survey on Port Scans and Their Detection Methodologies**

An attacker performs port scans of IP addresses to find vulnerable hosts to compromise. A survey on the existing Port Scan attack detection techniques was carried out. It includes the common port scan attacks, a general comparison based on its types, mode of detection, mechanism used for detection, etc. This survey reports the various standard datasets and some evaluation criteria for the detection approaches.

*The work was published in The Computer Journal , Oxford University Press DOI: 10.1093/comjnl/bxr035 , 2011.*

##### **Survey on Outlier Detection Methods in Network Anomaly Identification**

Outlier detection is an important anomaly detection approach. A survey on various distance-based, density-based and soft computing based outlier definitions, outlier detection methods and formulae for anomaly score was carried out. Also, the survey reports the various supervised and unsupervised anomaly detection methods using outlier approach. A general comparison among the various anomaly score and outlier detection methods are also reported.

*The work was published in The Computer Journal , Oxford University Press, Vol 54, no 4, pp 570-588, 2011*

##### **Anomaly Detection Analysis of Intrusion Data using Supervised & Unsupervised Approach**

A survey was carried out on various types of anomalies, types of input data and various proximity measures used for anomaly detection, various anomaly reporting techniques(supervised, semi-supervised and unsupervised) etc. Further, various approaches used for anomaly detection such as, Statistical, Bayesian Network, ANN, SVM, KNN, Knowledge based, GA Based, HMM based, Fuzzy Theoretic, and Clustering were also brought out.

The work was published in *International Journal of Convergence Information Technology (JCIT)* Volume 5, Number 1, February 2010.

### **Incremental Approaches for Network Anomaly Detection: Existing Solutions and Challenges,**

This survey presented incremental approaches for detecting anomalies in normal system or network traffic. The technological trends, open problems, and challenges over anomaly detection using incremental approach were also included.

The work has been submitted for publication in the *WIREs Review : Data Mining and Knowledge Discovery* (John Wiley & Sons).

### **(B) Dataset Generation**

The team explained their efforts to generate datasets utilising the test bed created. These datasets are used for testing of the algorithms developed by them.

#### **Packet-level network traffic capturing, pre-processing, feature extraction and dataset preparation**

To capture and prepare their own dataset, they used the *Nmap* tool to launch the various types of attacks (viz. SYN, ACK, FIN, XMAS, NULL, MAIMON) using test bed created. To accomplish its goal, using *Nmap*, they sent specially crafted packets to the target host and then analyzed the responses. They captured the normal as well as attack traffic using *Wireshark*. After necessary pre-processing/formatting the raw traffic data, they extracted various *basic, content-based, time-based* and *connection-oriented* features using *TCPtrace* tool and *c/perl* routines. After extraction of all the relevant features for port scan attack detection, they formatted the total feature dataset, which was used as a training dataset for anomaly detection.

#### **Flow-level network traffic capturing, pre-processing, feature extraction and dataset preparation**

To capture and prepare flow-level dataset, they used the *NFDUMP* and *NFSEN* tool to capture and visualize the flow level network traffic in the test bed. They considered a set of 14 features for analysis.

### **(C) Development of Algorithms, Validation and Testing**

The chief investigator informed the Committee that as per the project objectives, they have achieved the following works on offline and online anomaly detection modules based on incremental subspace clustering (both supervised and unsupervised) and outlier based method (semi-supervised). Also, the team explained their efforts on evaluation of association rule mining in misuse detection system.



- 6 -

### **Supervised Anomaly Detection using Clustering-based Normal Behavior Modeling**

This work deals with a clustering based supervised anomaly detection technique. A set of training data consisting of normal data only are divided into clusters which are represented by their profiles to form the normality model. Any deviation from the normality model is treated as attack. Methods for clustering, training and detection are provided. This technique was validated with KDD CUP 1999 datasets. Performance measuring methods like Recall, Precision, and  $F_1$  measure for good clustering are applied.

*The work was published in the International Journal of Advances in Engineering Sciences Vol 1, No ,1 pp 12-17, (2011).*

### **Supervised Classification of Network Anomalies using Clustering**

This work deals with a clustering based classification technique and application in network anomaly detection. A set of labeled training data consisting of normal and attack instances are divided into clusters which are represented by their representative profiles consisting of attribute-value pairs for selected subset of attributes. Each category of attack and normal instances are broken down into a set of clusters using a training algorithm based on a combination of unsupervised and supervised incremental clustering algorithms. The cluster profiles together with their class label form rules for labeling unseen testing instances. Methods for clustering, training and prediction are provided. Evaluation results on KDD' 1999 datasets for two-class (normal and attack), five class (normal, DoS, R2L, U2R and Probe) and all-attacks (normal and 37 attacks) showed good performance. The proposed method was also evaluated based on NSL-KDD dataset and their own dataset.

*The work was submitted for publication in the International Journal of Expert System with Application (Manuscript No. ESWA-D-11-00552, Elsevier).*

### **RODD: An Effective Reference-based Outlier Detection Technique for Large Datasets**

This work deals with an effective reference point based outlier detection technique (RODD) which performs satisfactorily in high dimensional real-world datasets. The technique was evaluated in terms of detection rate and false positive rate over synthetic and real-world datasets and the performance was found to be satisfactory.

*The work was published as Book Chapter in Advanced Computing, LNCS-CCIS (Springer-Verlag, Berlin Heidelberg), 2011, Volume 133, Part 1, 76-84, DOI: 10.1007/978-3-642-17881-8\_8*

### **NADO: Network Anomaly Detection using Outlier Based Semi-supervised Approach**

This work deals with an outlier based method for network anomaly detection. The method performs well over network intrusion datasets while comparing with existing algorithms. The performance of the method was evaluated with

5

KDDcup99 intrusion dataset and other real life datasets. It was observed that NADO has high detection rate and low false positive rate.

*The work was published in the Proc. of ACM ICCCS'2011, pp 531-536, February 12-14, 2011, India (ACM 978-1-4503-0464-1)*

### **Semi-supervised Outlier based Scheme for Network Anomaly Detection**

This work deals with an outlier based scheme for network anomaly detection that established the effectiveness over high dimensional real-world datasets and also KDDcup99 datasets. It selects relevant features from high dimensional datasets and used during cluster formation as well as during calculation of outlier score for network anomaly detection. It performs satisfactorily in high dimensional real-world datasets. The algorithms were evaluated in terms of detection rate, false positive rate, precision, recall, and F-measure over synthetic and real-world datasets and the performance was found to be satisfactory.

*The work has been submitted for publication in The Computer Journal (Oxford University Press)*

### **Unsupervised Anomaly Detection Using Incremental Subspace Clustering**

This work deals with an unsupervised anomaly detection technique using a subspace based incremental clustering algorithm. The clustering algorithm works for datasets with mixture of categorical and numeric attributes. For detecting anomalies based on the clustering result, they propose a model for behavior of attack and normal instances on the basis of sequences of connection records. This technique was validated with KDD CUP 1999 datasets.

*The work was published in the Proc. of NWNS'2010, pp 36-46, June, 2010.*

### **Evaluation of Association Rule Mining in Misuse Detection System**

Association Rule Mining (ARM) is a process of finding some relations among the attributes/attribute values of a huge database. Inside the huge collection of data stored in a database, some kind of associations/relationships among the various attributes may exist.

In this work, a multi-objective rule generation based on non-dominating sorting technique is presented to generate both the frequent and rare rules. It uses measures like *support count*, *comprehensibility* and *interestingness* as different objectives in evaluating a rule. The technique was validated over several real-life UCI ML repository datasets and has been found with a high classification accuracy. However, in the context of high dimensional larger data classification (like intrusion data), the performance has not been found always satisfactory.

### **(D) Integration of Dataset Generation and Supervised CatSub based Anomaly Detection modules**

It captures the packet level/ flow level live-network traffic data using their own test bed and then pre-processed and send into the feature extraction module for the extraction of various basic, content-based, connection based and time based features. Once the



6 ←

features are extracted, the feature data are fed to the supervised network anomaly detection module for testing in near real time (i.e., 2-minutes window). It reports alarm, if any deviation/non-conformity with respect to the defined normal behaviour is detected.

### **(E) Validation and Testing of modules with Datasets obtained from Other Agencies**

Validation and testing of the both offline and online modules were carried out using NSL-KDD, KDD99 and DARPA1998 Benchmark Datasets, data sets provided by CDAC Bangalore and datasets created utilising test bed and evaluated their performance in terms of FPR, DR, and F-measure.

### **(F) Evaluation of the proposed modules in comparison with the other data mining approaches**

The performance of the supervised, semi-supervised and unsupervised modules developed were evaluated in comparison to the known algorithms like C4.5, Bayesian Network (BN2), CART and ID3 and the results have been found satisfactory in terms of FPR, DR, and F-measure.

## **2.1.3 Funds utilisation**

- The total estimated cost of the project is Rs 47.69 lakhs which is in the form of grant-in-aid from DIT. A total amount of Rs 43.87 lakhs was released by DIT to the project in two instalments.
- The team informed that they utilised all the funds released to the project and there is no unspent balance. Any unspent balance at the closing of accounts will be returned to DIT.

## **2.1.4 Others**

### **2.1.4.1 Capital items**

The team presented the details of the capital items procured for the project. One servers, two workstations, ten PCs, one L3 Switch, two L2 switches, one Router, one UPS, one set mobile test bed (controller, access point and antenna) and accessories have been procured under the project. The team requested the Committee to recommend the retention of the capital items procured by them for continuation of the efforts.

### **2.1.4.2 Manpower**

The team informed the Committee that the manpower (02 JRAs) recruited were working in the project. The Committee took note of this.

## 2.2 Recommendations

The PRSG interacted closely with the project team both on technical and financial aspects related to the project and made the following recommendations:

- The Committee deliberated on the achievements in the project and felt that the team has developed expertise in the area of application of data mining approaches for network intrusion detection by way of developing specific clustering and association rule mining approaches, validating with both standard datasets as well as the dataset created by them. The team made additional efforts for creation of their own dataset using the test bed established. The PRSG considered that the project achievements were in line with the objectives of the project and recommended to DIT that the project may be closed formally. The Committee appreciated the efforts of the team.
- The Committee took note of the draft Project Completion report circulated by the team and suggested the team to update the same incorporating the summary of the experimentation carried out and the outcome and other aspects suggested by the Committee and submit it along with the necessary enclosures (refund of unspent balance and utilisation certificate etc) to DIT for formal closure of the project.
- The Committee recommended that the team may be permitted to retain the capital items procured in the project to continue the efforts.
- The Committee recommended continuation of the experimentation by the team and possible improvements in the data mining approaches developed.

The meeting ended with the vote of thanks to the Chair.



Form GFR 19-A

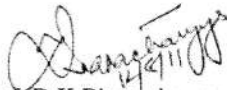
FORM OF UTILIZATION CERTIFICATE

Sl. No.	Letter No. & Date	Amount Rs.	Certified that out of the amount of <b>Rs. 3,76,000</b> received during the year <b>2010-11</b> in favour of <b>Dept. of CS&amp;E, Tezpur University</b> , sanctioned to <b>Registrar, Tezpur University</b> and <b>Rs. 6,80,041</b> on account of unspent balance of the previous year, a sum of <b>Rs. 8,37,669</b> has been utilized for the purpose for which it was sanctioned and that the balance of <b>Rs. 2,18,372</b> remaining unutilized so far will be spent within the current financial year <b>2011-2012</b> .
1	Approv. No 12 (9)/08-ESD dated 16-10-2008	40.11 Lakhs	
2	No. 12 (9)/2008-ESD(vol. II) dated 31-01-2011	3.76 Lakhs	
	<b>Total</b>	<b>43.87 Lakhs</b>	

2. Certified that I have satisfied myself that the conditions on which the amount was sanctioned have been duly fulfilled and that I have exercised required checks to see that the money was actually utilized for the purpose for which it was sanctioned.

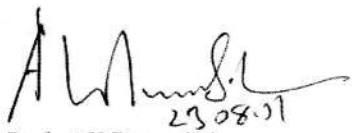
Kinds of Checks exercised

1. Ledger Book
2. Cash Book
3. Purchase orders
4. Stock Book Entry

  
Prof. D K Bhattacharyya  
Principal Investigator

**Professor**  
Department of Computer Science & Engg.  
Tezpur University

  
R R Borah  
Finance Officer  
Tezpur University  
TEZPUR UNIVERSITY

  
Prof. A K Buragohain  
Registrar  
Tezpur University  
Registrar  
Tezpur University





Form GFR 19-A


FORM OF UTILIZATION CERTIFICATE

Sl. No.	Letter No. & Date	Amount Rs.	Certified that out of the amount of Nil received during the year 2011-12 in favour of Dept. of CS&E, Tezpur University, sanctioned to Registrar, Tezpur University and Rs. 2,18,372 on account of unspent balance of the previous year, a sum of Rs. 2,18,372 has been utilized for the purpose for which it was sanctioned and that the balance of Nil remaining unutilized.
1	Approv. No 12 (9)/08-ESD dated 16-10-2008	40.11 Lakhs	
2	No. 12 (9)/2008-ESD(vol. II) dated 31-01-2011	3.76 Lakhs	
	<b>Total</b>	<b>43.87 Lakhs</b>	

2. Certified that I have satisfied myself that the conditions on which the amount was sanctioned have been duly fulfilled and that I have exercised required checks to see that the money was actually utilized for the purpose for which it was sanctioned.

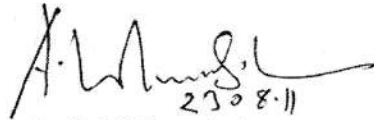
Kinds of Checks exercised

1. Ledger Book
2. Cash Book
3. Purchase orders
4. Stock Book Entry

  
Prof. D K Bhattacharyya  
Principal Investigator

**Professor**  
Department of Computer Science & Engg.  
Tezpur University

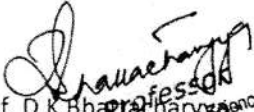
  
R R Borah  
Finance Officer  
Tezpur University  
Finance Officer  
TEZPUR UNIVERSITY

  
Prof. A K Buragohain  
Registrar Tezpur  
University  
Registrar  
Tezpur University



Statement of Expenditure  
DIT Funded Project

Sl. No	Head	Approved Budget Outlay (Rs.)	Expenditure (2008-2009) (Rs.)	Expenditure (2009-2010) (Rs.)	Expenditure (2010-2011) (Rs.)	Expenditure (2011-2012) Upto 12/8/2011 (Rs.)	Total Expenditure (2008-2012) Upto 12/8/2011 (Rs.)
1	Capital Equipment	27,40,000	4,67,584	18,83,672	3,42,554	-	26,93,810
2	Consumable Items / components	90,000	-	-	1,00,000	-	1,00,000
3	Manpower	4,51,200	49,548	2,92,800	3,12,000	15,413	6,69,761
4	Travel/Training	2,00,000	13,502	51,776	28,521	-	93,799
5	Contingencies	1,75,000	7,431	9,646	17,191	1,18,231	1,52,499
6	Overheads	7,30,640	-	5,55,000	37,403	84,728	6,77,131
	<b>Total</b>	<b>43,87,000</b>	<b>5,38,065</b>	<b>27,92,894</b>	<b>8,37,669</b>	<b>2,18,372</b>	<b>43,87,000</b>

  
Prof. D K Bhattacharya  
Principal Investigator  
Department of Computer Science & Engg.  
Tezpur University

  
R R Borah  
Finance Officer  
Tezpur University  
TEZPUR UNIVERSITY

