Quest for Research Excellence On Computing, Mathematics and Statistics

> Editors Kor Liew Kee Kamarul Ariffin Mansor Asmahani Nayan Shahida Farhan Zakaria Zanariah Idrus



Faculty of Computer and Mathematical Sciences

Conception

# Quest for Research Excellence on Computing, Mathematics and Statistics

**Chapters in Book** 

The 2<sup>nd</sup> International Conference on Computing, Mathematics and Statistics (iCMS2015)

Editors:

Kor Liew Lee Kamarul Ariffin Mansor Asmahani Nayan Shahida Farhan Zakaria Zanariah Idrus



# Quest for Research Excellence on Computing, Mathematics and Statistics

## **Chapters in Book**

The 2<sup>nd</sup> International Conference on Computing, Mathematics and Statistics

(iCMS2015)

4-5 November 2015 Langkawi Lagoon Resort Langkawi Island, Kedah Malaysia

Copyright © 2015 Universiti Teknologi MARA Cawangan Kedah

All rights reserved, except for educational purposes with no commercial interests. No part of this publication may be reproduced, copied, stored in any retrieval system or transmitted in any form or any means, electronic or mechanical including photocopying, recording or otherwise, without prior permission from the Rector, Universiti Teknologi MARA Cawangan Kedah, Kampus Merbok, 08400 Merbok, Kedah, Malaysia.

The views and opinions and technical recommendations expressed by the contributors are entirely their own and do not necessarily reflect the views of the editors, the Faculty or the University.

Publication by Faculty of Computer & Mathematical Sciences UiTM Kedah

ISBN 978-967-0314-26-6

# Content

# International Scientific Committee

Preface

CHAPTER 1	
CHAPTER 2	
<b>CHAPTER 3 </b>	
CHAPTER 4	
<b>CHAPTER 541</b> Dijkstra's Algorithm In Product Searching System (Prosearch) Nur Hasni Nasrudin, Siti Hajar Nasaruddin, Syarifah Syafiqah Wafa Syed Abdul Halim and Rosida Ahmad Junid	
CHAPTER 6	;

CHAPTER 7	
CHAPTER 8	
CHAPTER 9	
CHAPTER 10	
CHAPTER 11	
CHAPTER 12	
CHAPTER 13	
CHAPTER 14	
CHAPTER 15	

CHAPTER 16
CHAPTER 17
CHAPTER 18
CHAPTER 19
CHAPTER 20
CHAPTER 21213Estimating Philippine Dealing System Treasury (PDST)Reference Rate Yield Curves using a State-Space Representationof the Nelson-Siegel ModelLen Patrick Dominic M. Garces, and Ma. Eleanor R. Reserva
CHAPTER 22

CHAPTER 23
Partial Least Squares Based Financial Distressed Classifying Model of Small Construction Firms
Amirah-Hazwani Abdul Rahim, Ida-Normaya M. Nasir, Abd-Razak Ahmad, and Nurazlina Abdul Rashid
CHAPTER 24
CHAPTER 25
Data Mining in Predicting Firms Failure: A Comparative Study Using Artificial Neural Networks and Classification and
Regression Tree Norashikin Nasaruddin, Wan-Siti-Esah Che-Hussain, Asmahani Nayan, and Abd-Razak Ahmad
CHAPTER 26
Sanizah Ahmad, Norin Rahayu Shamsuddin, Nur Niswah Naslina Azid @ Maarof, and Hasfariza Farizad
CHAPTER 27
Kartini Kasim, Norin Rahayu Shamsuddin, Wan Zulkipli Wan Salleh, Kardina Kamaruddin, and Norazan Mohamed Ramli
CHAPTER 28
Outlier Detection in Time Series Model Nurul Sima Mohamad Shariff, Nor Aishah Hamzah, and Karmila Hanim Kamil
CHAPTER 29

CHAPTER 30
CHAPTER 31
CHAPTER 32
CHAPTER 33
CHAPTER 34
CHAPTER 35

CHAPTER 36	381
Technology Assistance for Kids with Learning Disabilities:	
Challenges and Opportunities	

Challenges and Opportunities Suhailah Mohd Yusof, Noor Hasnita Abdul Talib, and Jasmin Ilyani Ahmad

# CHAPTER 1 Towards Ameliorating the Problem of Packet Dropping in IDS using P System Model on GPU

#### Rufai Kazeem Idowu, Ravie Chandren M., and Zulaiha Ali Othman

Abstract. As the number of internet users grows exponentially everyday, the attacks and intrusions experienced on different networks equally multiply rapidly! Most of the current Intrusion Detection Systems (IDSs) are confronted with the challenge of coping with this high volume of traffics because they have low processing throughput. The overall consequence of this short-coming is what is known as packet dropping. P system, which is otherwise called membrane system, is a maximally parallel, non-deterministic and highly distributed model inspired by the functioning of living cells in Biology. Therefore, in exploring the parallelism advantage of both membrane system and Graphics Processing Unit (GPU), the paper presents an attack detection P system implemented on GPU to minimally decrease the negative impact of packet dropping which usually emanates from busy networks. On evaluation with KDD Cup dataset, our model achieves high throughput average of about 50000p/s and classification accuracy of 95.8%.

**Keywords:** Membrane Computing, Intrusion Detection, Network security, GPU.

Rufai Kazeem Idowu (🖂) • Ravie Chandren M

Centre for Software Technology & Management (SOFTAM), Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, (UKM) e-mail: ruffyk2001@yahoo.com, ravie@ukm.edu.my

# 1 Introduction

Within the cyberspace, every nefarious activity termed 'attack' or 'intrusion' is aim at: (i) compromising the integrity of the information system or/and (ii) denying its availability or/and (iii) rendering its performance inefficient. Based on the fore-going, the introduction of Intrusion Detection System (IDS) became expedient. Intrusion Detection System (IDS) which has been identified as a highly crucial element of Network/Computer Security (NCS) is saddled with the responsibility of providing protection within NCS's second layer [1]. In order to achieve this goal therefore, every IDS frequently keeps eagle's eye on the network environment or information system with a view to spotting and checkmating the impact of unusual activities which have high resemblance of an attack/intrusion [2], [3]. Therefore, an IDS which would be able to play the above required roles accordingly should exhibit among other traits, an extremely high throughput and accuracy in the face of the evergrowing network traffics. It has been shown that CPUs commonly become overwhelmed and invariably drop packets when their throughput could no longer cope with very high volume of traffic data [4], [5]. This challenge acts as a setback for IDS because it leads to inaccurate attack detection and increase in false positives. So, one of the ways of handling this within the classification domain is ensuring that the processing throughput is kept high.

Meanwhile, Membrane Computing (MC) which is a novel computational model, is an endowed parallel and distributed computing model having the inherent benefits of high understandability, communication advantage, dynamic feature, synchronization and non-linearity. Past researches have shown that MC is a very promising model for solving NP-hard problems and so has successfully been applied in several fields including biology, linguistics, medicine, economy, optimization, graphics and cryptography [6].

Furthermore, previous attempts to solve this problem of packet dropping in IDS had explored the hardware approaches of using multi-core, GPU or FPGA. So, this work employs a model which is designed in line with recognizer tissue P system by applying classification and symport communication rules on the objects contained within the membrane regions to ensure load balancing among the GPU processors. Ultimately, the implementation is premised on taking advantages of GPU and MC to mitigate packet dropping by having sustainable and greatly increased throughput under worst-case network traffic scenarios.

The remainder of the paper is structured under the following sections: Section 2 briefly discusses related works previous done on achieving high throughputs in IDSs and mitigating the impact of packet dropping. While Section 3 dwells on short description on how the attack detection P system works on GPU, Section 4 presents the simulations. Section 5 gives and discusses the obtained results. The final section 6 draws the conclusion.

## 2 Related Works

Several attempts to reduce the problem of packet dropping through the increase of processing throughput had been done in the past. These efforts include those of Vasiliadis et al. [7], Abhishek et al. [8], Song & Lockwood [9], Sun & Ricci [10] and Papadogiannakis et al. [4].

In 2005, Song & Lockwood [9] introduced BV-TCAM (Bit Vector-Ternary Content Addressable Memory architecture which combined the advantages of BV and TCAM. The BV-TCAM acted as a packet classifier and was able to flag multiple matches within extremely short network time. Their method which ultimately led to removal of prefix expansion of port ranges, achieved both throughput and very good storage efficiency. Similarly, Abhishek et al. [8] came up with principal component analysis (PCA) method implemented on FPGA with the primary aim of minimally reducing the network traffic dimensionality problem.

Also, from another perspective, Vasiliadis et al. [7], announced an IDS called MIDeA (Multi-Parallel Intrusion Detection Architecture) which has the ability to yield high processing speed without any drop in packet. Their work had three levels of parallelism which were done by adopting multi-queue NICs (Network Interface Cards), a set of multi-core CPUs and multiple GPUs. In MIDeA, packet processing and stateful analysis were done by mapping appropriate operations to distinct processing units hence unnecessary bottleneck and deadlock were avoided.

Furthermore, Sun & Ricci [10] proposed the use of extended Snap, which is a Click modular router-based approach. Their method yielded an increase in throughput of nearly 4x, thereby outperforming traditional software routers. Another remarkable progress was recorded by Papadogiannakis et al. in 2010 [4] where they implemented selective packet discarding in the Snort NIDS as a preprocessor that runs before the detection Engine. In their approach, NIDS could anticipate overload conditions and minimize their impact on attack detection by dynamically adjusting the number of packets that needed to be discarded.

# 3 Membrane System and the Attack Detection Model on GPU

#### 3.1 The Membrane System

Membrane systems or P systems are distributed, parallel and theoretical computing devices. A typical membrane system consists of rules which act on multiset of objects during a chemical reaction within an enclosed region. So,

depending on the nature of the rule, resultants objects are either consumed within the membrane or sent-out to the environment or to the adjoining membrane. An example of the Euler-Venn diagram representation of membrane systems and its equivalent tree structure can be seen in Fig.1 (a) and (b). It performs computation by applying the rules on the objects non-deterministically. The three basic variants of membrane systems are (1) Tissue-like, (2) Cell-like and (3) Neural-like.



Fig 1. (a) Membrane Structure (b) Membrane Tree

A *cell-like* membrane system is easily identified by the hierarchical arrangement of its cells. In this arrangement, the skin membrane is the protective layer shielding the entire membrane system from the environment. Within it are some other membranes (and the elementary membranes), objects and rules governing the relationship between the objects in the various regions. The evolution/re-writing and communication rules are usually applied in cell-like P systems [11]. In *neural-like* membrane system, while the only unique object called *spike* or *impulse* is recognized, its cells (neurons) are however linked through *synapses*. When a neuron spikes, its impulses are replicated and sent to all connecting neurons. The distance between consecutive spikes released into the environment constitute the result of the computation [12].

*Tissue-like* membrane system is a net of simple cells freely co-existing in a common environment – external space where there is compartmentalized communication and cooperation It is a one-membrane cells structure in which objects are moved in between adjacent (neighbouring) cells (provided the channels of communication are so defined) and the environment [13]. Most

often the symport/antiport rules of the type; (p, a/b, q) are applied for the movement of the objects in which p and q are cells' labels, a and b are multisets of objects.

#### 3.2 The Attack Detection P System Model

Membrane systems or Our Attack Detection P system ( $\prod_{AD_P}$ ) model was developed on the principle of recognizer tissue P systems introduced in [13]. A recognizer P system is often applied in decision making scenario-based problem such as True/False, Yes/No, On/Off in which only one option is applicable. In the model every region could evolve concurrently by sending the classified anomaly connection records to the environment. Formally, the  $\prod_{AD-P}$  is defined as a system of degree  $m \ge 1$  of the form:

$$\Pi_{AD_P} = (O, Y_1 \cdots Y_m, r, \beta, l)$$

(1)

Where:

- O is set of objects. An object represents a connection record in O, whereby O | ε [0, 4898430]. So, ⇒ O<sub>a</sub> ⊆ O where O<sub>a</sub> denotes arbitrarily many copies of anomalous connection record found in β.
- $Y_1 \cdots Y_m$  are membranes (cells) representing the zones of a network.
- r is a finite set of rules which is made up of types;  $r_1$  and  $r_2$  and defined thus:
  - (i)  $r_1$  are classification rules with guard and are of the type:
    - $$\begin{split} R_i &= a_{i23} a_{i6} a_{i27} \rightarrow s_{i1}; \\ &(a_{123} > 76.5 \, and \, a_{i6} \geq 40.5 \, and \, a_{i27} > 0.45); \\ &1 \leq i \leq MaxPac \end{split}$$

Where  $(a_{123} > 76.5 \text{ and } a_{i6} \ge 40.5 \text{ and } a_{i27} > 0.45)$ represents the conditional guard derived from the classification tree and  $s_i = \{1,0\}$  denotes the status of the connection record  $a_i$  which may either be intrusive (0) or non-intrusive (1) determined by the features 23, 6 and 27.

(ii)  $r_2$  are symport communication rules of the type:  $O_i \rightarrow (anomaly, \beta);$  This rule is applied to release *anomaly* traffics to the environment through the individual membranes. It implies that if anomalous connection record is detected, the rule would be used to transport affected object O within  $Y_1 \cdots Y_m$  to the environment,  $\beta$ .

Rules were used in non-deterministic and maximally parallel manner as tradition with computation in membrane systems. In each step, all objects and all cells which can evolve must evolve.

- $\beta = O \{Anomaly\};$  is the environment/zone. This external membrane environment is where the results of computation are obtained and so, it is called the output region. It does not hold any rule. Since the working packets are either *normal* or *anomaly*, hence the computation of  $\prod_{AD_{-}P}$  system halts in the accepting mode if only anomalous packets  $O_a$  (and strictly excluding normal connection records  $(O_n)$ ) are sent to the environment, otherwise, it is a rejecting computation. This stage signifies the end of computation (i.e final configuration). Please note that  $(O_n, O_a \subseteq O)$ .
- $l \subseteq \{1, 2, \dots, m\} x \{\beta\}$  which is a link (also known as channel or synapse) between the membranes and the environment,  $\beta$ .

#### 4 Simulations

#### 4.1 Dataset

For our simulations, we used the KDD Cup '99 dataset [14] which is the benchmark dataset for IDS. The dataset is made up of 4,898,430 labeled connection records whereby each data consists of 41 features which may either be quantitative or qualitative feature. Therefore, the first step in the implementation is the pre-processing of the data by ensuring that qualitative data were converted to numeric data.

#### 4.2 Methods and Materials

Simulation was done on a desktop computer which has the following configuration:

Core i7-3820 CPU and an NVIDIA GeForce GTX680 GPU. Furthermore, we Used CUDA extensions with GPU-enabled functions in MATLAB's parallel computing toolbox with CUDA extensions.

The first step in the implementation is the pre-processing of the data by ensuring that textual data were converted to numeric forms. The structure of the membranes and how the objects are represented in the  $\prod_{AD_P}$  model are depicted in fig.2. Several one-membrane cells (ovals) are considered as evolving in a common external environment ( $\beta$ ) where results are obtained. No direct communication exists in between the cells, but all the cells communicate with the environment since channels for transportation of such were specified in advance as  $l \subseteq \{1, 2, \dots, m\} x \{\beta\}$ . These ovals were labeled with  $1, 2, \dots, m$  and objects with distinctive embedded 41 elements (features) and applicable set of rules were equally specified. However, the arrows indicate that the decided instances of *`anomaly'* obtained by the application of the classification rules, leave the cells in maximal mode through the channels  $(1, \beta), (2, \beta), \dots, (m, \beta)$  to the external environment using symport rules. The dimension of the cells in the  $\prod_{AD-P}$  model is determined based on the number of thread blocks available on the GPU.



**Fig. 2**. Membrane/object representation in  $\prod_{AD-P}$  model

## 5 Results & Discussion

#### 5.1 Throughput

From Table 1, while columns 1 and 2 show all the tested packets and membrane numbers respectively, columns 5 and 6 show the throughputs for

both the GPU and CPU respectively. Since one of the key performance metrics for a network intrusion detection system is a sustainable throughput [5], the  $\prod_{AD-P}$  model achieved very good results in that regard. However, in Fig. 3, it could be observed that when the number of membranes is 16, the throughput for CPU is better than that of the GPU. This is because there is ineffective utilization of GPU resources in that instance when the number of membrane simulated on it is so small.

Packet Size (All	Membrane Number	Time(s) GPU	Time (s) CPU	Throughput (GPU)	Throughput (CPU)
Test)				Pac/Sec	Pac/Sec
314572	16	44.8	34.7	7014.3	9039.7
314572	32	23.1	34.7	13606.3	9039.7
314572	64	13.7	34.7	22814.5	9039.7
314572	128	9.3	34.7	33696.3	9039.7
314572	256	7.5	34.7	41853.9	9039.7
314572	512	6.6	34.7	47580.1	9039.7
314572	1024	6.1	34.7	50944.9	9039.7
314572	2048	5.9	34.7	53102.6	9039.7

**Table 1**: Throughput of CPU and GPU using  $\prod_{AD_P}$  model

Generally, the average throughput of the  $\prod_{AD_P}$  model on GPU is sufficient to checkmate packet drop/loss in an IDS. This is closely related to the increase in multiprocessor occupancy of the GPU which ultimately improves the system's efficiency.



Fig. 3. Comparison of CPU and GPU throughputs

#### 5.2 The Classification Accuracy

The classification accuracy is defined as the rate at which an IDS predicts the status (either normal or abnormal) of a traffic with a degree of certainty. It is usually obtained by the equation:

$$Classification \ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(2)

Where TP = True Positive, TN = True Negative, FP = False Negative and FN = False Negative. So, with the application of  $\prod_{AD_P}$  model, the classification accuracy rate obtained is 95.8%.

## 6 Conclusion

So far in this paper, we have presented our initial results on the proposed Attack Detection P System. By evaluation, it has shown that our model has the great potential of decreasing the problem of packet dropping usually encountered by IDS when the traffics become unimaginably large. The throughput and classification accuracy obtained are high which suggest that MC is a better alternative in tackling efficiency issue in IDS.

# References

- [1] Wu, Shelly Xiaonan, and Wolfgang Banzhaf. "The use of computational intelligence in intrusion detection systems: A review." *Applied Soft Computing* 10, no. 1: 1-35. (2010).
- [2] Ajith Abraham, Crina Grosan, and Yuehui Chen: Evolution of intrusion detection systems. School of Computer Science and Engineering, Chung-Ang University, Korea, 2005.
- [3] Uma M., Padmavathi G.: A Survey on Various Cyber Attacks and their Classification. International Journal of Network Security, Vol.15, No.6, PP.391-397, (2013).
- [4] Antonis Papadogiannakis, Michalis Polychronakis, and Evangelos P. Markatos."Improving the accuracy of network intrusion detection systems under load using selective packet discarding."In *Proceedings of the Third European Workshop on System Security*, pp. 15-21.ACM, (2010).
- [5] Schaelicke Lambert, and J. Curt Freeland."Characterizing sources and remedies for packet loss in network intrusion detection systems."In

Workload Characterization Symposium, 2005. Proceedings of the IEEE International, pp. 188-196. IEEE, (2005).

- [6] Ciobanu, G., & Păun, Gh.: *Applications of membrane computing* (Vol. 17). Berlin: Springer (2006).
- [7] Giorgos Vasiliadis, Michalis Polychronakis, and Sotiris Ioannidis. Midea: a multi-parallel intrusion detection architecture. In ACM Conference on Computer and Communications Security, pages 297–308, (2011).
- [8] Abhishek Das, Sanchit Misra, Sumeet Joshi, Joseph Zambreno, GokhanMemik, and AlokChoudhary. "An efficient fpga implementation of principle component analysis based network intrusion detection system." In *Proceedings of the conference on Design, automation and test in Europe*, pp. 1160-1165. ACM, (2008).
- [9] Haoyu Song, John W. Lockwood: Efficient Packet Classification for Network Intrusion Detection using FPGA. International Symposium on Field-Programmable Gate Arrays (FPGA '05), Monterey, CA, Feb 20-22, (2005)
- [10] Weibin Sun, Robert Ricci: "Fast and Flexible: Parallel Packet Processing with GPUs and Click," in Proc. of ANCS, pp. 25-36, (2013).
- [11] Păun Gh & Rozenberg G: A guide to membrane computing, Theoretical Computer Science, 287, 73–100, (2002).
- [12] Ibarra O.H, Păun Gh.: Membrane Computing: A General View, The European Academy of Sciences, (2007).
- [13] Carlos Mart-Vide, Gheorghe Păun, Juan Pazos, Alfonso Rodriguez-Paton: Tissue P systems Theoretical Computer Science 296, 295 – 326 (2003).
- [14] KDD Cup99 Dataset, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html(1999).





View publication stat