Malek-Ashtar University of Technology
Faculty of Electronics and Computer

M.Sc Thesis

Title

Provide a Novel Sophisticated Hybrid method for Intrusion Detection Using the Artificial Immune System (AIS)

By

Ehsan Farzadnia

Supervised by

Dr. H. Shirazi
Dr. A. Nowroozi

31 December 2018

In the name of God

Malek-Ashtar University of Technology
Faculty of Electronics and Computer

Provide a Novel Sophisticated Hybrid method for Intrusion Detection Using the Artificial Immune System (AIS)

A Thesis
Submitted in partial fulfilment of the requirements
for the degree of Master of Science (M.Sc) at Secure Computing

By
Ehsan Farzadnia

Evaluated and approved by the Thesis Committe, on 31 Dec 2018.

No.	Title	Responsibility	Signature
1	Dr. H. Shirazi	Supervisor(1)	Gh.M.Shiraz
2	Dr. A. Nowroozi	Supervisor(2)	
3	Dr. B. Minaie	Examiner (External)	
4	Dr. K. Dadashtabar	Examiner (Internal)	
5		Department Graduate Coordinator	

Lib No. : MUT 18609
Degree: Master of Science at Secure Computation
Title: Provide a novel sophisticated hybrid method for intrusion detection using the artificial immune system
Author: Ehsan Farzadnia
Supervisor(1): Dr. H. Shirazi[1]
Supervisor(2): Dr. A. Nowroozi[2]
Examiner(External): Dr. B. Minaie[3]
Examiner(Internal): Dr. K. Dadashtabar[4]
Department: Information Security and Cyber Defenses
Faculty: Faculty of Electronics and Computer
Date: 31 December 2018

Abstract

Anomaly Detection is one of the significant and complex problems in computer networks security. Artificial Immune Systems are among the novel methods which have been mostly used by researchers in recent years. This system is a good prototype to develop Machine Learning techniques, and it can be a proper candidate for designing the next generation of intrusion detection systems owing to its many potentials. Two of its paradigms, the Danger Theory and the Negative Selection mechanisms, are inspired by the Human Immune System (HIS). These methods are being widely used by scholars in the Intrusion detection field. The general purpose of the current research is to establish a type of artificial life through simulation and incorporation of two innate and acquired immunity mechanisms with a function similar to HIS in order for achieving an immune and stable intrusion detection system. In other words, the goal is to attain security from immunity.

In this thesis, we provide a hybrid method with two lines of defense based on improvement and incorporating two real-valued negative selection and dendritic cell (DC) algorithms. In our approach, distributed cooperating of DCs with mature effective detectors act as a stimulator leading to produce more effective detectors and regulate memory pool dynamically and immunity maturation

---

[1] Faculty member at Malek-Ashtar University of Technology
[2] Postdoc researcher, Sharif University of Technology
[3] Faculty member at Iran University of Science and Technology
[4] Faculty member at Malek-Ashtar University of Technology

which means security control via immunity establishment. Simulation of such schema required an artificial life to be executed in the MATLAB environment. Evaluation of the final experiment shows that the proposed hybrid schema is more efficient than other 17 prevalent techniques found in literature. Moreover, improvement rate (in percent) for the proposed hybrid method is obtained 9.25 for detection, 93.91 and 19.75, respectively for FN and FP, 22.22% for detection of unseen attacks, 1.13% for accuracy, and 5.89% for cc, comparing with two artificial immune-based methods of FtRNSA and IO-RNSA, at the end of 19^{th} cycle. Capability of Intrusion Detection (CID) rate was also high. Consequences demonstrate that the proposed hybrid method is able to achieve relative stability in controlling security by passing more cycles and covering maximum wee holes in long-term.

Keywords: Network anomaly detection, Artificial immune system, Artificial life, Mature detector, Dendritic cell

Publications

[1]. **Farzadnia, Ehsan**, Hossein Shirazi, and Alireza Nowroozi. "A novel sophisticated hybrid method for intrusion detection using the artificial immune system." *Journal of Information Security and Applications* 58 (2021): 102721.

[2]. **Farzadnia, Ehsan**, Hossein Shirazi, and Alireza Nowroozi. "A new intrusion detection system using the improved dendritic cell algorithm." *The Computer Journal* 64.8 (2021): 1193-1214.

[3]. Shirazi, Hossein, **Ehsan Farzadnia**, and Alireza Norouzi. "Scrutinizing and evaluating intrusion detection approaches based on the artificial immune system." *C4I Journal* 2.1 (2019): 26-49.

[4]. **Farzadnia, Ehsan**. Shirazi, Hossein. Norwoozi, Alireza, “Scrutinizing & Evaluating the Defense approaches of Bio-immune inspired Intrusion Detection Systems.” (Oral presentation in persian), 4^{th} *International Conference on New Studies of Computer and IT*, February 24-25, 2018 at *Sadjad University of Technology, Mashhad.*

Technical Reports

[1]. **Ehsan Farzadnia**, Supervisors: Hossein Shirazi, Alireza Nowroozi', “Generating Uniformly Random Numbers within a Hyper Shape (Super-Ring) in N-dimensional Space : A MATLAB Simulation*”, Dept.of Communication and Information Security, Malek Ashtar University of Technology*, May, 23, 2018.

[2]. **Ehsan Farzadnia**, Supervisors: Hossein Shirazi, Alireza Nowroozi, “The Real Valued Negative Selection Algorithm (RNSA) : A MATLAB Simulation”, *Dept.of Communication and Information Security, Malek Ashtar University of Technology*, May, 25, 2018.

[3]. **Ehsan Farzadnia**, Supervisors: Hossein Shirazi, Alireza Nowroozi, “The Black Hole Clustering Algorithm : A MATLAB simulation”, *Dept.of Communication and Information Security, Malek Ashtar University of Technology*, September, 21, 2017

About me

Ehsan Farzadnia was born on 22nd of September 1987 in Urmia, Iran. He studied Computer Engineering (Software) and obtained his Bachelor in this field from Urmia University in March 2011. After doing his military service, while working for a private computer company, he decided to pursue his career with a Master degree. Due to this, he participated in National Organization of Educational Testing (Konkur), and was able to get an admission offer in Computer Engineering (field of interest: Secure Computation) from Malek Ashtar University of Technology in September 2015. He continued his research in intrusion detection systems and succeeded in defending his thesis on 31th December 2018. His research interests include but are not limited to Steganography, Immunoinformatics, Quantum computation, Anomaly detection, Machine learning and Data mining. He is currently working with Python and conducting interdisciplinary studies in order to provide a high quality proposal for PhD.


بسم الله الرحمن الرحيم

بسمه تعالی

مجتمع دانشگاهی برق و کامپیوتر

پایان نامه دوره کارشناسی ارشد رشته مهندسی کامپیوتر – گرایش رایانشِ امن

با عنوان

ارائه روشی خِبره برای تشخیص نفوذ با استفاده از سیستم ایمنی مصنوعی

توسط

احسان فرزادنیا

اساتید راهنما

دکتر حسین شیرازی

دکتر علیرضا نوروزی

۱۰ دیماه ۱۳۹۷

بسمه تعالی

تعیین سطح طبقه‌بندی پایان‌نامه

عنوان پایان‌نامه: **ارائه روشی برای تشخیصِ نفوذ با استفاده از سیستمِ ایمنیِ مصنوعی**

نام و نام خانوادگی استاد راهنما اوّل : **دکتر حسین شیرازی** نام و نام خانوادگی استاد راهنمای دوّم : **دکتر علیرضا نوروزی**

نام و نام خانوادگی دانشجو: **احسان فرزادنیا** شماره دانشجویی: **943832067**

رشته: **مهندسی کامپیوتر** گرایش: **رایانشِ امن**

از نظر استاد راهنما این پایان‌نامه: 1- طبقه‌بندی دارد ☐ 2- طبقه‌بندی ندارد ■

نام و امضاء استاد راهنما:

"تبصره: با توجه به طبقه بندی دار بودن پایان نامه، تا تاریخ ..... این پایان نامه، فقط با نظر این واحد دانشگاهی، در اختیار متقاضی قرار گیرد."

ریاست گروه علمی

مدیرآموزش و تحصیلات تکمیلی مجتمع
دکتر کوروش داداش تبار

ریاست مجتمع
دکترعلی جبار رشیدی

تأییدیه صحت و اصالت نتایج پایان‌نامه

اینجانب **احسان فرزادنیا** به شماره دانشجویی **943832067** دانشجوی رشته **مهندسی کامپیوتر** گرایش **رایانشِ امن** مقطع تحصیلی کارشناسی ارشد تأیید می‌نمایم که کلیه نتایج این پایان‌نامه حاصل کار اینجانب و بدون دخل و تصرف است و موارد نسخه‌برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع آورده‌ام.

در صورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم (قانون حمایت از حقوق مؤلفان و مصنفان و قانون ترجمه و تکثیر کتب و نشریات و آثار صوتی، ضوابط و مقررات آموزشی، پژوهشی و انضباطی...) با اینجانب رفتار خواهد شد و حق هرگونه اعتراض در خصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش صلب می‌نمایم. در ضمن، مسئولیت هر گونه پاسخگویی به اشخاص اعم از حقیقی و مراجع ذیصلاح (اعم از اداری و قضایی) به عهده اینجانب خواهد بود و دانشگاه هیچگونه مسئولیتی در این خصوص نخواهد داشت.

نام و نام خانوادگی دانشجو: احسان فرزادنیا

امضاء و تاریخ: ۹۷/۱۱/۱۹

بسمه تعالی

دانشگاه صنعتی مالک اشتر

مجتمع دانشگاهی برق و کامپیوتر

صورتجلسه دفاعیه پایان نامه

با تاییدات خداوند متعال، جلسه دفاعیه پایان نامه آقای /خانم **احسان فرزادنیا** تحت عنوان: **ارائه روشی برای تشخیصِ نفوذ با استفاده از سیستمِ ایمنیِ مصنوعی** در رشته: **کامپیوتر** گرایش: **رایانشِ امن** در مورخ **97/10/10** تشکیل و پس از دفاع با نمره ۹۷,۵ از100 (**معادل** ۱۷,۵۵ **از** 18) مورد ارزیابی قرار گرفت.

1- استاد راهنمای اوّل : **دکتر حسین شیرازی**

2- استاد راهنمای دوّم: **دکتر علیرضا نوروزی**

3- استاد داور: **دکتر بهروز مینایی**

4- استاد داور: **دکتر کوروش داداش تبار**

با توجه به 2 نمره مربوط به سنوات، ارایه مقاله و گزارش های پیشرفت پروژه، نمره نهایی دانشجو ۱۹,۵۵ و با درجه عالی مورد تصویب قرار گرفت.

5- مدیر آموزش و تحصیلات تکمیلی مجتمع : **دکترکوروش داداش تبار**

6- رئیس/ معاون آموزش و پژوهش مجتمع : **دکتررضا فاطمی مفرد**

«مَن لَم یَشکُر المَخلوق، لَم یَشکُر الخالِق»

خداوند متعال را شُکر می گویم که توفیقِ تحقیق و تدوینِ این پایان نامه را نصیب اینجانب نمود. در راستای انجام این اثر ، افرادی به طور مستقیم و یا غیر مستقیم نقش داشته اند که نام بُردن از همه آنها در اینجا میسّر نیست. بنابراین ، از همه کسانی که در این راستا به نحوی همکاری داشته اند نهایتِ تشکّر و قدردانی را می نمایم.

این پایان نامه بر اساسِ مطالعات گسترده از منابع علمیِ معتبر جهان در سه حوزه تخصصیِ ایمنی شناسی، هوش مصنوعی و امنیّت فناوری اطلاعات و زیر نظر اساتیدِ محترمِ راهنما جنابِ دکتر حسین شیرازی و جنابِ دکتر علیرضا نوروزی به انجام رسیده است. جا دارد از زحماتِ ایشان در جهت دهی به مطالعات و ارائه تجربیات علمی و تحقیقاتیِ ارزشمند خود به اینجانب در راستای به ثمر رسیدنِ اهداف این تحقیق سپاسگزاری نمایم.

در پایان امیدِ آن دارم که در آینده روندِ تحقیقات بیش از پیش به سمتِ موضوعات بین رشته ای چون الهام گرفتن از شگفتی های طبیعت در جهت ایده پردازی ، مُدلسازی و مهندسی ایده ها سوق پیدا نموده و در راستای حلّ مسائلِ پیچیده علوم کامپیوتر بویژه حوزه ی تخصّصیِ امنیّت فناوری اطلاعات موثّر واقع گردد.

پائیز ۱۳۹۷ خورشیدی - احسان فرزادنیا

با سپاس ویژه از

دکتر حسین شیرازی

دکتر علیرضا نوروزی

تایید صحّت ، اصالت و رعایتِ امانت در پایان نامه

اینجانب احسان فرزادنیا دانشجوی رشته مهندسی کامپیوتر گرایش رایانشِ امن مقطع تحصیلی کارشناسی ارشد به شماره دانشجویی ۹۴۳۸۳۲۰۶۷ ، صحّت ، اصالت و رعایتِ امانَت در پایان نامه را تائید و اعلام می نمایم کلیه نتایج این پایان نامه و نشریات مرتبط با آن (از قبیل مقالات استخراج شده) حاصل کار اینجانب و بدون هر گونه دَخل و تصرّف بوده و موارد نسخه برداری شده از آثار دیگران را با درج کامل مشخصات منبع ذکر کرده ام.

علاوه بر این ، هر گونه آثار مرتبط با این پایان نامه را با رعایت امانَت منتشر نموده و یا خواهم نمود. بنابراین ، در تمامی آثار مرتبط (نظیر مقاله ، ثبت اختراع ، شرکت در جشنواره ها و کنفرانسها و مواردی از این قبیل) حقوق مالکیت دانشگاه صنعتی مالک اشتر تهران (از جمله حقوق اساتید محترم راهنما و مشاور) را رعایت می نمایم و هر گونه اثری را با هماهنگی و درج نام مبادی ذیصلاحِ دانشگاه (از جمله اساتید محترم راهنما) منتشر خواهم نمود.

در صورت اثبات خلاف مندرجات فوق و یا هر گونه تخلّفی که صحّت ، اصالت و رعایتِ امانَت در پایان نامه را مخدوش نماید ، دانشگاه می تواند مطابق با ضوابط و مقرّرات حاکم (قانون حمایت از حقوق مؤلّفان و مصنفان و قانونِ ترجمه و تکثیرِ کُتُب و نشریات و آثارِ صوتی ، ضوابط و مقرّراتِ آموزشی ، پژوهشی و انضباطی) با اینجانب رفتار نماید. بنابراین ، حق هر گونه اعتراض در خصوصِ احقاقِ حقوقِ مکتسب و تشخیص و تعیینِ تخلّف و مجازات را از خویش سلب می نمایم. همچنین ، مسئولیّتِ هر گونه پاسخگویی به اشخاص اعم از حقیقی و حقوقی و مراجعِ ذیصلاح (اعم از اداری و قضایی) به عهده اینجانب خواهد بود و دانشگاه هیچ گونه مسئولیّتی در این خصوص نخواهد داشت.

آبان ۱۳۹۷ خورشیدی
نام و نام خانوادگی دانشجو
اثر انگشت و امضاء

این پایان نامه را با نهایتِ احترام تقدیم می کنم به :

مادرم ، این اقیانوسِ محبّت و مهربانی

پدرم ، این دشتِ سبزِ ایثار

و

جویندگانِ راهِ راستینِ حقیقتِ علم و دانش

آنان که علم را به هیچ بَهایی نمی فروشند

چکیده

تشخیص ناهنجاری یکی از مسائل مهم و پیچیده در امنیّت شبکه های کامپیوتری می باشد. از جمله روشهای نوینی که در سالهای اخیر جهت این موضوع بیشتر مورد استفاده محققان قرار گرفته اند سیستم های ایمنی مصنوعی می باشند. این سیستم ، نمونه اوّلیه مناسبی برای توسعه یادگیری ماشین بوده و به دلیل وجود پتانسیلهای فراوان می تواند کاندیدای مناسبی در طراحی سیستمهای تشخیص نفوذ آینده باشد. دو الگوی فکریِ مهمّ این سیستم ، تئوری خطر و انتخاب منفی ، سازوکارهای ایمنی الهام گرفته شده از سیستم ایمنی بدن انسان هستند. این دو متد به طور گسترده در زمینه تشخیص ناهنجاری مورد استفاده محققان قرار میگیرند. هدف اصلی این پژوهش ، برقراری نوعی حیات مصنوعی از طریق شبیه سازی و تلفیقِ دو سازوکار مصونیّت ذاتی و تطبیق پذیر با کارکردی مشابه ایمنی بدن انسان در جهت دستیابی به یک سیستم تشخیص نفوذِ پایدار و ایمنِ شبکه می باشد. به عبارت بهتر ، هدف ، رسیدن از ایمنی به امنیّت است.

در این پایان نامه ، یک روش هیبریدی با دو خط (لایه) دفاعی بر مبنای بهبود و تلفیقِ دو الگوریتم انتخاب منفیِ حقیقی مبنا و الگوریتم سلّول دِندریت ارائه نموده ایم. در این روش ، مشارکت توزیع شده سلّولهای دِندریت با تشخیص دهنده های بالغ به عنوان محرّک عمل نموده و سبب تولید تشخیص دهنده های موثّرتر و تعدیل پویای حافظه و بلوغ ایمنی می گردد که این به معنای کنترل امنیّت از طریق برقراری ایمنی می باشد. شبیه سازی چنین سیستمی، نیازمند ایجاد نوعی حیات مصنوعی بود که در محیط نرم افزار متلب اجرا گردید. ارزیابیِ نتایج نهاییِ آزمون ، کارآمدیِ روش پیشنهادی را در مقایسه با ۱۷ دسته بند رایج نشان می دهد. بعلاوه ، میزان بهبودِ سیستم پیشنهادی در مقایسه با دو روش ایمنی مصنوعی (FtRNSA و IO-RNSA) در پایانِ سیکل ۱۹ اُمِ حیات، دست کم ۹,۲۵ ٪ برای تشخیص ، ۹۳,۹۱ و ۱۹,۷۵ درصد به ترتیب برای خطای منفی و مثبت کاذب، ۲۲,۲۲ ٪ برای شناسایی حملات ناشناس، ۱,۱۳ ٪ برای صحّت و ۵,۸۹ ٪ برای ضریب همبستگی بدست آمده و استعدادِ تشخیص نیز بالا بود. نتایج بدست آمده حکایت از آن دارند که روش پیشنهادی با پیمودن سیکل های بیشتر و با پوشش حداکثریِ فضای حَفره ها در دراز مدّت قادر است به یک ثبات نسبی در کنترل امنیّت دست یابد.

واژه های کلیدی: تشخیص ناهنجاری شبکه ، سیستم ایمنیِ مصنوعی ، حیاتِ مصنوعی ، تشخیص دهنده بالغ ، سلّول دِندریت

عبارت های اختصاری

جدول صفر - عبارت های اختصاری استفاده شده در طول گزارشِ پایان نامه

عبارت اصلی	عبارت اختصاری	عبارت اختصاری به انگلیسی	توضیحات
تشخیص دهنده[1]	تش.د.	D	آنتی بادی نابالغ
آنتی ژن غیر خودی	آژ.غ.خد.	nsAg	-
تشخیص دهنده بالغ	تش.د.بغ.	MD	آنتی بادی بالغ
بالغ موثّر	بغ.م.	EM	-
تشخیص دهنده بالغ موثّر	تش.د.بغ.م.	EMD	آنتی بادی بالغی که در طول عمر خود حداقل یکبار تشخیص موفق انجام داده و موثر واقع شده است. این آنتی بادی بالغ ، پتانسیل تبدیل شدن به آنتی بادی بالغ موثّر حافظه را دارد.
تشخیص دهنده بالغِ موثّرِ حافظه	تش.د.بغ.م.حظ.	MEMD	آنتی بادی بالغی که با شناسایی آنتی ژن غیر خودی تکثیر شده و تبدیل به حافظه شده است.
بالغِ موثّرِ حافظه	بغ.م.حظ.	MEM	-
سلّول دندریت	سل.دن.	DC	گونه ای از یاخته خواران که به نمونه برداری از سلولهای خودی بافتهای بدن مشغول هستند تا با نابودیِ آژ.غ و تحویل آنها به لنفوسیتِ T مانع تخریب سلولی شوند.
مصونیّتِ ذاتی	مص.ذت.	InIm	واکنشِ ایمنی ذاتی که در لایه نخست دفاعیِ زیر پوست جاندار در برابر آنتی ژن غیر خودی (عوامل غیر خودی) بوسیله یاخته خواران بوقوع می پیوندد.
مصونیّت تطبیق پذیر	مص.تط.	AdIm	واکنشِ ایمنی تطبیق پذیر (اکتسابی [2]) که در لایه دوّم دفاعی در برابر عوامل نفوذی بوسیله
سلّول دندریتِ بالغ	سل.دن.بغ.	MDC	سلولهای دندریت به محض آنکه سیگنالها را نمونه برداری نمودند بر اساس پردازشی که بر روی سیگنالهای سالم یا تخریب دریافت شده از بافت های سلولی بدن انجام می دهند به این وضعیت در آمده و اقدام به مهاجرت به گروه سلولهای دندریت بالغ می نمایند. اصطلاحاً به سلول دندریت به وضعیت بالغ در می آید. این وضعیت نشان دهنده وجود خطر در بافتهای سلولی می باشد.
دندریتِ بالغ	دن.بغ.	-	-
سلّول دندریتِ نیمه بالغ	سل.دن.نبغ.	SMDC	سلولهای دندریت به محض آنکه سیگنالها را نمونه برداری نمودند بر اساس پردازشی که بر روی سیگنالهای سالم یا تخریب دریافت شده از بافت های سلولی بدن انجام می دهند به این وضعیت در آمده و اقدام به مهاجرت به گروه سلولهای دندریت نیمه بالغ می نمایند. اصطلاحاً به سلول دندریت به وضعیت نیمه بالغ در می آید. این وضعیت نشان دهنده وضعیت امن و به نوعی سیگنالی آماده باش به لایه دفاعی دوّم (لنفوسیتهای T و B) ارسال می کند تا آماده مقابله با حملات باشند.
دندریتِ نیمه بالغ	دن.نبغ.	-	-

[1] Detector

[2] Acquired Immunity

فهرست مطالب

عنوان	صفحه

بخش دوّم : ادبیّات تجربی

فصل سوّم : ارائه روش پیشنهادی

بخش نخست : تئوری خطر

بخش دوّم : انتخاب منفی

پیوست ب - جداول ، اشکال و نمودارها

فهرست شکل ها

فهرست نمودارها

فصل نخست

Generalities of Research

کلّیاتِ پژوهش

۱-۱- مقدمه

به منظور حلّ مسائلِ پیچیده در دنیای واقعی ، طبیعت ، همیشه الگوی مناسبی برای الهام بوده است. در سالهای اخیر کارهای تحقیقاتی در حوزه امنیّت شبکه و خصوصاً تشخیصِ ناهنجاری به سمت الهام از حقایق موجود در طبیعت متمایل شده اند تا در جهت حلّ مسائل مشکل و پیچیده ی این حوزه بتوانند ایده ها و راهکارهای نو را ارائه نمایند. انواع الگوریتمهای فراابتکاری در مسائل بهینه سازی [۱]و دسته بندها[۲] ، بیوروباتها[۳] در مسائل واقعی روزمرّه از جمله ی موفّق ترینِ این راهکارها و ایده ها هستند.

امروزه با گسترش شبکه های ارتباطی ، مفهوم شبکه از سطح کامپیوتری بودن به مفاهیم بسیار گسترده تری چون شبکه های اینترنت- چیزها ، شبکه های ابری [۴]، شبکه های اجتماعی [۵] ، شبکه های زیرساخت های حسّاس کنترل صنعتی [۶] ، شبکه های بی سیم[۷] و غیره توسعه پیدا کرده است. بنابراین کاربرد عبارت "شبکه های ارتباطی" کاملاً به جا می باشد. از طرفی با این گسترش احساس می شود که دنیای شبکه به دلیل پویاتر شدن کاربردهای گوناگون مبتنی بر هوش مصنوعی و ادغام سایر علوم در آن رفته رفته تبدیل به دنیایی

---

[1] Bio – inspired Algorithms (Meta – Heuristic optimization techniques)
[2] Classifiers
[3] Nature – inspired Robots (Biorobots)
[4] Cloud Based Networks
[5] Social Networks
[6] SCADA
[7] Wireless Networks

زنده می شود. دنیایی که در آن هر گره هوشمند دارای شناسه (IP) خاص هست. به نظر می رسد نیاز به نظارت و کنترل عامل انسانی رو به زوال رفته است.

از اینرو یکی از بزرگترین مشکلات جوامع امروزی که در سالهای اخیر تاثیر آن به مراتب بیشتر شده است افزایش شدّت و تعداد حملات رایانه ای به شبکه های ارتباطی سازمانها و شرکتها بوده است. علّت عمده ی این تاثیر مربوط به وجود نقاط ضعف و حفره های بالقوّه ی امنیّتی و آسیب پذیر در سیستم های رایانه ای بوده که متاسفانه سالانه قربانیان زیادی را هدف حملات و نفوذ نفوذگران و سوء استفاده کنندگان قرار می دهد. شدّت این حملات با رشد و توسعه ی شبکه ها و افزایش ارتباط بین رایانه ها به صورت نمایی افزایش می یابد. پدیده نفوذ را مجموعه ای فعالیتها تعریف می کنند که نفوذگران تلاش می کنند تا جامعیّت [1]، قابلیّت اطمینان [2] و یا دسترس بودن یک منبع[3] را به خطر بیندازند.

تعریف – به فرایند نظارت بر وقایع رخ داده در یک شبکه و یا سیستم های کامپیوتری در جهت کشف موارد انحراف (ناهنجاری) از سیاست امنیّتی تشخیص نفوذ (ID) گفته می شود. [۶۵]

با رخداد نفوذ شبکه و وجود ترافیک آنومالی ، کار شناسایی آن توسط عامل انسانی بسیار مشکل بوده و در بیشتر موارد به دلیل الگوهای متنوّع در ترافیک حملاتی که از طریق شبکه رخ می دهند حتّی غیر ممکن به نظر میرسد. به همین دلیل نیاز به وجود یک عامل هوشمند ماشینی حس می شود تا با استفاده از تکنیکهای یادگیری ماشین و داده کاوی بتواند ضمن شناسایی الگو و یادگیری و کسب تجارب لازم ، این پدیده را شناسایی نموده و جلوی دسترسی مشکوک به سیستم هدف را به سریعترین شکل ممکن بگیرد. سیستم های تشخیص و جلوگیری از نفوذ [4]، مجموعه ای از عملیاتی را بر مبنای استراتژی یادگیری خود مدیریت ، برنامه ریزی و پردازش می کنند که قابلیّت اطمیان و یکپارچگی منابع اطلاعاتی [5]را تضمین نمایند. یک IDPS در واقع یک سیستم مدیریت امنیّتِ اطلاعات سیستم های رایانه ای نیز هست.

بزرگترین چالشی که عمدتاً مدیران و متخصّصان هوش مصنوعی و امنیّت سایبری با آن مواجه اند در حال حاضر مسئله سخت " تشخیص بلادرنگ ناهنجاری ناشناخته ی شبکه [6] با کمترین هزینه ی ممکن" می باشد. به نظر میرسد بهره برداری و تقلید هوشمندانه از مفاهیم نُوینی همچون سیستم دفاعی ایمنی زیستی (BIS) و مُدلسازی و مهندسی دستاوردهای آن تاحدودی توانسته آغازگر راهی باشد که ضمن رفع بسیاری از این چالشها نحوه نگرش به رویکردهای جاری در حلّ این مسائل را از جنبه های گوناگون با چالشهای جدیدی مواجه سازد. با این مقدّمه در این فصل به بیان مسئله و اهداف این پژوهش می پردازیم.

---

[1] Integrity
[2] Reliability
[3] Availablity
[4] Intrusion Detection and Prevention Systems (IDPS)
[5] Integration of Resources
[6] Unknown Attack (due to Zero Day vulnerability)

۲-۱- بیانِ مسئله

با نگاهی عمیق به سیستم های دفاعی موجودات زنده در طبیعت ، می توان از الگوهای موفّق زیستی در فرایند تشخیص و پاسخ هوشمندانه به حملات و رفتارهای محرّک زیست محیطی به خوبی الهام گرفت و در علم امنیّت شبکه و خصوصاً توسعه IDS ها به کاربست. کاری که هم اکنون دانشمندانی در جهان مانند نینا فِفِرمَن[۱]به دنبال کشف حقایق و رازهای طبیعت و کاربرد دستاوردهای حاصل شده در علوم کامپیوتر هستند. مثلاً برای حل پاره ای از مسائل بهینه سازی مسیریابی بسته ها در شبکه از مدل رفتاری مورچه ها در یافتن غذا و استراتژی خاص آنها در کشف مسیر رفت و برگشت طعمه به لانه و بالعکس که با ترشّح دنباله های ماده ی شیمیایی فِرومون در طول مسیر صورت میگیرد ، تقلید، مدلسازی ریاضی و نهایتاً مهندسی شده و الگوریتم[۲]آن نیز ارائه شده است. بنابراین الهام از رفتار های خاص هوشمندانه و ذاتی جانداران در طبیعت در موقعیتی خاص می تواند راهگشای مشکلی بزرگ باشد.

انسانها و بطور کلّی تمامی سیستم های مبتنی بر هوش طبیعی و مصنوعی [۳]بدون استثناء در تمامی امور پردازشی با چالش دسته بندی مواجه هستند. زمانی این چالش بیشتر می شود که نوع و الگوی داده ی ورودی به عامل هوشمند، جدید بوده و برای سیستم ناشناخته باشد بطوریکه سیستم تاکنون تجربه ی مواجهه با آن و تحلیل آنرا را نداشته باشد (دسته بندی بدون نظارت). بخصوص اگر حجم داده ورودی نیز بالا بوده و به پردازش آنی نیاز باشد سختی کار به مراتب بیشتر شده و چالشهای مهمتری همچون پردازش داده عظیم [۴]و وجود منابع پردازشی و حافظه کافی نیز بوجود می آید.

البته در کارهایی پژوهشی مانند [۱]محقّق امکان پیاده سازی و رفع چالش پیچیدگی محاسباتی و فضای حافظه لازم برای یک سیستم تشخیص نفوذ مبتنی بر سیستم ایمنی مصنوعی را در محیط ابر بررسی نموده و با کاربرد سازوکار انتخاب منفی در سیستم تشخیص عملکرد آنرا بهبود داده است. ارزیابی نتایج این تحقیق نشان میدهد که بحث چالش محاسباتی NSA و پیاده سازی موازی آن با راه حلّ محاسبات ابری می تواند مرتفع گردد. تشخیص نفوذ یک مسئله ی سخت [۵]دسته بندی می باشد. سختی این مسئله بدان علّت است که هیچ تابع/ فرمول ریاضی و آماری خاصی تاکنون اثبات نشده که بتواند بر زیر مجموعه مقادیرِ ویژگیهای ترافیک، اعمال شده و ویژگیها را به برچسب واقعی کلاس بدرستی نگاشت کند. علاوه بر این درجه ی سختی این مسئله با بالارفتن تعداد خوشه ها نیز بیشتر می شود. برای مثال خوشه بندی با $K > 2$ (k تعداد مرکزیّتها) یک مسئله $NP-Hard$ می باشد. هنوز راه حلّ جامع و پایداری از سوی محقّقان و متخّصصان ارئه نشده که در تمامی شرایط امنیّتی بتواند

---

[۱] Dr. N. Fefferman ، استادِ دانشگاه تِنِسی ایالات متّحده امریکا (University of Tennessee) و از پیشروان و محقّقان به نام در زمینه موضوعی بیولوژیِ فراگشتی و مدلسازی ریاضی از علم بیولوژی در جهت حلّ مسائل علوم کامپیوتر به شمار می رود.

[2] ACO (Ant Colony Optimization technique)

[3] AI based systems

[4] Big Data

[5] NP – Hard Problem

چالشهای مربوطه را رفع نماید. چالشهایی مانند تشخیص بلادرنگ ناهنجاری ، وجود خطاهای کاذب ، دقّت و بطور کلّی کیفیّت پایین دسته بندی. البته چالش غیر قابل اجتناب بودن از رخداد خطای دسته بندی و دلایل آن از سه دیدگاه نظری ، تجربی و فلسفی قابل بررسی و اثبات است. در این پژوهش چالش مذکور از هر سه دیدگاه بررسی شده است.

از طرفی جریان ترافیک شبکه به دلیل تنوّع الگوها و ابعاد بالای مسئله (تعداد ویژگیها) ، کار آنالیز و دسته بندی را در سیستم های تشخیص نفوذ بسیار دشوار می کند. امروزه IDS ها و سامانه های کنترل امنیّت شبکه تجاری و صنعتی بسیاری تولید می شوند که نقاط ضعف زیادی دارند. استفاده از این ابزارها محدود به شبکه هایی خاص با شرایطی خاص است و بسته به شرایط امنیّتی سازمان ها و شرکت ها می بایست پیوسته نظارت کاملی از سوی متخصصان امنیّت بر آنها صورت گرفته و بدرستی پیکربندی شوند. تحقیقات در حوزه ی تشخیص نفوذ بحث بسیار گسترده ای است و از جنبه های مختلف معماری ، تکنیک پیاده سازی ، متد و استراتژی تشخیص می توان آنها را بررسی نمود و تحقیق و توسعه داد.

سیستم ایمنی زیستی تاکنون در هیچ یک از کارهای پژوهشی با هدف تولید یک سیستم تشخیص نفوذ جامع به طور کامل شبیه سازی یا پیاده سازی عملی نشده ولی در تعدادی از مقالات مانند [۲ و ۳][۴۸] به وجود پتانسیلهای قابل توجّه آن در شناسایی و پیشگیری از نفوذ و لزوم انجام تحقیق در مورد امکان توسعه ی چنین سیستمی تاکید شده است. تعدادی از مقالات نیز مانند [۴] صرفاً با استفاده از مفهوم واکنش ذاتی سیستم ایمنی زیستی که رویکرد دسته بندی آن یادگیری نظارت نشده [۱]می باشد به ارائه روشی برای تشخیص نفوذ پرداخته اند. مصونیّت ذاتی صرفاً یک واکنش سریع بوده و توانایی بررسی دقیق و عمیق کشف نفوذ را ندارد بنابراین واکنش ایمنی تطبیق پذیر[۲] به سیستم ایمنی بدن کمک می کند تا حافظه ی خود را بهبود بخشیده و با تولید آنتی بادیهای بالغ و موثّر حافظه بتواند عامل نفوذی را کشف و اثر آن را در بافتهای سلولی بدن از بین ببرد. از اینرو مصونیّت تطبیق پذیر و خصوصاً متد انتخاب منفی که الهام گرفته شده از این پدیده ی طبیعی می باشد با هدف ارائه ی IDPS های کارآمد بیشتر مورد توجه پژوهشگران بوده است.

به منظور رفع چالشهای حوزه تشخیص نفوذ شبکه ، بیشتر کارهای پژوهشی مانند [۱۱-۵][۴۴] هر یک با رویکردی متفاوت به بهبود سیستم ایمنی مصنوعی و الگوریتم انتخاب منفی پرداخته اند. در تمامی این تحقیقات، مصونیّت ثانویه (واکنش ایمنی تطبیق پذیر در سیستم ایمنی زیستی) مبنای الهام بوده است.

در این پژوهش ، سعی نموده ایم با الهام از BIS و تجزیه و تحلیل دو واکنش ذاتی [۳]و تطبیق پذیر و با تکیه بر مفهومی نو به نام تئوری خطر[۴] ، نوعی حیات مصنوعی را در محیط نرم افزار متلب شبیه سازی نموده

---

[۱] خوشه بند

[2] Adaptive immunity Response
[3] Innate immunity Response
[4] Danger Theory

و چالشهای مسئله ی سخت دسته بندی ترافیک شبکه را در طول این حیات بررسی نماییم. به منظور انجام شبیه سازی ها از متدهای ایمنی مصنوعی (NSA و DCA) استفاده نموده و در طول این پژوهش آنها را بهبود و توسعه دادیم. در نهایت با کسب تجربه ی حاصل از نتایج آزمایشات در شبیه سازی ها، یک روش هیبریدی برای تشخیص آنومالی ترافیک شبکه با دو خط دفاعی پیشنهاد گردید که دارای ویژگیهای پایه ایِ یک سیستم ایمنی از قبیل خودیادگیری و امکان رشد و تعدیل پویای حافظه ، خود ایمنی [1]، خود ترمیمی[2]و خود سازماندهی[3] بوده و نمونه ای ساده از شبیه سازی کامل سیستم ایمنی زیستی می باشد.

سیستم تشخیص پیشنهادی ترکیبی بوده و از دو خط دفاعی تشکیل شده است. خط نخست دفاعی واکنش ایمنی ذاتی داشته و به صورت نظارت نشده / نیمه نظارت شده ، ترافیک شبکه را دسته بندی می نماید و خط دوّم دفاعی نیز با دقت بیشتر و با پتانسیل خود یادگیری، به دسته بندی نظارت شده ترافیک شبکه و شناسایی نفوذ می پردازد.

۳-۱- ضرورت انجام پژوهش

علم بر اساس آمار و احتمالات است. در علم هوش مصنوعی و رایانش امن (امنیّت) که هر دو گرایش هایی خاص از علوم کامپیوتر هستند در مورد مسائلی که با احتمالات حل شده و خروجی آنها عدد حقیقی (پیوسته) است قطعیّتی وجود ندارد. بدین معنی که اگر از علم آمار و احتمال برای پاسخ به یک مسئله در علوم کامپیوتر استفاده شود نتیجه حاصل شده به صورت قطعی (گسسته) نخواهد بود و اصولاً تجربه نشان داده که رویکرد های پیوسته همچون منطقِ فازی به جای منطق باینری در حل مسائلی خاص از علوم کامپیوتر که نیازمند پیش بینی مبتنی بر دانش از قبل آموخته هستند بهتر جواب میدهند.

علاوه بر این در حوزه ی امنیّت شبکه و تشخیص نفوذ ، رشد حملات شبکه و تنوّع الگوهای مخرّب ترافیکی که دستاورد هوش انسانی است NIDS ها [4]را با چالشهای بسیاری در شناسایی و برچسب زنی صحیح آنومالی رو به رو کرده است. در واقع، وجود این تنوّع باعث شده تا مسئله تشخیص نفوذ به صورت احتمالی حل شده و برچسب خروجی نمونه های آزمون صرفاً یک عدد حقیقی در بازه ی $[0.1]$ باشد بطوریکه از 0 تا 0.49 نشان دهنده نرمال بودن و از 0.5 تا 1 نیز معرّف آنومالی باشد.

در حالت کلّی مسئله ی تشخیص نفوذ را به دو شیوه پیوسته و گسسته می توان حل نمود. از طرفی به نظر میرسد به دلیل اینکه پدیده ی نفوذ دستاورد هوش غیر ماشینی است در مجموع رویکردهای پیوسته که خروجی

[1] Self – immunity
[2] Self – healing
[3] Self – organizing
[4] Network Anomaly detection system

آنها حقیقی مبنا بوده و به تحلیل انسانی در حل مسائل نزدیکترند مسئله ی تشخیص نفوذ را بهتر درک و حل می کنند.

به منظور حلّ مسئله دسته بندی و تشخیص آنومالی غالباً علم دانش کاوی [1]و از تکنیکهای یادگیری ماشین استفاده می گردد. این روشها در موارد محدودی توانسته اند برخی از چالشهای تشخیص نفوذ را رفع نمایند. در این پژوهش مسئله ی تشخیص آنومالی ترافیکِ شبکه را به شیوه ی پیوسته حل نمودیم به دلیل آنکه روش پیشنهادی ما در نهایت میبایست روشی ترکیبی باشد، از اینرو به نظر می رسد صرفاً تنها با این شیوه می توان نتایج خروجی دسته بندها (برچسب های حقیقی) را بر مبنای استراتژیِ تصمیم گیریِ مشخصی ترکیب نمود. بدین منظور در کلّیه مراحل تحقیق، متدهای ایمنی مصنوعیِ حقیقی مبنا [2] به کار رفته و در طول روند پژوهش بهبود و توسعه داده شده اند.

ضمن اینکه با ارائه ی روشی جدید با پیچیدگی سطح بالا [3]مبتنی بر BIS ، به عنوان یک سیستم خبره در راستای دستیابی به هدف اصلی در جهت رفع برخی از چالشهای زیر نیز کار شده است. از جمله :

- پیچیدگی محاسباتی فضا و زمان بالا در شناسایی و نیاز به منابع پردازشی و حافظه زیاد.
- بروز خطاهای کاذب .[4]
- تشخیص ناهنجاری های از نوع U2R و R2L [5]
- شناسایی بلادرنگ [6]نفوذ و اهمیّت برقراری نوعی موازنه [7]میان امنیّت و سرعت.
- نرخ تشخیص و دقّت دسته بندی پایین حملات.
- پردازش داده ی حجیم.
- کُنترل همروندی [8] تشخیص در حین یادگیری و بالعکس.
- جستجو و انتخاب زیر مجموعه ی بهینه ای از ویژگیها (کاهش ابعاد ، به عنوان یک مسئله ی سخت)
- دسته بندی نظارت نشده با $K > 2$ (تعداد خوشه ها بیشتر از دو ، به عنوان یک مسئله ی سخت)
- پیش بینی وقوع حملات.
- امنیّت خود سیستم تشخیص.

---

[1] Knowledge mining
[2] Real Valued based Artificial Immune Methods
[3] Sophisticated method
[4] False Positive . False Negative
[5] مجموعه ی حملات User to Root و Remote to Local ، چالشی که اکثر سیستمهای تشخیص آنومالی با آن مواجه اند.

[6] Real time
[7] Trade – off
[8] Concurrency control

در نتیجه لزوم پرداختن به تحقیقات بین رشته ای چون پژوهش جاری که در زیر مجموعه ی پژوهش در حوزه علم هوش محاسباتی و محاسبات نرم [1] قرار میگیرند از دیدگاه های مختلفی همچون نظری ، تجربی و فلسفی نیز قابل بررسی بوده که ضمن ارائه راه حلّ در خصوص بسیاری از مسائل سخت در زمینه علوم کامپیوتر مانند مسئله مهم تشخیص نفوذ ، پتانسیل بالقوّه ای در جهت تسریع نیل به اهداف چشم انداز توسعه علمی کشور و پیشبرد مرزهای دانش در خود دارد.

۴-۱- اهداف پژوهش

هدف اصلی این پژوهش ، برقراری نوعی حیات مصنوعی از طریق شبیه سازی مصونیّت های ایمنی ذاتی و تطبیق پذیر با کارکردی مشابه در BIS جهت دستیابی به یک سیستم تشخیص نفوذ پایدار و ایمنِ شبکه می باشد. به عبارت دیگر هدف ، رسیدن از ایمنی به امنیّت است.

انتظار می رود که برقراری چنین حیاتی در قالب یک سیستم تشخیص نفوذ بتواند به رشد مصنوعی این سیستم منجر شده و ضمن اعطای نوعی پویایی به آن، پتانیسلهای BIS را در شناسایی عوامل آنتی ژنیک در خود داشته باشد. لازمه ی وجود چنین حیات مصنوعی آنست که سیستم همواره در حال یادگیری و شناسایی بوده و این دو فرایند بصورت همروند و به موازات هم پیش بروند بطوریکه تولید و تکثیر تشخیص دهنده های بالغ و تعدیل حافظه و فضاهای خودی-غیر خودی هیچ موقع در سیستم متوقّف نشوند. این پایان نامه امکان ایجاد یک چنین سیستمی با دو خط دفاعی را از طریق برقراری چنین حیاتی ، بررسی ، اثبات علمی و شبیه سازی نموده و بحث پیاده سازی و عملی بودن این طرح و توجیه صنعتی آن موضوع و هدف این پژوهش نیست.

یک توصیف از این سیستم بدین صورت است که قادر است با نمونه های ترافیک مشکوک شبکه همانند آنتی ژنها در BIS رفتار نموده و با آنها زندگی کند در نتیجه عوامل سیستم که همان تشخیص دهنده های بالغ حافظه و در اصل ناظران درونی سیستم [2]هستند خواهند توانست به مرور زمان و با کسب تجاربِ حاصل از شناسایی نفوذ، رشد و نمو [3]نموده و ایمنی درونی سیستم را خود کنترل نمایند (خود- ایمنی). در اصل، این

[1] Soft Computing

[2] Probes

[3] واژه ی رُشد به مفهومِ تَعدیلِ این ناظران می باشد. تعدیلِ یک ناظر (تشخیص دهنده ی حافظه) به مفهوم آنست که پتانسیلِ آن تشخیص دهنده بر حسبِ شناسایی عوامل نفوذ و یا نرمال ، به مرور زمان ارتقا پیدا می کند. یعنی یا پتانسیلِ ناظر در جهت شناسایی عوامل نفوذ بیشتر شده و یا برعکس. کنترل این وضعیت (پتانسیل نرمال یا آنومالی) بر اساس پارامتری به نام AVote ممکن است که در فصول پنجم مطرح و اثبات شده است.

تشخیص دهنده ها هستند که کار پردازش را به شکل توزیع شده و مشارکتی [1] انجام داده و نوعی پویایی و در واقع حیات را به سیستم می بخشند.

ذکر این نکته ضروری است که اهداف معمول هر سیستم دسته بندی و تشخیص نفوذ ، ارتقا و بهبود عملکرد دسته بندی (افزایش نرخ های تشخیص و دقّت دسته بندی ، کاهش خطا) می باشد که البته جزو اهداف جانبی این پژوهش نیز هست.

۵-۱- مجموعه سؤالات و فرضیه های پژوهش

- **حیات مصنوعی** : دستیابی به درجه ویژه ای از مقاومت (امنیّت) و تضمین نسبی آن از طریق برقراری حیات مصنوعی و حفظ ایمنی درون سیستم تشخیص به چه صورت امکان پذیر است ؟ اصولاً این حیات مصنوعی چه نوع حیاتی است ؟ آیا روش هیبریدی پیشنهادی می تواند پایداری امنیّت شبکه را از نظر شناسایی ترافیک آنومالی آنگونه که کارکرد سیستم ایمنی زیستی میباشد در دراز مدّت تضمین نماید؟
- **هیبریداسیونِ تئوری خطر و انتخاب منفی** : از دیدگاه ایمنی شناسی ، مصونیّت های ذاتی و اکتسابی در BIS به صورت ترکیبی کار می کنند. آیا امکان ترکیب نسخه های بهبود یافته ی دو الگوریتم DC و NS به منظور ایجاد سیستمی با دو خط دفاعی وجود دارد ؟ در صورت مثبت بودن پاسخ ، روش ترکیبی تا چه اندازه می تواند در کاهش نرخ خطاهای مثبت و منفی کاذب موثر باشد ؟ علاوه بر این روش هیبریداسیون پیشهادی چه تاثیری بر بلادرنگ بودن تشخیص خواهد داشت ؟
- **کاهش ابعاد مسئله** : معمولاً استفاده از متد ها و استراتژیهای کاهش ابعاد مسئله در افزایش سرعت اجرا و دقّت دسته بندی تاثیر گذارند. بررسی این تاثیر در مورد متدهای دسته بندیِ ایمنی مصنوعی تاکنون در هیچ یک از کارهای پژوهشی انجام نشده است. به دلیل آنکه این متدها مسئله ی دسته بندی را به شیوه ای متفاوت از سایر دسته بندهای معمول و بر مبنای معیار فاصله بین نمونه ها حل می کنند بنابراین تاثیر کاهش ابعاد در این مورد هنوز یک ابهام به نظر میرسد. کاربرد الگوریتم جستجو و انتخاب ویژگی دَه - پا (CFA) در [۱۴] چه تاثیری بر عملکرد دسته بندی روش پیشنهادی می تواند داشته باشد؟ تاثیر این کاربرد در عملکرد هر یک از خطوط دفاعی سیستم پیشنهادی ، در تولید و تکثیر تشخیص دهنده ها و خصوصاً در فاز نمونه برداری از آنتی ژنها در الگوریتم خط نخست دفاعی (DCA) به چه میزان است؟
- **تخصیصِ سیگنالها** : آیا کاربرد الگوریتم حفره سیاه [2]در [۱۳] به عنوان یک خوشه بند می تواند در یافتن بهترین زیر مجموعه ویژگیها و نگاشت (تخصیص) سیگنالهای ورودی موثر باشد ؟ با توجه به

[1] Distributed and Cooperated approach (Decentralized)
[2] Black Hole Clustering Algorithm (BHA)

اهمیت تعیین سیگنالهای ورودی الگوریتم DC و تاثیر آن بر بلوغ و کیفیت مهاجرت سلولهای دندریت ، آیا ایده استفاده از یک خوشه بند مثل الگوریتم خوشه بند حفره سیاه در مقایسه با دو متد معمول نگاشت، تخصیص و محاسبه سیگنالهای ورودی (انتخاب ویژگی و استفاده از تجارب متخصّصان) می تواند عملکرد کلّی سازوکار DCA را از نظر بهبود و مهاجرت صحیح سلولها بهبود ببخشد ؟ اصولاً کدام یک از این سه روش پتانسیل بالقوّه ای را در تخصیص بهتر سیگنالهای ورودی DCA دارند ؟

- **شناسایی حملات ناشناخته :** شناسایی حملات ناشناخته به دلیل بروز آسیب پذیریهای روز صفر (حملات روز صفر) یک چالش بسیار بزرگ است که همواره درصد بالایی از رخداد خطای منفی کاذب بدین علّت است. در دراز مدّت ، روش هیبریدی پیشنهادی در شناسایی این حملات به چه میزان می تواند موفّق باشد؟ فلسفه ی وجودی این نوع حملات چیست و آیا اصولاً صد در صد می توان با رخداد خطاهای کاذب مقابله نمود و نرخ وقوع آنها را به صفر کاهش داد ؟ اگر خیر به چه دلیل ؟

۱-۶- مهمترین کارهای پژوهشی مطالعه و شبیه سازی شده به عنوان مبنای انجام پژوهش

به منظور انجام این تحقیق ، کارهای پژوهشی [۸-۶] [۱۵-۱۳] به طور کامل در نرم افزار متلب شبیه سازی گردیده اند.

۱-۷- روش انجام پژوهش

۱-۷-۱- روش گردآوری اطلاعات

- تحقیق و تفحّص در منابع پژوهشی ، مقالات ISI ، کُتُب و پایان نامه های دانشگاهای معتبر از طریق دسترسی به پایگاه های اطلاعاتی و علمی آنلاین مثل موتور جستجوی گوگل اسکولار [1].
- طرح پرسش (سوالات آگاهانه) از طریق پیام رسان علمی تحقیقاتی ریسرچ گیت [2] و انجام تبادلات علمی با سایر پژوهشگران بنام در حوزه علم ایمنی شناسی زیستی ، علوم امنیّت رایانه و هوش مصنوعی صرفاً در حوزه ی تحقیقاتی پایان نامه و کمک گرفتن از آنها در جهت پی بردن به مسائل و روشهای علمی در حلّ آنها زیر نظر اساتید راهنما.

[1] Google Scholar Search Engine
[2] https://www.researchgate.net/Ehsan_Farzadnia2

۲-۷-۱- قلمرو پژوهش

الف) موضوعی : امنیّت فناوری اطلاعات و ارتباطات

ب) زمانی : بررسی در بازه زمانی ۱۳۹۰ تا ۱۳۹۶ ، به دلیل آنکه موضوع مورد مطالعه [۱]، خاص بوده و در بازه زمانی مربوطه تحقیقات زیادی بر روی آن انجام نگرفته است ، در بعضی موارد همچون پایان نامه های دانشگاه های خارجی و یا کتب ، بازه زمانی فوق مربوط به ده سال گذشته می باشد.

ج) مکانی : دانشگاه صنعتی مالک اشتر تهران (پژوهشکده امنیّت اطلاعات و ارتباطات)

۳-۷-۱- کاربردهای پژوهش

ردیف	نام سازمان	نوع استفاده
۱	دانشگاه صنعتی مالک اشتر تهران	تحقیقاتی
۲	سازمان پدافند غیر عامل	سیاست گذاری در امنیّت شبکه ملّی
۳	شرکت ارتباطات زیر ساخت مخابرات ایران	کاربردی و توسعه ای
۴	هر سازمان یا شرکتی که به نوعی با شبکه درگیر است	

۴-۷-۱- روش و مراحل پژوهش

ماهیّت این تحقیق از نوع توسعه ای بوده و به روش تجربی و با استفاده از دادگان ترافیک شبکه در ابزار شبیه سازی متلب انجام شده است. به منظور انجام این تحقیق نیاز بود تا در سه حوزه ی علمی ایمنی شناسی زیستی ، هوش مصنوعی و امنیّت کامپیوتر مطالعات گسترده ای صورت گیرد. به دلیل ماهیّت زیستی این تحقیق ، پیش زمینه مطالعات تئوری آن قریب به بیش از شش ماه زمان بُرد و طی مراحل زیر به اتمام رسید:

- مطالعه عمیق در حوزه علم ایمنی شناسی زیستی و مشخصاً بررسی BIS و بسط نتایج مطلوب این بررسی به منظور الهام گیری ، ایده پردازی و مُدلسازی ریاضی آن.
- کار با تکنیکهای یادگیری ماشین ، دسته بندها ، خوشه بندها و آموختن علم داده کاوی پیشرفته در نرم افزار مهندسیِ متلب.

[۱] بررسی و ارزیابی امکان حل مسئله تشخیص آنومالی شبکه با الهام از سیستم ایمنی زیستی

- شبیه سازی دو الگوریتم ایمنی مصنوعی (انتخاب منفی و سلولهای دندریت) در نرم افزار متلب و بررسی امکان بهبود و ترکیب آنها با ایده پردازی و الهام از مصونیت های بیولوژیکیِ ایمنی ذاتی- تطبیق پذیر و ارزیابی نتایج بدست آمده در دادگان نفوذ.
- بررسی امکان ایجاد یک سیستم تشخیص با دو خط دفاعی بگونه ای که سیستم شبیه سازی شده حدالامکان کارکردها ، پتانسیلها و استراتژیهای دفاعی BIS را به ارث ببرد. سپس ارائه ایده مناسب جهت تشکیل ماژول تصمیم گیری نهایی و هیبریداسیونِ نتایج برچسب زنی حاصل از خطوط دفاعی و در نهایت مهندسی ایده های شبیه سازی شده به منظور ارائه روش هیبریدی پیشنهادی.

۸-۱- روشهای استفاده شده برای تجزیه و تحلیل اطلاعات و آزمونِ فرضیه ها

سیستم تشخیص نفوذ در واقع یک نوع دسته بند است. ارزیابی هر سیستم دسته بندی و کلّیه تکنیکهای یادگیری ماشین با معیارهای تشخیص و کارائی [۱]صورت میگیرد. ماتریس اغتشاش[۲] و شاخص منحنیِ ROC و سطح زیر آن (AUC) ، شیوه های رایج برای نمایش نتایج ارزیابی دسته بند و آزمونِ فرضیه هستند. در این پژوهش ضمن استفاده از این معیارها از دو معیار مهم ضریب همبستگی[۳]و استعداد تشخیص[۴] (CID) نیز جهت ارزیابی آزمون فرضیه ها استفاده شده است.

ضریب همبستگی یا cc به منظور پیش بینی عملکرد سیستم دسته بند تشخیص نفوذ در مواجهه با حجم ترافیک بالای شبکه مورد استفاده قرار میگیرد. همچنین معیار CID نیز در مواقعی کاربرد دارد که منحنی های دسته بندهای مورد ارزیابی همدیگر را در نقاطی قطع نمایند. در این گونه مواقع نمی توان با محاسبه بزرگیِ سطح زیر منحنی (AUC) ارزیابی صحیحی از عملکرد ها داشت.

جهت آزمون فرضیه ها و انجام آزمونهای مربوطه نیاز به ابزار آزمون و همچنین دادگان یادگیری ماشین می باشد. بدین منظور کلّیه شبیه سازی ها در نرم افزار متلب نسخه MATLAB R2017b. در رایانه ای با مشخصات فنّی زیر انجام شده و از دو دادگان استاندارد نفوذ ISCX NSL – KDD | UNB و UNSW – NB15 استفاده گردید. دادگان دوّم در سال ۲۰۱۵ از سوی محقّقان مرکز تحقیقات پیشرفته دانشگاه نیوساوت وِلز استرالیا ارائه گردید و حاوی حملات جدید و Benchmark می باشد. [۱۶] دادگان اوّل نیز نسخه بِروز

---

[۱] نرخ تشخیص مثبت و منفی صحیح (TP و TN) ، خطای مثبت و منفی کاذب (FP و FN) ، خطا (Error) ، صحّت (Accuracy) و دقّت (Precision) ، حساسیّت (Sensitivity) ، میانگین هارمونیک F – Measure ، Specificity.

[2] Confusion Matrix

[3] Correlation Coeficient (cc)

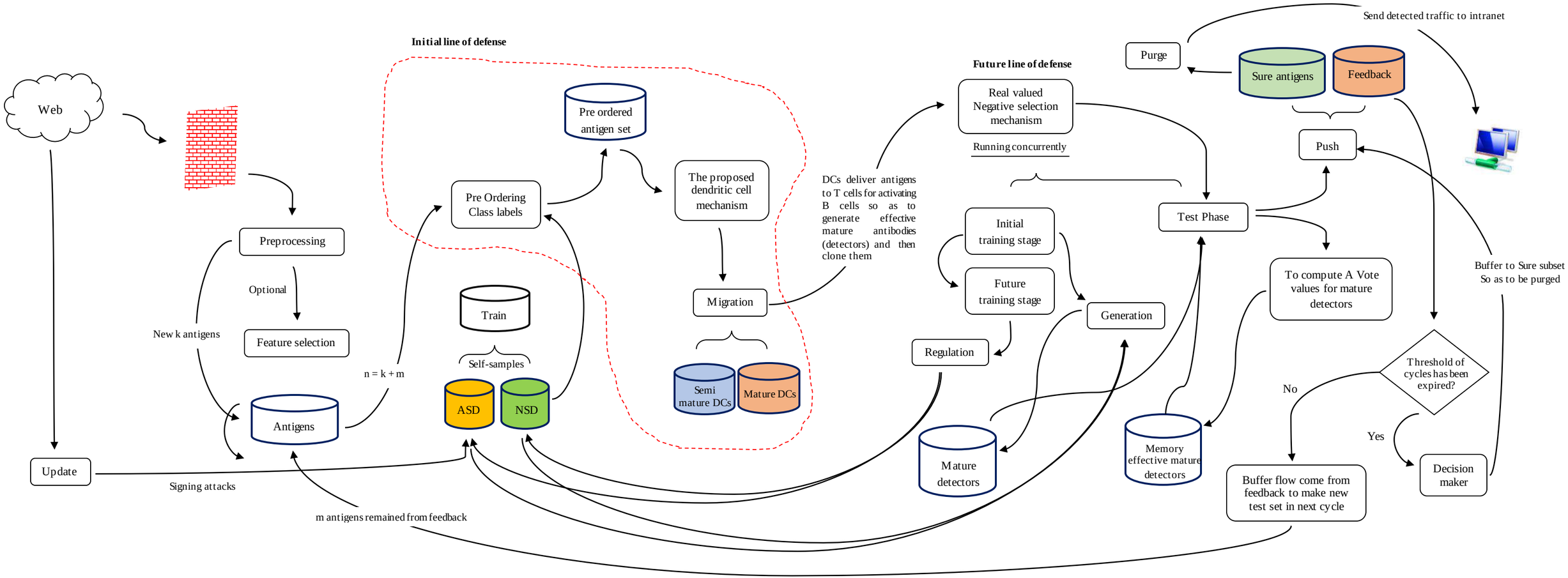
[4] Capability of Intrusion Detection

رسانی شده دادگان معروف KDDCup99 متعلّق به دارپای آمریکاست که توسط مرکز تحقیقاتی دانشگاه UNB کانادا پیوسته بروز رسانی و مورد استفاده قرار می گیرد.

جدول ۱- مشخصّات فنّیِ رایانه به کار رفته برای انجام شبیه سازی ها و آزمونها

Processor	*Intel® Core™ i5-3230M CPU @ 2.60 GHZ 2.60 GHZ*
RAM	*4.00 GB (3.87 GB usable)*
System Type	*64-bit Operating System, x64 based processor*
OS	*Windows 10 Enterprise © 2017*
MATLAB Usage Memory	*919*

۹-۱- چکیده تصویری


Initial line of defense
Future line of defense
Web
Preprocessing
Optional
Feature selection
New k antigens
Antigens
Update
Signing attacks
n = k + m
Pre Ordering Class labels
Pre ordered antigen set
The proposed dendritic cell mechanism
Migration
Semi mature DCs
Mature DCs
Train
Self-samples
ASD
NSD
m antigens remained from feedback
DCs deliver antigens to T cells for activating B cells so as to generate effective mature antibodies (detectors) and then clone them
Real valued Negative selection mechanism
Running concurrently
Initial training stage
Future training stage
Regulation
Generation
Mature detectors
Test Phase
Purge
Sure antigens
Feedback
Send detected traffic to intranet
Push
To compute A Vote values for mature detectors
Memory effective mature detectors
Threshold of cycles has been expired?
No
Yes
Buffer flow come from feedback to make new test set in next cycle
Decision maker
Buffer to Sure subset So as to be purged

۱۰-۱- مروری اجمالی بر محتوای فُصولِ بعدی

فصل نخست این پژوهش به بیان مسئله ، اهداف اصلی و روش تحقیق اختصاص دارد. فصل دوّم ، ادبیّات تحقیق، پیش زمینه زیستی پژوهش را از دیدگاه علم ایمنی شناسی به دقّت مرور و تشریح نموده است. بخش اوّل این فصل به این مبانی نظری اختصاص دارد. در بخش دوّم ، ادبیّات تجرِبی ، پیش زمینه تحقیقات صورت گرفته و تجربیّات محقّقان در استفاده از سیستم ایمنی مصنوعی و متدهای آن در جهت ارائه روشهای تشخیص نفوذ بررسی و ارزیابی شده و با دیدگاه نقّادانه ، چالشها و پتانسیلهای برخی از مهمترین کارهای پژوهشی این حوزه با هدف کسب پیش دانشی برای ارائه روش پیشنهادی واکاوی شده اند. فصول بعدی به ارائه ی روش تشخیص نفوذ پیشنهادی و گزارش آزمایشهای مربوطه اختصاص دارند.

در فصل سوّم ، در بخش نخستِ آن ، ضمن معرّفی تئوری خطر (DT) ، الگوریتم DC به منظور کاربرد آن به عنوان یک سِپَر دفاعی پیشگیرانه در شناسایی بالقوّه آنومالی ترافیک شبکه بررسی شده و در نهایت خط نخست دفاعی با الهام از مفهوم مصونیّت ذاتی، شبیه سازی شده است. همچنین در این فصل الگوریتم جستجو و انتخاب ویژگی دَه – پا جهت بررسی تاثیر کاهش ابعاد مسئله بر عملکرد روش پیشنهادی و الگوریتم BH به منظور کاربرد آن در نگاشت زیر مجموعه ویژگیهای بهینه (تخصیص) به سیگنالهای ورودی الگوریتم DC مورد بررسی واقع شده اند.

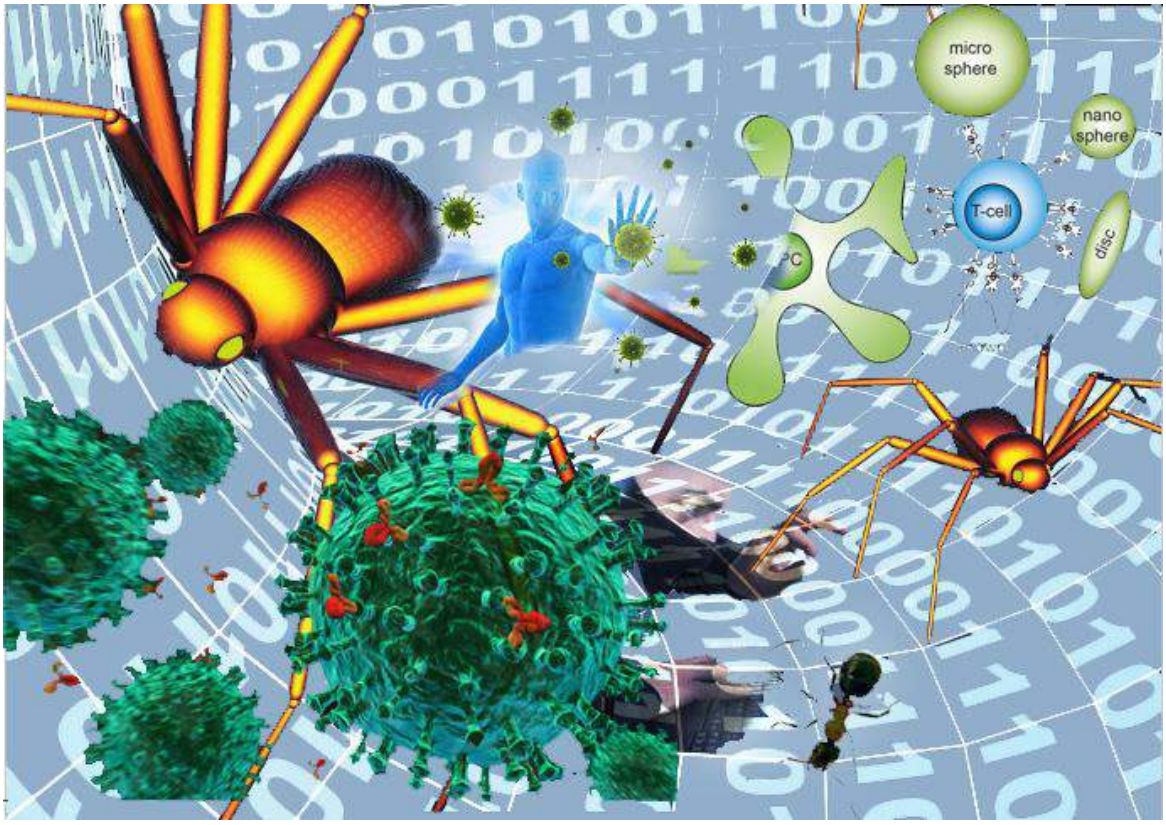
در بخش دوّم فصل سوّم ، در ادامه روند ارائه روش تشخیص نفوذ پیشنهادی ، خط دوّم دفاعی با الهام از واکنش ایمنیِ اکتسابی توسعه داده شده است. بدین منظور RNSA استاندارد (V-detector) شبیه سازی و مورد آزمون قرارگرفت و در نهایت ایده پیشنهادی برای ترکیب هیبریدیِ[1] نسخه های بهبود یافته این دو الگوریتم (DC و Real Valued NS) به عنوان خطوط دفاعی سیستم تشخیص نفوذ پیشنهادی ارائه و مورد آنالیز و ارزیابی قرار گرفته است.

فصل پایانی این پژوهش نیز به جمع بندی و بحث در خصوص نتایج حاصل از آزمایشات و یافته های پژوهش در فصول قبل اختصاص دارد. در ادامه ی این فصل به نتیجه گیری کلّی از این پژوهش و بررسی دستاوردهای حاصل از آن در پاسخ به پرسشهای مطرح شده در فصل اوّل پرداخته شده و چشم انداز توسعه تحقیقاتی - صنعتیِ سیستم هیبریدیِ تشخیص نفوذ پیشنهادی به عنوان کارهای تحقیقاتی آینده به خوبی ترسیم شده است.

[1] Hybridation

فصل دوّم

Research Literature



BIS

پتانسیل ها و چالشهای الهام گیری از سیستمِ ایمنیِ زیستی

جهت کاربرد در تشخیص نفوذ

بخش نخست
ادبیّاتِ نظری

۱-۱-۲- مقدمه

طبیعت همیشه الگویی مناسب برای انسان ها بوده تا بهترین ایده ها و مکانیزم ها را از دل آن بیرون کشیده و با مهندسی ، مدل حاصل شده را به صورت مصنوعی در ماشین آلات و دست ساخته های خود به کار ببرند. همیشه در طبیعت، کارایی ارجحیت دارد و هیچ الگوی پایداری نیست که نتوان از آن چیزی آموخت.

تفکّر بیولوژیکی منبعی سرشار از الهام گیری را برای مهندسان و محققان در راستای کشف راهکارهای نا ممکن برای حل مسائل پیچیده به همراه داشته است. از مدل حرکت حیوانات گرفته تا ساختار بدن جانوران و گیاهان، هر چیزی در موقعیتی خاص می تواند راهگشای مشکلی بزرگ باشد، تنها کافی است با نگاهی عمیق به طبیعت نگریست و پدیده مورد نظر را مدلسازی ریاضی نمود. الهام گرفتن از پدیده های طبیعی در علم شبکه نیز برای طراحی استراتژیهایی برای حمله و دفاع استفاده شده اند. با مشاهده رفتارها و خصوصیات بیولوژیک جانداران ، ایده هایی به ذهن می رسد که در صورت داشتن پتانسیل بهره برداری، می توانند بررسی و مدلسازی شوند. به روشهایی که به مدلسازی و مهندسی این مدلها کمک نموده و ابزاری برای ترویج و توسعه کاربرد تفکّر بیولوژیکی هستند NIC [۱]گویند. خروجی NIC یک الگوریتم زیستی می باشد.

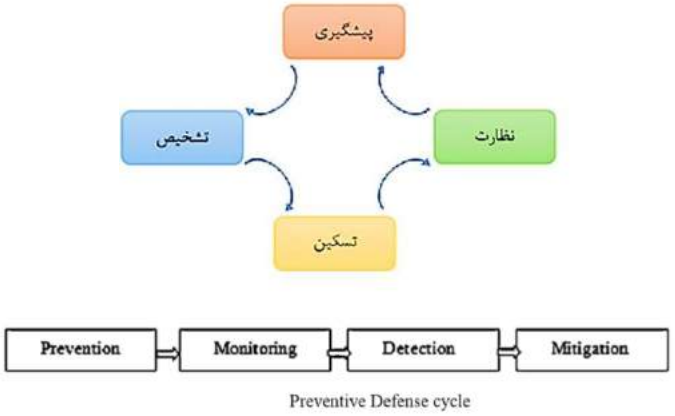
---

[۱] محاسبات الهام گرفته شده از طبیعت Nature Inspired Computations

از دیدگاه دانشمندان ، BIS دارای پتانسیل ها و قابلیتهای بسیاری می باشد. [۱۷] اصطلاح الهام گیری از بیولوژیک به تقلید از خصوصیات زیستی و رفتارهای خاص دفاعی جانداران ، مدلسازی و مهندسی آنها به منظور کاربرد در زمینه حل مسائل سخت و پیچیده اطلاق می شود. این رفتارهای مورد نظر ویژگیهایی مانند انعطاف پذیری ، استحکام [۱]، خود سازماندهی و پویایی در اتخاذ استراتژیهای دفاعی مقابله با تهاجم را باید داشته باشند. رفتار [۲]، در واقع پاسخ موجود زنده به محرّک محیطی و در اصل واکنشی در برابر کنش بیرونی است. در علم رایانش امن ، از جمله مسائل مهمی که توجه دانشمندان را به خود جلب کرده اند چالشهایی هستند که در انتهای فصل قبل بیان شدند. هدف عمده الگوریتمهای الهام گرفته شده از زیست در تشخیص نفوذ ارائه راهکارهایی برای رفع این چالشهاست.

۲-۱-۲- چرخه ی دفاع

بنا به استناد به مقاله [۱۸] که به بررسی رفتارهای پاسخ به نفوذ ، الهام گرفته شده از سیستم دفاعیِ بیولوژیکیِ گیاهان در برابر تهاجم پرداخته ، می توان پروسه ی دفاع را به صورت چرخه ی زیر در نظر گرفت و این پروسه را در امنیّت شبکه اقتباس نمود. ما این چرخه را چرخه ی حیاتِ دفاعِ پیشگیرانه نامیده ایم.



شکل ۱ – چرخه ی دفاع پیشگیرانه [۳]، برگرفته از [۱۸]

وجودِ این چرخه به عنوان یک استراتژی و در واقع الگویی برای توسعه سیستم های تشخیص و جلوگیری از نفوذ مبتنی بر اصولِ دفاع در عمق [۴]می تواند مطرح باشد. زیرا اصلِ دوّم امنیّت - دفاع در عمق همواره بر این

[1] Robustness
[2] Behavior
[۳] شکل بالا رویکردِ زیستی این چرخه و شکل پایین سیکل دفاع پیشگیرانه را از دیدگاه امنیّتی آن نشان می دهد.
[4] Defend in Depth

نکته استوار هستند که امنیّت کلّ سیستم وابسته به آنست که هر لایه ی امنیّتی وظیفه خود را به طور مستقّل انجام داده و در مجموع این لایه های امنیتی مکمّلِ برقراری و تضمین امنیّت باشند. (پ-الف-۱)

در نتیجه بر مبنای رویکرد دفاع در عمق هر کدام از این مراحل در چرخه بالا را می توان مکمّل یکدیگر دانست که می بایست در توسعه ی هر IPS/IDS این اصول را رعایت نمود. هدف اصلی در امنیّت شبکه ، حفظ این چرخه و پایداری آن است ولی مسئله ای که در این پژوهش مطرح می شود نحوه ی حفظ این چرخه با استفاده از سیستم ایمنی مصنوعی می باشد.

۳-۱-۲- مُدلسازی و مهندسی

به منظور الگوبرداری و توسعه ی گام به گامِ ایده های الهام گرفته شده از سازوکارهای بیولوژیک و کاربرد موثر آن در زمینه حل مسائل پیچیده گامهای اصولی باید برداشته شود. در زمینه امنیت اطلاعات هرگونه الهام گیری از مدلهای بیولوژیک به شکل فرایند سه مرحله ای زیر امکان پذیر خواهد بود. [۱۹] ساختار این پژوهش نیز اساس همین فرایند می باشد.

شکل ۲ – فرایند مدلسازی و مهندسی از مفاهیم بیولوژیک ، برگرفته از [۱۹]

ابتدا مشابهت های بین سیستم هدف و سیستم بیولوژیکی و ضرورتِ الهام گیری از آن انجام می شود ، مثل سیستم های تشخیص نفوذ در امنیت شبکه و محاسباتی[1] که ضرورتاً باید انجام شوند. سپس رفتارِ بیولوژیکی که می تواند پتانسیلی برای کاربرد در جهت حل مسئله باشد به درستی مدلسازیِ فکری می گردند و در نهایت مدل ایجاد شده مهندسی شده و روشی مبتنی بر رفتارهای زیستی ارائه می گردد. بنابراین ضروری است که اصول بیولوژیکی حاکم به درستی فهمیده شوند تا رفتارها به طرز صحیحی مدلسازی شده و مدل ایجاد شده

[1] Biological Immune Computing

را مهندسی نمود. مهندسیِ مدل ایجاد شده در نهایت منجر به ارائه ی الگوریتمِ زیست مبنا می گردد مانند الگوریتم ژنتیک.

۴-۱-۲- سیستم ایمنی[1]

سیستم ایمنی بیولوژیکی بدن انسان (HIS) ، سیستمی با ساختارِ زیستی است که پاتوژن [2]های خارجی و عوامل محرّک بیماری زای محیطی را به محض ورود به بدن شناسایی نموده و از آن در برابر این عوامل محافظت می کند. اولین لایه دفاعی از ورود پاتوژنها (عاملهای عفونت) به بدن به گونه ساده ای جلوگیری می کنند. این سطح شامل پوست است که برای بیشتر پاتوژنها سوراخ شدنی نیست و ترشحات مخفی مایع (مانند بزاق دهان) دارد که خصوصیات آنتی بادی (پادتَن) دارند. اگر یک پاتوژن برای شکستن این موانع و ورود به بدن مدیریت کند ، با سیستم ایمنی ذاتی(طبیعی) در دفعه بعدی روبه رو می شود.

سیستم ایمنی ذاتی (طبیعی) می تواند تفاوت بین خودی و غیر خودی (خارجی) را تشخیص دهد. در واقع می تواند مشخص کند که کدام سلّولها قسمتی از بدن بوده و کدامیک نیستند. [۵۸]

مکانیزم این سیستم هر اندام بدن را قادر می سازد تا در برابر هر تهدیدی که از محیط رو به رو شود پس از شناسایی صحیح نفوذ در برابر آن مقاومت و عکس العمل مناسب نشان دهد تا زنده بماند. ویژگی منحصر به فرد این سیستم و تمامی سیستم های ایمنی زیستی[3]، تشخیص عوامل خودی [4]از غیر خودی در بدن هر جانداری می باشد. [۵۱]

۴-۱-۲- سیستم ایمنی بیولوژیک (BIS)

هدف سیستم ایمنی زیستی ، مقابله با پاتوژنها و عوامل آنها (آنتی ژنها) و تشخیص هر گونه رفتار غیر خودی در بدن جاندار می باشد. این نوعِ تشخیص یکی از مهمترین و اساسی ترین وظایف این سیستم است. محققان در ایالات متحده شباهت قطعی میان این سیستم و مکانیزم امنیتی کامپیوتر پیدا کرده اند. شباهت به این ترتیب است که با استفاده از فراخوانی های سیستمی به عنوان داده بازرسی (audit trail) ، مجوزهای

---

[1] Immune System

[2] Pathogen ، عوامل بیماری زا هستند مثل انواع آنتی ژنها (عوامل نفوذی به سلولهای بدن) ، میکروبها ، ویروسها و غیره.

[3] سیستم ایمنیِ ذاتی ، ارگانیسمی وسیع و متشکل از گروه بزرگی از سلولهای ایمن که با مکانیزمهای مولکولی جهت محافظت در برابر آلودگی (پاتوژنهای بیماری زا) همواره در فعل و انفعال هستند. در این سیستم ها پروتئین ها عامل اصلی انتقال اطلاعات (یا سیگنالهای هشدار بیماری) بین سلولهای خودی هستند.

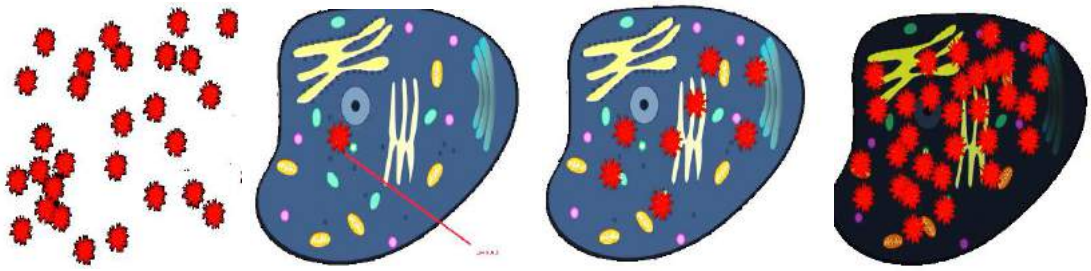
[4] Self – cells

فرایندهای پروفایلهای نرمال در حال اجرای سیستم به طور پیوسته بررسی می شوند. بطوریکه تفاوت فاصله میان الگوی این فرایندهای نرمال با الگوی آنها در حالت عادی آنها مقایسه می شوند تا آنومالی را تشخیص دهد.

آزمایشات نشان می دهند که سیستم تشخیص آنومالی مبتنی بر اصول ایمنی زیستی می تواند بسیاری از حملات را تشخیص دهد. مخصوصاً حملاتی که از مجوز فرایندها استفاده می کنند. سَندیپ کومار [1] دانشمندی است که در این زمینه تحقیقات زیادی انجام داده است. [۴۵]

پرسش - موقع نفوذِ ویروس به بدن چه اتفاقی می افتد؟ ویروسها ، باکتریها و عوامل ایجاد بیماری معمولاً به این سادگی نمی توانند از طریق نخستین لایه دفاعی بدن یعنی پوست وارد شوند ولی در صورت نفوذ ، ابتدا با سیستم ایمنی ذاتی روبه رو می شوند. [۵۸]

مطابق مطالبی که در [پ-الف-۱۰] در قسمتِ واکنشِ ایمنیِ سلّولی بیان شده است (شکل ۱۱) این سیستم از انواع مکانیزمها و سلولها (به عنوان عوامل دفاعی) مانند APC ها (یاخته خواران) تشکیل شده و می تواند تفاوت بین خودی و غیر خودی را تشخیص دهند. در بخش بعد بیان شده است که چگونه یاخته خواران، آنتی ژنها را به عنوان عامل نفوذی بلعیده وپس از نابودی اثرِ آنها، به فعالسازی سیستم دفاعیِ انطباقیِ بدن (ایمنی اکتسابی[2]) به عنوان پیشرفته ترین سطح دفاعی اقدام می کنند. تشریح دقیق سازوکار لایه های دفاعی BIS در مثال " قلعه" در قسمت زیر بیان شده است.



شکل ۳ – فرایند نفوذ ویروس به سلول ، تکثیر و نابودی آن (از راست به چپ)

وقتی ویروسی به هر طریقی وارد بدن شد و از سدّ لایه های دفاعی نیز عبور کرد بلافاصله وارد سلول هدف شده و در داخل سلول تکثیر می شود. با این فرایند تعداد بسیار زیادی ویروس مشابه تولید و بدین ترتیب همه با هم سلول را از بین می برند. با تخریب سلول، این ویروسها هر یک در بافت بدن به سمت سلولهای دیگری حرکت و به آنها نیز نفوذ کرده و خود را در آنجا تکثیر می کنند و اصطلاحاً بیماری ویروسی در بدن گسترش

[1] Sandeep Kumar
[2] Accuired immunity

می یابد (اِپیدِمی). در واقع تمام این ویروسها با هم موجب از بین بردن بافت سلولی جاندار شده اند و اینکار نیازمند فرایند خارق العاده تکثیری سلولی بوده است.

به منظور درک چگونگی سازوکار سیستم ایمنی زیستی بدن مثالی در زیر آورده شده است که شباهت های زیادی به BIS دارد. قلعه ها دژ های مستحکمی هستند که برای محافظت از اماکن مهم و استراتژیک در برابر نفوذ و حمله دشمن ساخته شده و در گذشته به کار می رفتند. این ساختار دارای سه سد (لایه) دفاعی قوی می باشد.

۱- دیوارهای بلند / خندقهای عمیق دور تا دور دیوارها که از بیرون قلعه حفر شده اند.
۲- سربازان : بالای قلعه نگهبانی می دهند.
۳- جاسوسان : امور را به شکل نامحسوس نظارت می کنند و در صورت لزوم در سطح بالایی واکنش نشان می دهند.

اولین لایه دفاعی وجود دیوارهای بلند با خندقهای عمیقِ حفر شده در دور تا دور قلعه اند که مانع محکمی در برابر نفوذ می باشند. سربازانی که بر روی قلعه نگهبانی می دهند و پُست مشخصی دارند به عنوان لایه دوم دفاعی شناخته می شوند. سربازان گهگاهی نیز به دشمن بیرونی پاسخ داده و مانع ورود دشمن احتمالی به قلعه می شوند. مشابه عملکرد یاخته خواران در BIS که در بالا بیان گردید ، اینکار اغلب با دستگیری همراه بوده و با سلاح صورت میگیرد و خساراتی به بار می آورد. ضمن آنکه اگر کوچکترین رِخنه ای در نفوذ به قلعه بوجود آمده باشد بهمراه تمام رویدادها به مسئولین بالاتر که جاسوسان هستند گزارش می کنند.

شکل ۴ – قلعه

سومین سد دفاعی مهم ، جاسوسان قلعه بوده که عواملی سرّی هستند که بر تمام فعالیتهای رخداده در محافظت از قلعه نظارت می کنند. در واقع همانطور که در ادامه خواهیم دید عملکردِ جاسوسان بسیار حیاتی بوده و شبیه به سیستم دفاعیِ واکنشِ تطبیق پذیر و واکنشِ ویژه به ترتیب در لنفوسیتهای T و B می باشد.

سیستم دفاعی قلعه شباهت زیادی به سیستم ایمنیِ زیر پوست بدن یک جاندار دارد. زیرا پوست بدن نیز به عنوان سدّ دفاعی مهم در برابر نفوذ بیماری ها ، باکتریها و ویروسها عمل می کند. پوست نیز از لایه های دفاعی

مختلفی تشکیل شده است. مویرگهای بر روی پوست ، عصبها ، مایعاتی جمع شده [1]و تمامی این اجزا با همکاری یکدیگر کار مراقبت از بدن را انجام می دهند.

۲-۴-۱-۲- سازوکارِ سیستم دفاع بیولوژیک

به طور کلی دو نوع مصونیّت (پاسخِ ایمنی) در بدن موجودات زنده وجود دارد : ایمنی ذاتی و تطبیق پذیر. سیستم ایمنی ذاتی در شناسایی نفوذ و نابودی آن سرعت عملِ بیشتری نسبت به نوع تطبیقی دارد. به طور مثال عملکرد دفاعی یاخته خواران که در بخش قبل بیان گردید روایتی از سازوکار ذاتی می باشد. در این بخش به تشریح سیستم ایمنی تطبیقی می پردازیم.

لنفوسیتها سلولهای پروتئینی و دو نوعِ T و B هستند. هر کدام واکنش ایمنی خاص خود را دارند. مشابه مفهوم دفاع در عمق که در سطوح مختلف سیستم ایمنی یک قلعه می توان مشاهده نمود BIS نیز با مصونیّت اکتسابی خود ، شامل سه نوع واکنش ایمنی می باشد که به آنها واکنش مُعیّن[2]نیز گفته می شود. (شکل ۵)

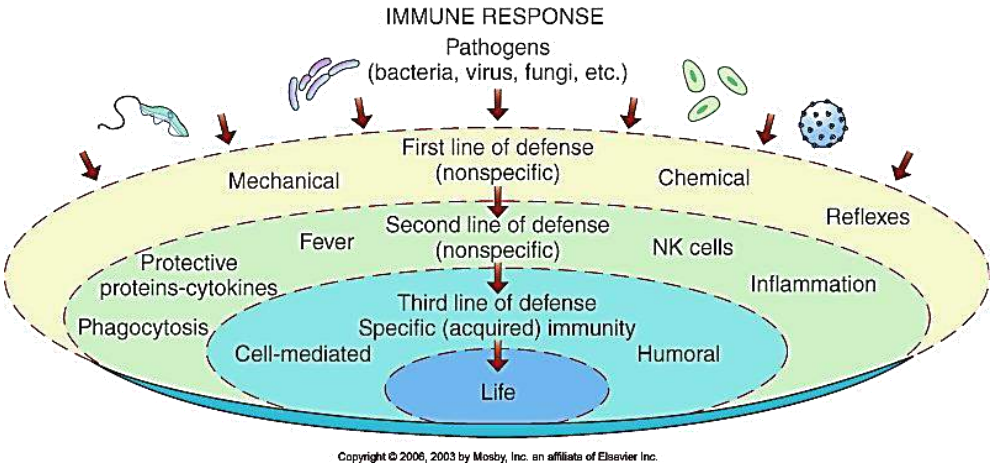
این واکنش ها عبارتند از [۵۸]

- واکنش هیومورال[3] (مربوط به لنفوسیتهای B)
- واکنش ایمنی ویژه [4]، مربوط به واکنش آنتی بادی ها در برابر آنتی ژنها.
- واکنش ایمنیِ سلّولی [5] (مربوطه به لنفوسیت های T)

شرح جزئیّات مربوط به این سه واکنش در [پ-الف-۱۰] موجود است.

[1] وجود موادی به شکل مایع بر روی پوست که از ورود باکتری و و آلودگی به بدن جلوگیری می کنند. بطوریکه PH بدن را متناسب نگه می دارند. (شبیه به کار یک سرباز در قلعه)

[2] Specific immune Response
[3] Humoral
[4] Specific immune response
[5] Cell – mediated



شکل ۵ – لایه های مختلف دفاعی سیستم ایمنی بیولوژیک در حالت کلّی (مشابهت با مفهوم دفاع در عمق امنیت شبکه)

۵-۱-۲- سیستم ایمنی مصنوعی

سیستم های ایمنی مصنوعی یک سری از متدولوژیهایی هستند که از سیستم ایمنیِ بدن انسان تقلید شده اند و شاخه ای از علم هوش محاسباتی هستند. [۲۰] AIS ، در واقع یک اصطلاح علمی است که به سیستمی اطلاق می گردد که از علم ایمنی شناسی[1] و با هدف حل مسئله الهام گرفته شده باشد. اساساً یک سیستم ایمنی مصنوعی از سه مرحله تشکیل شده است. [۱۹] شکل زیر اصطلاحات مربوط به مدلسازی محاسباتی مفاهیم ایمنی شناسی در جهت ایجاد سیستم ایمنی مصنوعی نشان می دهد.

Immunological Terms	Computational Modelling Terms
T cells, B cells, and antibodies	Detectors, clusters, classifiers, and strings
Self-cells, self-molecules, and immune cells	Positive samples, training data, and patterns
Antigens, pathogens, and epitopes	Incoming data, verifying data samples, and test data
String-matching rule Complementary rule and other rules	Distance and similarity measures Affinity measure in the shape–space

شکل ۶ – مدلسازی مفاهیم بیولوژیک در امنیت شبکه

نمونه هایی از کاربردهای شگفت انگیز سیستم های ایمنی مصنوعی در [۵۹] بیان شده اند. تمام کنشها و واکنشهایی که در سیستم های ایمنی الهام گرفته شده از بیولوژیک مطرح اند بدون هیچ کنترل متمرکزی رخ

[1] Immunology

می دهند بطوریکه رفتار های جمعی سلولها با همدیگر به طور مستقل از هم منجر از بین بردن آلودگی در بدن می شود. سوال اینجاست که این مکانیزمها چه گونه با اصطلاحات امنیت شبکه جایگزین و نگاشت گردند؟

پاسخ به ترتیب زیر است [۱۸] :

۱- ابتدا با کنار زدن فایروالها و گرایش به سمت کاربرد مفهوم پیشگیری از نفوذ در شبکه در اولین نقطه ورود ترافیک به شبکه
۲- تشخیص و ثبت هر فعالیت مشکوک صرف نظر از اینکه آیا نشانه حمله است یا خیر
۳- راه اندازی یک زیر سیستم به منظور آزمودن فعالیت های مشکوک و اتخاذ عکس العمل فوری برای مقابله با نفوذ

در سال ۱۹۹۴ دانشمندی به نام جِفری کِیفارت[1] از شرکت آی بی اِم یکی از دو طرح اولیه فوق را پیشنهاد داد. یک تشابُه میان سیستم ایمنی ذاتی با سیستمی که وی طراحی کرده بود و می توانست نفوذ را با دو متد تشخیص دهد اینگونه بود :

بررسی جامعیت برنامه ها و داده ها ، یک مانیتور کننده فعالیت که به فعالیت مشکوک عکس العمل نشان می داد. مشکلی که در سیستم وی وجود داشت کاربرد مفهوم خودی (Self) به تقلید از BIS بود زیرا کاربرد این مفهوم به دلیل اینکه فایلها توسط کاربران به طور پیوسته تعدیل و اصلاح شده و یا نرم افزار جدیدی نصب می شود یا از نصب خارج می شود امکان پذیر نبود. بنابراین محقق به جای این رویکرد اقدام به طراحی و راه اندازی سیستمی جهت مانیتورینگ و پایشِ ترافیک شبکه نمود و همزمان مورد سوّم در بالا را نیز اندیشید. به گونه ای که توانست از مفهوم APC ها در برنامه های تَله (طُعمه) استفاده کند. همچنین برای ماشین های شبکه نیز طرحی را ارائه داد که زمانی که گره ای با ویروسی مقابله کرد و الگوی مقابله را یاد گرفت به دیگر همسایه نیز در صورت ناشناخته بودن بیاموزد.

این تفکّر همان تفکّر مشارکت در تکثیر لنفوسیتهای فعال شده جهت فعال سازی سایر سلولهای مشابه به منظور مقابله با ویروس احتمالی در بدن است. خلاصه ای از کارهای انجام شده در AIS برای بهبود و توسعه IDS در بخش نهم از کتاب [۲۰] بررسی شده اند.

۶-۱-۲- رویکردهای نوین در تشخیص نفوذ

به طور کلی توسعه مدلهای مبتنی بر AIS را می توان به دو نسل تقسیم کرد. نسل سنّتی که مدلهای ساده الهام گرفته شده از BIS بودند شامل دو مصونیّت ذاتی و اکتسابی در جهت مقابله با نفوذ.

[1] Jefry kephart

اما در نسل دوم با توسعه ی تحقیقاتِ میان رشته ای عملاً موجب شکل گیری دو تئوری گردید. این تئوری ها در کنار روشهای نسل اول توانستند در بهبود کارائی تشخیص نفوذ مدرن نقش مهمی را ایفا نمایند. به تدریج از سال ۲۰۰۴ میان دانشمندان و محققان حوزه های ایمنی شناسی و علوم رایانه همکاری هایی انجام گرفت تا از این روشها به منظور توسعه الگوریتمهای الهام گرفته شده از ایمنیِ زیستی استفاده شده و در تشخیص نفوذ به صورت عملی به کار روند. [۲۰] برای کلیه الگوریتمهای فرا ابتکاری الهام گرفته شده از طبیعت رابطه زیر حاکم است. [۱۷] بطوریکه الگوریتمهای مهندسی شده از AIS را می توان در دسته دوم قرار داد.

$$\text{SI} - \text{Based} \quad \subset \quad \text{Bio} - \text{inspired} \quad \subset \quad \text{Nature} - \text{inspired}$$

از این رابطه به خوبی مشخص است که تمام الگوریتمهای الهام گرفته شده از طبیعت، زیستی نیستند. در این بین الگوریتمهایی وجود دارند که از ویژگیها و رفتارهای جاندارن با زندگی اجتماعی و از هوش مربوط به زندگی اجتماعی آنان [۱]جهت حلّ مسائل پیچیده استفاده می گردد. این الگوریتمها زیر مجموعه بیولوژیک نیز هستند. اما الگوریتمهای AIS و الگوریتمهایی مثل گرده افشانی گل[۲] و غیره ، نیز در دسته هوش ازدحامی قرار نمی گیرند حتّی با وجود اینکه الهام گرفته شده از بیولوژی هستند.

از طرفی برخی الگوریتمهای فَرا ابتکاری نیز موجودند که زیر مجموعه الگوریتمهای الهام گرفته شده از طبیعت بوده اما بیولوژیک نیستند و منشاء زیستی ندارند مانند : الگوریتمهای جاذبه ، حفره سیاه ، بیگ بنگ ، جستجوی هارمونی و ... این الگوریتمها مبتنی بر اصول و قوانین فیزیک و شیمی کار می کنند. ممکن است الگوریتمهایی در آینده بوجود بیایند که از هیچ یک از اصول و نظم موجود در گونه های طبیعت پیروی ننمایند بله ساخته و پرداخته بشر و یا حتی از قوانین فرا طبیعی پیروی کنند. در بین موارد بالا الگوریتمها یا سیستمهایی با منشاء زیستی مانند AIS و همچنین الگوریتمهای مبتنی بر هوش ازدحامی جانداران مهمترین رویکردهایی هستند که در حل مسائل تشخیص آنومالی و نفوذ به کار میروند. این مسائل عمدتاً همان چالشهای مهمی هستند که طبق بررسی ها و مطالب فصل قبل یک سیستم تشخیص نفوذ امروزی با آنها مواجه است.

۱-۶-۱-۲- مدلهای مبتنی بر مصونیّت

این قسمت به بررسی چهار رویکرد سیستم ایمنی مصنوعی اختصاص دارد. دو مدل نخست (انتخاب منفی و تکثیری) از مفاهیم واکنش ایمنی ذاتی و ایمنی اکتسابی موجود در سیستم ایمنی زیستی الهام گرفته شده و

[1] Swarm Intelligence

[۲] الگوریتم گل تلاش می کند از خصوصیات و ویژگیهای گرده در گیاهان گل تقلید کند و در پیوند با گل که سازگار با بعضی از حشرات گرده افشان است این کار را انجام دهد و یاد بگیرد.

جزو نسل نخست محسوب می گردند. سازوکار آنها بگونه ایست که از جنبه بیولوژیکی مکمّل یکدیگر می باشند. (سازوکار دو واکنش هیمورال و ویژه) موارد سوّم و چهارّم نیز نسل دوّم این مدلها بوده که برای بهبود و رفع نقاط ضعف مدلهای نسل اوّل و بیشتر جهت تکمیلِ تمام مصونیّتهای سیستم ایمنی مصنوعی ارائه شده اند. به عنوان نمونه تئوری خطر به منظور توسعه واکنش ایمنی ذاتی ارائه شده است که جزء جدایی ناپذیر سیستم ایمنی زیستی و به عنوان لایه نخست دفاعی جاندار می باشد. به شرح کامل چهار سازوکار ایمنی مصنوعی در [پ-الف-۱۳] پرداخته شده است.

- مدل نخست (نسل اوّل) : انتخاب منفی
- مدل دوّم (نسل اوّل) : انتخاب تکثیری
- مدل سوّم : تئوری خطر
- مدل چهارّم : شبکه ایمنی مصنوعی

جدول زیر نگاشت این مدلها را بهمراه خلاصه ای از کاربرد آنها در حل برخی مسائل حوزه امنیت بیان کرده است. [۲۰]

Immunological Concepts and entities	Immunity based models	Computer problems
Self/no Self T cells recognition	Negative selection algorithm	Errors, anomaly detection and change
Idiotypic networks Immunological memory, B cell	Immune networks theory	Supervised and Unsupervised learning
Clonal Expansion, maturation, B cell	Clonal selection algorithm	Search and Optimization
Innate Immunity	Danger Theory	Defense strategy

شکل ۷ - نگاشت مدلهای مبتنی بر مصونیت و مفاهیم و موجودیتهای ایمنی شناسی به حل مسائل امنیّت [۲۰]

۷-۱-۲- بررسی سیستم ایمنی مصنوعی و نگاشت مفاهیم آن به تشخیص نفوذ

روشهای نسل اوّل و دوّمِ AIS در عملکرد تشخیص دارای نوعی پویایی هستند. این پویایی همان توزیع وظیفه تشخیص نفوذ به آنتی بادیهای بالغ به شکل کنترل غیر متمرکز صورت میگیرد. این عوامل (لنفوسیتهای بالغ) پیوسته تکثیر یافته و تبدیل به حافظه می شوند و از طریق جریان خون همواره در حال سرکشی در بافت های بدن هستند. این تکثیر خود فرایندی دارد به این صورت که تنها آن دسته از لنفوسیتهای بالغی تکثیر پیدا میکنند که در طول عُمر خود حداقل یک بار تشخیص موفق انجام داده باشند و این کار (میزان تولید مثل و تکثیر) بستگی به میزان دفعات تشخیص موفّق آنتی بادیهای بالغ دارد که از آن در بخش های قبل این فصل تحت عنوان "حدّ وابستگی" یاد شد. ما از این مفهوم در فصول بعد در ارائه ی ایده پیشنهادی الهام گرفته ایم.

آنها در عین حال می توانند خود را احیاء نموده و یا با مرگ طبیعی و از پیش تعیین شده خود ، زمینه لازم برای بازتولید لنفوسیتهای هم نوع را فراهم کنند. این احیاء با جهش ژنتیکی ممکن است. علاوه بر این آنها می توانند به محض تطبیقِ اشتباه و چسبیدن به آنتی ژن خودی ، محکوم به فَنا شده و به مرگ برنامه ریزی دچار شوند. این دو ، خط قرمز یک لنفوسیت بالغ در بدن هستند یعنی عدم برخورد و تطبیق الگو با خودی ها و انجام حداقل یک برخورد موثر (با غیر خودیها) در طول عمر خود.نگاشت این مفاهیم به تشخیص نفوذ و تزریق نوعی حیات مصنوعی به دنیای ساز و کار این سیستم ها می تواند علیرغم پیچیده تر کردن کار تشخیص و کلاً افزایش ابعاد حل مسئله ، مزیّتهای بسیاری را با خود به همراه داشته باشد. کاری که ما در فصول بعدی با موفّقیت آن را انجام داده و به آثار مفید آن در دراز مدّت پی بردیم.

اگر وظایف زیر سیستم شناسایی و تطبیق الگو در یک سیستم تشخیص نفوذ امروزی که به شکل متمرکز بوده و دارای سیاستهای اتخاذ شده و کنترل و بروز رسانی قوانین آن توسط مدیر امنیت شبکه یک سازمان است را از حالت متمرکز خارج سازیم طبیعتاً این کار بوسیله ی یک سری از عوامل مانند سنسورهای تعبیه شده در نقاط حسّاس شبکه ممکن است صورت گیرد. یا حتّی خود IDS ها ممکن است در این نقاط به نحوی توزیع شوند که عملاً کار یک سامانه ی کشف و جلوگیری از نُفوذ توزیع شده را انجام دهند. آیا این پیکربندیها به اندازه کافی پویا خواهند بود که بتوانند با هوشمندی خود تصمیم بگیرند بطوریکه بدون نیاز به کنترل کافی توسط یک مدیر ، از عهده احیاء و ترمیم و بازسازی خود در مواقع ضروری بر آیند؟

بنابراین در حالت کلّی بحث اصلی در کاربرد مفاهیم و رویکردهای ایمنی مصنوعی در تشخیص نفوذ ، ایجاد سیستم هایی است که ضمن توزیع شدگی و کنترل غیر متمرکز ، وظیفه مهم تشخیص نفوذ را به شکل خود سازمانده ، خود احیاء[1]، خود آموز[2]انجام دهند. با این توصیف دو رویکرد نسل اوّل AIS را میتوان به صورت زیر با مفاهیم IDS تطبیق داد. البته در این راستا چالشهایی نیز مطرح می شود که در بخش انتهایی بررسی شده اند.

۱-۷-۱-۲- انطباق [3]

وقتی ویروسی وارد بدن می شود مانند آن است که حمله ای (ویروس، کرم ، بد افزارو ...) از طریق ترافیک شبکه، سیستم میزبان یا شبکه هدف را مورد هجوم قرار دهد. یک شبکه در صورت پیکربندی صحیح و رعایت استانداردهای امنیتی اَکتیو و پَسیو ، دارای پتانسیل لایه های دفاعی مختلفی می باشد. بدین معنا که مثل ساختار لایه های دفاعی بدن و سیستم ایمنی ذاتی و تطبیقی که در بخش قبل بیان گردید یک شبکه نیز از

[1] Self repairing
[2] Self learning
[3] Adaptation

این اولویت های دفاعی برخوردار بوده بطوریکه در لایه اوّل مشابه با عملکرد سیستم بدن ، لایه دیوار آتش و اینترفیس شبکه در روترِ لَبه وجود دارد. لایه ی دوّم که با رفتار واکنش سریع سلّولهای طبیعی مانند APC ها عملکردی مشابه با فاز شناسایی سیستم های تشخیص نفوذ شبکه دارند. لایه بعدی به عنوان مهمترین فاز مشابه فعالیتِ لنفوسیتها واکنش های معینی را بر مبنای از قبل آموخته های خود برای شناسایی دقیق عامل نفوذی و نابودی آن انجام می دهند. مطابق مطالبی که در بخش های قبل بیان شد ، لنفوسیتها به نوعی کار دسته بندی را انجام می دهند. در واقع وقتی که این سلولها در بافت های بدن در فعالیت هستند با هر سلولی نمی توانند برخورد کنند بلکه فقط در شرایطی این برخورد حاصل می شود که سر شاخکهای آنتی بادیها و پُرزهای برآمده آنتی ژنها با هم تطبیق داشته و این تطبیق از حدآستانه معین تعیین شده بهتر و بالاتر باشد. این مشابه یک دسته بند تشخیص انومالی می باشد که اگر الگوی ترافیک شبکه ی دریافت شده با الگوی مد نظر سیستم مستخرج از پایگاه داده یکسان بود تطبیق رخ داده و آلارم به صدا در می آید.

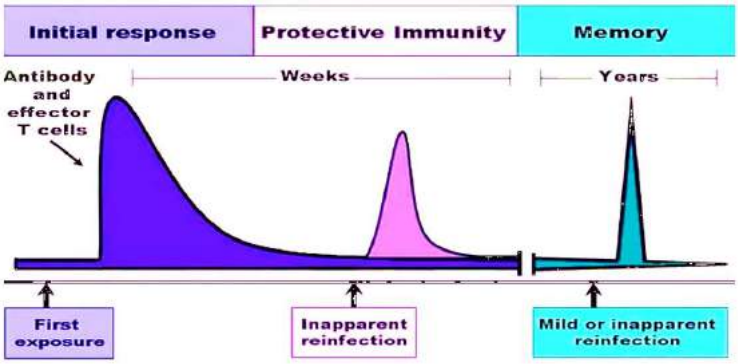
فورِست ، در رویکرد انتخاب منفی خود این مفهوم را به شکل رشته هایی بیان کرده است. مطابق شکل ۱۶ پیوست [پ-الف-۱۳] تطبیق الگو در صورتی رخ می دهد که دو رشته بیت منطبق ، در حداقل r بیت متوالی با هم Xor نموده و حاصل این عمل صفر باشد. بنابراین با توجه به این که در سرکشیِ لنفوسیتها در بافت های سلّولی بدن نیز به ندرت آنتی ژنی مشاهده می شود که با دیواره سلولهای تی برخورد نموده و تطبیق انجام شود، این عدم تطبیق به آن علت است که پایگاه پروفایلهای رفتاری سیستم تشخیص به اندازه کافی با رفتارهای نرمال بروز نشده اند که این منجر به عدم یادگیری کافی سیستم تشخیص و عدم تطبیق الگو شده است.

به عبارت دیگر مشابه لنفوسیتهای T که قبلاً در غُدّه ی تیموس با الگوهای خودی آموزش می دیدند در کاربردِ NSA در تشخیص نفوذ نیز کار تطبیق الگوهای این رشته های تصادفی با الگوهای نرمالِ ترافیکِ شبکه که قبلاً توسط A − NIDS [۱] گرد آوری و به صورت پروفایل هایی در آمده اند انجام می شود. در نتیجه NSA رشته هایی را به شکل تصادفی تولید می نماید تا با این پروفایلها چک شوند و از بین این رشته ها آن دسته که دارای الگوهای تطبیق کارآمدتری هستند برای تکثیر انتخاب می شوند. انتخاب تکثیری نیز به عنوان رویکرد نسل اوّل AIS به تکثیر آن دسته از لنفوسیتهای B مربوط است که آنتی بادیهای بر روی سطح این لنفوسیتها ، قبلاً کار شناسایی آنتی ژنها را با موفقیت انجام داده و فعّال شده باشند. حاصل این فعّال سازی تکثیر آن دسته از لنفوسیتهای از نوع B حافظه است که پس از بالغ شدن ، حداقل با یک آژ.غ.خد. برخورد انجام داده و در طول عُمر خود با سلولهای خودی تطبیق نخورده باشند. در سیستم تشخیص نفوذ نیز مفهوم انتخاب تکثیری و کلاً واکنش هیمورال سلولهای B حافظه می تواند الگویی برای دسته بندی IDS باشد. بطوریکه با وجود این دسته بندها ضمن کار دسته بندی ترافیک نرمال از آنومالی و ذخیره سازی الگوی آنومالی شناخته

[۱] سیستم های تشخیص ناهنجاریِ شبکه (ملاک تشخیص در این سیستم ها وابسته به امضاءهای حملات نبوده و صرفاً یک وضعیت غیر عادی از حالت نرمال می تواند منجر به شناسایی ناهنجاری گردد)

شده در پایگاه داده حملات و امضای آن ، سیستم تشخیص می آموزد که برای دفعات بعدی در مواجهه با حملاتی با چنین الگوهایی کار دسته بندی با دقّت بیشتری انجام شود. اصولاً به منظور نگاشت موثر این مفاهیم ایمنی زیستی به تشخیص نفوذ باید کارهای زیر بنایی زیادی صورت گیرد. کار مهمی که ما در این پایان نامه بدان پرداخته ایم. از طرفی BIS با واکنش های ذاتی و تطبیق پذیرِ خود به مرور زمان قادر است امنیت را تضمین و بیمه نماید. این عمدتاً مدیون وجود مفاهیمی مانند خود یادگیری و به یاد سپاری رخدادها و الگوهاست. این کار توسط لنفوسیتهای بالغ و آموزش دیده صورت میگیرد. بنابراین سیستم به مرور زمان بوسیله واکنشهای خود در برابر آنتی ژنها ضمن ثبت رخداد و نوع عامل غیر خودی می آموزد که مثلاً لنفوسیتها (نوع B) برای دفعات بعدی با مواجهه با چنین الگویی آیا الگوهای آنتی بادیهای بر روی سطح خود را اصلاح ژنتیکی یا تکثیر نمایند؟

نکته مهم - عمل اصلاح ژنتیکی در نسخه های حقیقی مبنای الگوریتم NS به مفهومِ تغییر اندکی در موقعیّت تش.د. مربوطه و پوشش حفره های جدید اطراف آن می باشد. همچنین در رویکرد حقیقی مبنا مطابق با آنچه که در فصول بعد در ارائه ی روش پیشنهادی انجام شده است تکثیر نیز به مفهوم تثبیت موقعیّت تش.د. مربوطه به کار رفته است. در نمودار زیر امکان رشد مصونیّت سیستم در برابر عوامل نفوذی بدن را نشان می دهد. بطوریکه سیستم در اوّلین رخداد مواجهه با آنتی ژنی خاص به دلیل اینکه الگوی متناسب با آن آنتی ژن خاص را تاکنون تجربه ننموده و از سوی سیستم شناسایی نشده است بنابراین تولید لنفوسیتهای دارای آنتی بادیهایی با الگوهای متناسب ، زمان بر خواهد بود. در نتیجه به محض واکنش سلولهای T و به جریان افتادن و تسریع روند تولید و تکثیر آنتی بادیهای بالغ و حافظه سازی به عنوان عوامل تاثیر گذار در امنیت سیستم (effector ها) ، خطر رفع شده و تطبیق ها با موفقیت در بافت سلولی صورت می گیرد. بعد از مدتی دوباره در صورتی که چنین خطری با همان الگو مجدداً بخواهد به سیستم نفوذ کند، زمان لازم (از لحظه فعالسازی و واکنش سریع لنفوسیت ها تا تولید و تکثیرِ آنتی بادیهایی با الگوهای منطبق و در نهایت نابودی آنتی ژن) کاهش یافته و به این ترتیب پس از چندین بار تجربه سیستم از کشف نفوذ و تطبیق الگوی موفق ، عملاً در مدت زمان بسیار کوتاه و بسیار سریع با عوامل نفوذی مقابله خواهد کرد.



شکل ۸ - نمودار مصونیّت در برابر بیماری (بیمه بدن) به مرور زمان و نقش لنفوسیتها [۵۸]

۲-۱-۸- طبقه بندی و آنالیز رویکردهای دفاعی

در جدول زیر مهمترین کارهای پژوهشی سالهای اخیر به ترتیب سال ارائه ، طبقه بندی و بررسی شده اند. در این مقالات ترکیبی از رویکردهای مختلف ایمنیِ زیستی به کار رفته اند. حاصل این ترکیب در برخی موارد مزیّت هایی نیز بهمراه داشته است. مزیّتهایی مثل بهبود نرخ تشخیص و افزایش دقّت با کاربرد الگوریتمهای زیستی در انتخاب ویژگی مانند CFA . در تعدادی از مقالاتی که بر روی برخی متدهای ایمنی مصنوعی کار کرده اند مثل انتخاب تکثیری ، انتخاب منفی و تئوری خطر ، مزیّت های مهم توزیع پذیری ، پویایی ، استحکام، انطباق و تنوّع مشاهده می شوند. بطوریکه با این مزیتها نتیجه اثر بخش تری را در کارآمدی عملکرد دسته بندی در شناسایی رفتار آنومال ارائه نموده اند.

جدول ۱ – طبقه بندی رویکردهای ترکیبی AIS و هوش ازدحامی به کار رفته در تشخیص نفوذ شبکه

مرجع	رویکرد تشخیص (الگوریتم استفاده شده)			رویکرد انتخاب/کاهش ویژگی	استراتژی جستجو	نتیجه بررسی
[21]	AIS (DCA و NSA)	بهره اطلاعات	-	-	-	۱- تغییر مقادیر بیت های $r-$ continues bit بر کارائی دسته بندی تاثیر بسیاری می گذارد. با افزایش مقدار r تا 6 ، به سختی دسته بند خوب کار می کند. ۲- DCA نتایج رضایت بخش $FAR=0$ (این بدان معناست که یکی از اهداف تشخیص آنومالی حاصل شده است). ۳- NSA خطای مثبت بالا و غیر مطمئن است بطوریکه با هر بار تغییر پارامترهای خروجی و مقدار r نتایج بسیار متفاوتی در دسته بندی مشاهده می شود.
[53]	مدل تشخیص نفوذ مبتنی بر ابر با استفاده از DCA به عنوان هسته اصلی IDS			Machine Learning Module	-	مشکلات IDS های سنتی : متد تشخیص واحد خود یاگیری ضعیف و توانایی انطباق ضعیف کنترل غیر دقیق دو معیار نرخ خطای مثبت و منفی مزایای IDS مبتنی بر ابر با استفاده از DCA : تشخیص بلادرنگ و موازی نفوذ.
[22]	CS (Clonal Selection)	-	-	-	-	مقایسه AIS با NN : CS مزیتهایی مانند حافظه ، یادگیری ، همبستگی و مشارکت ، بازیابی برای حل مسئله شناسایی و وظایف دسته بندی دارد. در مقایسه با BP (شبکه عصبی سنتی) عملکرد بهتری داشته است. سیستم ایمنی خصوصیات توزیعی ، پویایی ، انطباقی ، استحکام ، تنوع را دارد و می تواند خود به خود یاد بگیرد و تشخیص دهد و حافظه دارد. بیشتر سیستم های تشخیص نفوذ موجود خیلی از این ویژگیها را ندارد. بنابراین سیستم ایمنی مصنوعی قادر به تشخیص رفتار تشخیص آنومالی است که نتیجه (اثر) بهتری دارد.

۲-۱-۸-۱- پُتانسیلها

AIS دارای پتانسیلهای بالقوّه بسیاری است. مهمترین پتانسیل این سیستم قابلیت واکنش پویا و توان تطبیق پذیری بالای آنها با شرایط است که موجب اتّخاذ روشهای مقابله موثّر با عوامل نفوذ می شود. عواملی که موجب حفظ پویایی این سیستم ها شده اند قابلیت هایی مانند خود یادگیری ، توزیع شدگی ، قابلیت به یاد سپاری الگو و نحوه مقابله با عوامل غیر خودی هستند. در زیر برجسته ترین نقاط قوّت این سیستم ها که در رویکردهای جاری و سنّتی تشخیص نفوذ به عنوان چالشهایی مطرح بوده اند بررسی نموده ایم:

- ترکیب دو رویکرد خوب های شناخته شده [1] و بَدهای شناخته شده [2] در فاز آموزش تش.د. ها تا بُلوغ.
- توزیع شدگی: نوعی مشارکت توزیع شده در فعالیت تشخیص دهنده های بالغ.
- هوشمندی : چهار متد انتخاب منفی ، تکثیری ، تئوری خطر و شبکه ایمن موجب ایجاد سیستمی هوشمند شده و از دیدگاه تئوری چرخه دفاع را به طور کامل اجرا می کنند.
- قابلیت پیشگیری و پیش بینی وقوع حملات : تئوری خطر امکان شبیه سازی مص.ذت. و همچنین امکان پیش بینی خطرِ نُفوذ را بواسطه سیگنالها فراهم نموده است.
- تشخیص حملات ناشناخته.
- قابلیت تشخیص در حین آموزش و آموزش در حین تشخیص.
- امنیت خود سیستم تشخیص : سیستم های ایمنی مصنوعی در صورت کاربرد در IDS ، خود ایمن بوده زیرا در هر شرایط امنیتی ، توانایی خود ترمیمی و خود سازماندهی را دارند. از دیدگاه تئوری ، چهار متد AIS در کنار هم منجر به ایجاد سیستمی با قابلیتهای بسیار بالا می شوند.

۲-۱-۸-۲- چالشها

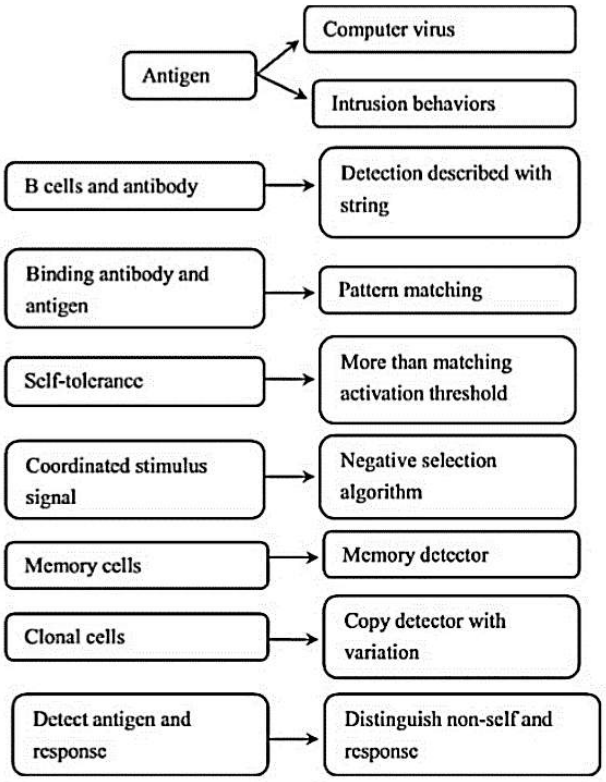
با نتیجه گیری ازبررسی های فوق ، مهمترین چالشهای سیستم های تشخیص نفوذ مبتنی بر ایمنی مصنوعی را می توان به شکل زیر بیان کرد:

- **پیچیدگی مُحاسباتی بالا و نیاز به منابع پردازشی و حافظه زیاد :** فرایند تولید تش.د. ها ، آموزش کافی به آنها و آزمون و خطا در حین آموزش ، تشخیص نفوذ و به خاطر سپاری رخداد و نوع الگوی حمله، بلوغ لنفوسیت ها ، تکثیر و همگی این فعالیتها منابع بالایی را می طلبند زیرا غالباً در شرایط احتمال وقوع حملات همه این فعالیتها با هم و مبتنی بر چرخه دفاع انجام می شوند.
- **خطا در تشخیص** : با تطبیق اشتباه الگوها منجر به مرگ ناخواسته لنفوسیت می شود و خطا را در پی دارد. اما پس از هر خطایی سیستم به خوبی قادر به تغییر ژنتیکی در لنفوسیت مُرده نموده و به این ترتیب یک خطا در سیستم دوبار تکرار نخواهد شد و این مزیت خوبی است.

[1] Known Goods
[2] Known Bads

- بروزرسانی های منظّم پایگاه داده های سیستم : کشف الگوهای غیر خودی و به خاطر سپاری آنها و تولید تصادفی الگوهای خودی به نحوی که تمامی رفتارهای صحیح را پوشش دهند. به عبارت دیگر ، دو پایگاه مهم خودی به ترتیب پایگاه حملات امضاء شده [1]و پروفایلهای سیستم هستند.
- **تشخیص بلادرنگ نُفوذ** : مِتدهای نسل اوّلِ AIS عمدتاً این مشکل را دارند ولی به مرور زمان و کسب تجربه و آموزش کافی در کوتاهترین زمان قادرند با واکنش سریعِ خود ، حمله را شناسایی و با آن مقابله کنند. البته رویکردِ تئوری خطر نیز ضعف های عمده این مِتدهای اصلیِ سیستم ایمنی را پوشش داده است. نقاط ضعفی مثل پیشگیری و پیش بینی به موقع نفوذ ، تولید تش.د. های موثر به میزان کافی و متناسب با توزیع و شدت حمله، بطوریکه که نوعی موازنه را بر قرار ساختند.
- **تعیین پارامترهای ورودی** : غالباً بسته به شرایط و کمیت نفوذ رخ داده در شبکه ، پارامترهای ورودیِ الگوریتمهای ایمنی مصنوعی به شکل بهینه ای باید تعیین گردند. البته از الگوریتمهای جستجوی مبتنی بر هوش ازدحامی نیز می توان در بهینه سازی پارامترهای ورودی بهره برد. از پارامترهای مهمِ متد انتخاب منفی می تواند به حد آستانه میزان وابستگی (r) جهت تطبیق الگو و حدّ آستانه لازم برای تکثیر در انتخاب تکثیری را می توان نام برد. بطوریکه با هر بار تغییر پارامتر ، نتایج متفاوتی در خروجی مشاهده خواهد شد. نگاشت میان مفاهیمِ ایمنی زیستی و امنیّت شبکه را می توان به صورت اینفوگرافیک زیر بیان کرد.

[1] Signature Base



شکل ۹ - اینفوگرافیکِ نگاشت مفاهیم ایمنی زیستی و مفاهیمِ امنیت [۲۲]

بخش دوّم
ادبیّاتِ تجربی

در این بخش تاریخچه ی پژوهشهای به انجام رسیده در زمینه ی AIS و کاربرد آن در شناسایی آنومالی شبکه و تجرِبه محقّقان را در پیشنهادِ سیستم های تشخیص نفوذِ مبتنی بر این سیستم بهمراه شیوه ی ارزیابیِ مقالاتی که از این سیستم در کارهای خود استفاده نموده اند به دقّت مُرور و بررسی و ارزیابی مقایسه ای گردیده اند.

۱-۲-۲- تاریخچه ای مختصر درباره پیدایش نفوذ و شیوه های شناسایی آن

تحقیق و توسعه ی IDS ها از اوایل دهه ۸۰ میلادی آغاز شد. زمانی که شبکه ی اینترنت هنوز به شکل امروزی آن توسعه نیافته بود. وجود آسیب پذیری در سیستم های عامل هایی که هنوز در آن زمان سیستم هایی نوپا بودند موجب شد تا مهاجمین به قصد سوء استفاده از یک سیستم به سیستم دیگر حرکت نموده و به این ترتیب تشخیص مهاجم / نفوذ ضرورت یافت. به تدریج با گسترش شبکه اینترنت، ابزارهای حمله نیز در دسترس عموم قرار گرفتند و انواع حملات پدید آمدند و امنیّت به عنوان یک علم مطرح شد. پس آغاز عصر نفوذ را باید ۱۹۸۵ دانست. زمانی که برای نخستین بار دانشمندان امنیّت کامپیوتر بر سر راهکارهای مقابله با نفوذ اقداماتی را انجام دادند.

با گذشت زمان کشف آسیب پذیری های جدید در هسته برخی از سیستم های عامل از سوی محقّقان و سوء استفاده گران موجب گردید تا این سوء نیّت (شَرّ) در شبکه های رایانه ای از فردی به فرد دیگر و از شبکه ای به شبکه ی دیگر بِخَزَد. این خَزِش تا به امروز نیز ادامه یافته و روز به روز به پیچیدگی آن افزوده می شود.

گذشته از مسائل تاریخی اصولاً یک اِکوسیستم شبکه ایمن باید سه اصل امنیّتی دسترس پذیری ، محرمانگی و جامعیت (CIA) را برای هر سیستم و کاربران آن به ارمغان آورد. [۶۵] از بین این سه اصل مهم ، دسترس پذیری و جامعیّت اهمیت فراوانی در تشخیص نفوذ شبکه دارند. از طرفی توسعه منابع شبکه ، خطری پنهان را برای تضمین این اصول امنیّتی به دنبال دارد. با توسعه این منابع روشهای تشخیص نیز بالاجبار باید پیچیده تر و منعطف تر شده و امنیّت شبکه و بطور کلّی تفکّر چگونگی مقابله با حملات می تواند چالشهایی را در پی داشته باشد.

۲-۲-۲- نفوذ و دفاع

به منظور دستیابی به درجه ویژه ای مقاومت در برابر نفوذ و تضمین نسبی امنیّت در یک شبکه ، استفاده از دیوار آتش و سایر مکانیزم های جلوگیری از نفوذ مانند اعمال سیاستهای امنیتی در روتر میزبان یا لَبه شبکه به شکل سنّتی آن دیگر کافی نبوده بطوریکه در طول دو دهه گذشته تابه حال این نیاز احساس شد تا از سیستم های دیگری به نام سیستم های تشخیص نفوذ استفاده گردد.

IDS/IPS ها صرفاً یک لایه امنیّتی از لایه های مختلف به کار رفته در استراتژی دفاع در عمق شبکه هستند. البته بحث دفاع در عمق از این جهت مطرح می شود که این سیستمها به عنوان بازوی اصلی تشخیص و مقابله با حملات عمل نمی کنند بلکه در کنار سدّ دفاعی اصلی شبکه یعنی فایروالها (دیوارهای آتش) ، پراکسی ها ، سرورهای احراز هویّت / احراز اصالت و سایر ابزارهای محافظتی به عنوان مکمّل تضمین سیاستهای امنیّتی شبکه عمل می نمایند. بنابراین کاربرد این سیستمها برای نخستین بار به این دلیل توسعه پیدا کرد که فایروالها

اغلب قادر به شناسایی الگو حملاتی با رفتارهای ناهنجار و ناشناخته نبودند و این نقطه ضعف دیوارهای آتش سَفت افزاری [1]موجب میشد که نفوذگر بتواند با دور زدن قوانین پیکربندی شده فایروال ، از سدّ دفاعی فایروال و از لایه دفاعی دوّم یعنی آنتی ویروس و برنامه های امنیّتی کاربردی تحت کنترل سیستم عامل نیز عبور کرده و به سیستم قربانی دسترسی کامل (روت) پیدا کند. در بخش نخست این فصل در قسمت چرخه ی دفاع بر مبنای زیستیِ آن تاکید گردید. در واقع IDS/IPS ضمن اینکه به عنوان یک لایه ی امنیّتی ، خود جزء کوچکی از استراتژی دفاع در عمق در شبکه هستند بلکه به نظر میرسد که این استراتژی را باید در توسعه موتور داخلی آنها و بر مبنای چرخه دفاع به کار ببرند. این نکته حائز اهمیّتی است. ما استراتژی دفاع در عمق را در توسعه هسته اصلی روش تشخیص نفوذ پیشنهادی به کار بردیم.

۳-۲-۲- آنومالی چیست ؟

آنومالی [2]، مشاهده ای است که از سایر مشاهدات انحراف بسیار زیاد و بیش از حد معمول تعیین شده (حد آستانه آنومالی) داشته باشند. طبق این تعریف ویژگیهای هر الگوی داده ایِ غیر معمول که توسط سیستم تشخیص نفوذ به ثبت می رسد هر گاه با ویژگیهای خوش تعریف و سابقه الگوی نرمال داده ها مطابقت نداشته باشد نشان دهنده یک نوع آنومالی است. [۲۳] گرچه آنومالی توسط محققان بسیاری بسته به کاربرد تعاریف متفاوتی ارائه شده است اما گفته «هاوکینگ» در مورد آنومالی که در سال ۱۹۸۰ ارائه شد اینگونه است:

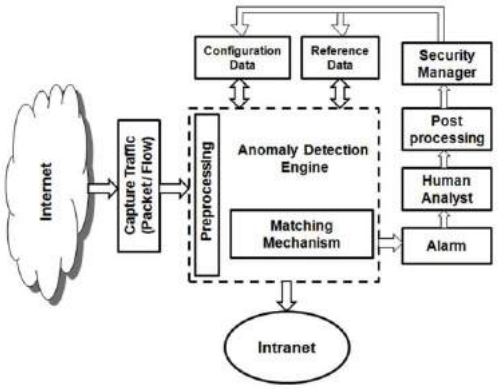
« آنومالی مشاهده ای است که انحراف بسیار زیادی از سایر مشاهدات (نمونه های داده) داشته باشد بطوریکه این شک را بر انگیزد که آن مشاهده با مکانیزم متفاوتی تولید شده است.» انواع آنومالی ها بنا به استناد به [۲۴-۲۳] به شرح [پ-الف-۱۱] به سه دسته کلی تقسیم می شوند.

۴-۲-۲- رویکردهای توسعه سیستمهای تشخیص آنومالی

تعریف - سیستم تشخیص نفوذ ، سَفت افزاری است که ترافیکِ شبکه را در مبادی ورودی به شبکه/ میزبان به دقّت آنالیز نموده و الگوهای مشکوکی را که ممکن است منجر به حمله شوند کشف می کند. در حوزه امنیت شبکه ، آنومالی مفهوم نفوذ را دارد اما در سایر علوم از آن به عنوان بی نظمی یاد میشود که ممکن است مطلوب نیز باشد. سیستمهای تشخیص آنومالی بر مبنای دانش [3]حاصل از تجارب مکتسب از یادگیری الگوهای رفتاری داده نرمال کار می کنند. [۲۵]

[1] Hardware & Software
[2] Anomaly
[3] Anomaly based Intrusion Detection System (AIDS)

برای مثال وقتی سیستم تشخیص آنومالی بسته هایی که آدرس مقصد یا پورت مقصد آنها نامعتبر است را مشکوک در نظر می گیرد ، در واقع دانشی از این الگوی ترافیک داده ای نامعتبر کسب خواهد کرد که در مواقع بعدی شبیه به این وضعیت به فاز تشخیص آنومالی سیستم کمک خواهد کرد. شکل زیر معماری پایه هسته یک سیستم تشخیص نفوذ شبکه مبتنی بر تشخیص آنومالی (رفتار ناهنجار) را نشان میدهد.



شکل ۱۰ – معماری کلّی یک سیستم تشخیص آنومالی ، برگرفته از [۳۸]

گرچه رویکردهای توسعه ی A – NIDS مختلفی وجود دارند ولی در حالت کلی تمامی آنها از مراحل پایه زیر تشکیل شده اند :

- پارامتر سازی (PreProcessing Phase): نمونه های ترافیک شبکه مشاهده شده به صورت یک پروفایل پارامترسازی می شوند. این گام آماده سازی و تخصیص ویژگیها می باشد.
- گام آموزش (Training Phase) : تشخیص رفتار عادی از غیر عادی در این گام صورت میگیرد و مدل متناظر ایجاد می شود. این فاز را با عنوان گام پردازش نیز می شناسند. تکنیکهای مختلفی در پردازش استفاده می شوند که در زیر به صورت بولتی مشاهده می کنید.
- گام تشخیص (Detection Phase): پس از آنکه مدل متناظر با رفتار در گام آموزش (فاز پردازش) ایجاد شد ، ترافیک سِنس شده (مشاهده شده) بر اساس این مدل ارزیابی می گردد. اگر میزان انحراف فراتر از حد آستانه معین بود یک آلارم جِنریت می شود.

طبق مقاله [۳۸] سیستم های تشخیص آنومالی براساس نوع پردازش به سه شاخه اصلی دسته بندی می شوند:

- تکنیکهای مبتنی بر آمار (Statistical based Techniques)
- تکنیکهای مبتنی بر دانش (Knowledge based Techniques)
- تکنیکهای مبتنی بر یادگیری ماشین (ML based Techniques)

ضمن اینکه ابزارهای نرم افزاری تشخیص نفوذ تجاری نیز مانند Tripwire، Bro، Snort، ISS که مبتنی بر کاربرد هستند نیز موجودند. این ابزارها سربار محاسباتی زیادی داشته و بر کارائی و بازدهی سیستم عامل تاثیر منفی می گذارند. این سیستم ها هر یک برای کاربرد خاصی طراحی شده اند تعدادی از آنها مانند Bro مبتنی بر شبکه هستند و تعدادی نیز مانند Tripwire مبتنی بر میزبان. این ابزارها نمی توانند رفتار آنومالی را شناسایی کنند. اما از زمان تشخیص بلادرنگ برخوردارند زیرا از یک سری قوانین و سیاستها تبعیت می کنند. برای مثال Tripwire یک ابزار بررسی جامعیت فایل است و تحت کنترل سیستم عامل کار می کند. به عنوان نمونه ای دیگر ، ISS یک سیستم تشخیص نفوذ مبتنی بر هردو میزبان و شبکه است که به صورت بلادرنگ کار می کند اما نمی تواند آنومالی را تشخیص دهد. Bro نیز ابزاری مبتنی بر شبکه است که بر اساس سیاست خود به بررسی صحّت پروتکلها در شبکه به صورت بلادرنگ می پردازد. امروزه تقریباً بیشتر سیستم های تشخیص نفوذ رویکردی ترکیبی دارند و به این دلیل نیازمند منابع محاسباتی و حافظه ای بالایی هستند. ترکیب رویکردها ضمن پیچیده شدن و بروز چالش مزیّتهایی را به همراه دارد که انتهای فصل بیان شده اند.

۵-۲-۲- وظایف سیستم های تشخیص نفوذ

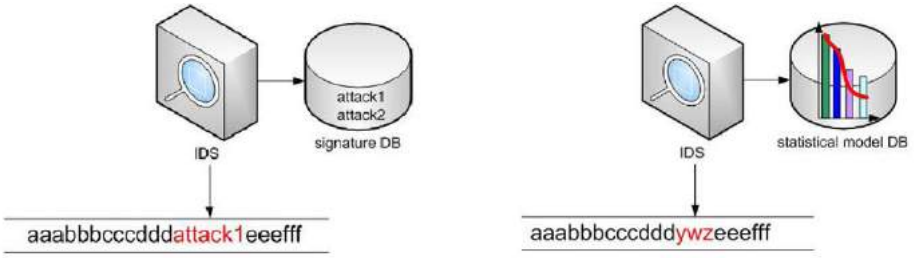
وظایف اصلی سیستمهای تشخیص نفوذ به شکل زیر است. [۶۳] هدف از عمل به این وظایف حصول اصول سه گانه امنیّت (محرمانگی ، دسترس پذیری و جامعیّت) می باشد.

- پایشِ مستمر فعّالیتهای سیستم و کاربر (پایِشِ رخدادهای شبکه از طریق سِنسورها و نظارت بر فعالیتهای سیستم عامل)
- پیش پردازش و اعمال فرایند کاهش ویژگی و حذف نویز.
- اعمالِ داده کاوی بر ترافیک شبکه و کشفِ دانش.
- بروز رسانی پایگاه امضاءهای حملات
- حفظ جامعیت فایلهای خودی (مخصوصِ HIDS ها)
- پاسخ خودکار به تشخیص فعالیتها (IPS)
- گزارشگیری از نتیجه تشخیص (صدور هشدارها و عکس العمل مناسب برای رفع به موقع آسیب پذیریها با اطلاع رسانی به مدیر سیستم جهت پیکربندی و اعمال سیاستهای مناسب)

۶-۲-۲- طبقه بندیِ سیستمهای تشخیصِ نفوذ از جنبه های مختلف

تاکنون دسته بندی های متعددی از روشهای تشخیص نفوذ ارائه شده اما هنوز هیچ طبقه بندی کلی از آن پذیرفته نشده و پیوسته در حال تغییر و تحوّل بوده اند. تکنیکهای تشخیص نفوذ از نظر متدولوژی آنالیز و

پردازش به دو دسته کلّی مبتنی بر سوء استفاده (مبتنی بر امضاء) و مبتنی بر کشف ناهنجاری تقسیم بندی شده اند. به زبانِ عامیانه ، رویکرد شناسایی نخست را اصطلاحاً بدهای شناخته شده و دوّمی را نیز خوبهای شناخته شده نامیده اند. (شکل زیر)



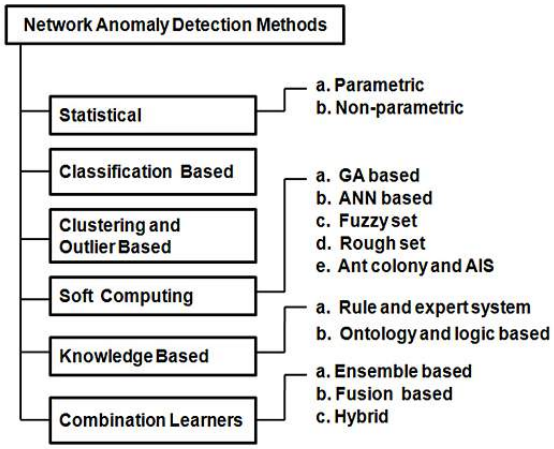
شکل ۱۱ - سمت چپ ، متدولوژیِ تشخیص مبتنی بر سوء استفاده و سمت راست متدولوژیِ مبتنی بر تشخیص ناهنجاری

تشخیص سوء استفاده بر پایه تکنیکهای تطبیق الگو است و سیستم تشخیص نفوذ شامل یک پایگاه داده امضاهای حملات شناخته شده بوده و تلاش می کند که این امضاها را با ترافیک آنالیز شده مطابقت دهد. هرگاه یک تطبیق یافت شد به منزله نفوذ قلمداد شده و یک هشدار به صدا در خواهد آمد. به طور رسمی این متدولوژی بر پایه مدل $M_S \subseteq I_D$ می باشد. (I_D کل حملات و M_S نیز حملات امضاء شده هستند) اگر یک عنصر ورودی مشکوک i عضو M_S باشد آنگاه IDS هشدار تولید خواهد کرد. از جمله چالشهای مهم این متدولوژی ، عدم بروز بودن پایگاه امضاهاست که منجر به عدم شناسایی حملات ناشناخته و بالا رفتن نرخ خطای منفی کاذب می گردد. از سوی دیگر در رویکرد تشخیص ناهنجاری ، به دلیل آنکه آمارِ رفتارهای نرمال معمولاً از رفتارهای ناهنجار بیشتر است لذا حجم کاری سیستم های تشخیص مبتنی بر سوء استفاده کمتر می باشد. [۲۳] در واقع در رویکرد مبتنی بر تشخیص ناهنجاری ، IDS ، همواره به کسب دانش حاصل از شناسایی رفتارهای عادی و نرمال سیستم و ذخیره این رفتارهای نرمال (پروفایل سازی) مشغول می باشد. مهمترین چالش این متدولوژی نیز بالا بودن نسبی نرخ خطای مثبت کاذب آن می باشد. مهمترین مزیّت آن نیز شناسایی نسبی حملات ناشناخته و تاحدودی غلبه بر حملات روز صفر است.

یکی از چالشهای تشخیص نفوذ ، امکان ترکیب این دو متدولوژی می باشد که به نظر میرسد نتایجِ این ترکیب مفید باشد. تکنیکهای Hybrid و Ensemble نمونه هایی از استراتژیها برای ترکیب این سیستمها هستند. در این رابطه ، سیستم های ایمنی مصنوعی ذاتاً دارای پتانسیلهای بالقوّه زیادی هستند.

۷-۲-۲- طبقه بندی سیستمهای تشخیص نفوذ از نظر منبع اطلاعات

سیستم های تشخیص را می توان بر اساس معیارِ منبعی که مورد پردازش قرار می دهند به دو گروه مبتنی بر میزبان و مبتنی بر شبکه به صورت جدول زیر تقسیم بندی و پتانسیلها و چالشهای هر یک را جداگانه بررسی نمود. به طور کلّی از دیدگاه کلان ، IDS ها را می توان بر اساس معیارهای مختلفی مثل تکنیک پیاده سازی، روش و متدولوژی تشخیص ، منبع اطلاعاتی مورد پردازش (شبکه یا میزبان) ، استراتژی تشخیص ، تشخیصِ بلادرنگ به طبقه های مختلفی تقسیم بندی نمود. (شکل ۱۳) همچنین سیستم های مبتنی بر تشخیص ناهنجاری شبکه (A – NIDS) نیز از نظر روشهای تجزیه و تحلیل رفتار نرمال تا کسب دانش و یادگیری با تکنیکهای موجود در شکل ۱۲ ، قابلیّت توسعه را دارند. سیستم ایمنی مصنوعی و متدهای آن که در این پژوهش استفاده شده اند، زیر مجموعه جدیدی از محاسبات نرم محسوب می گردند.

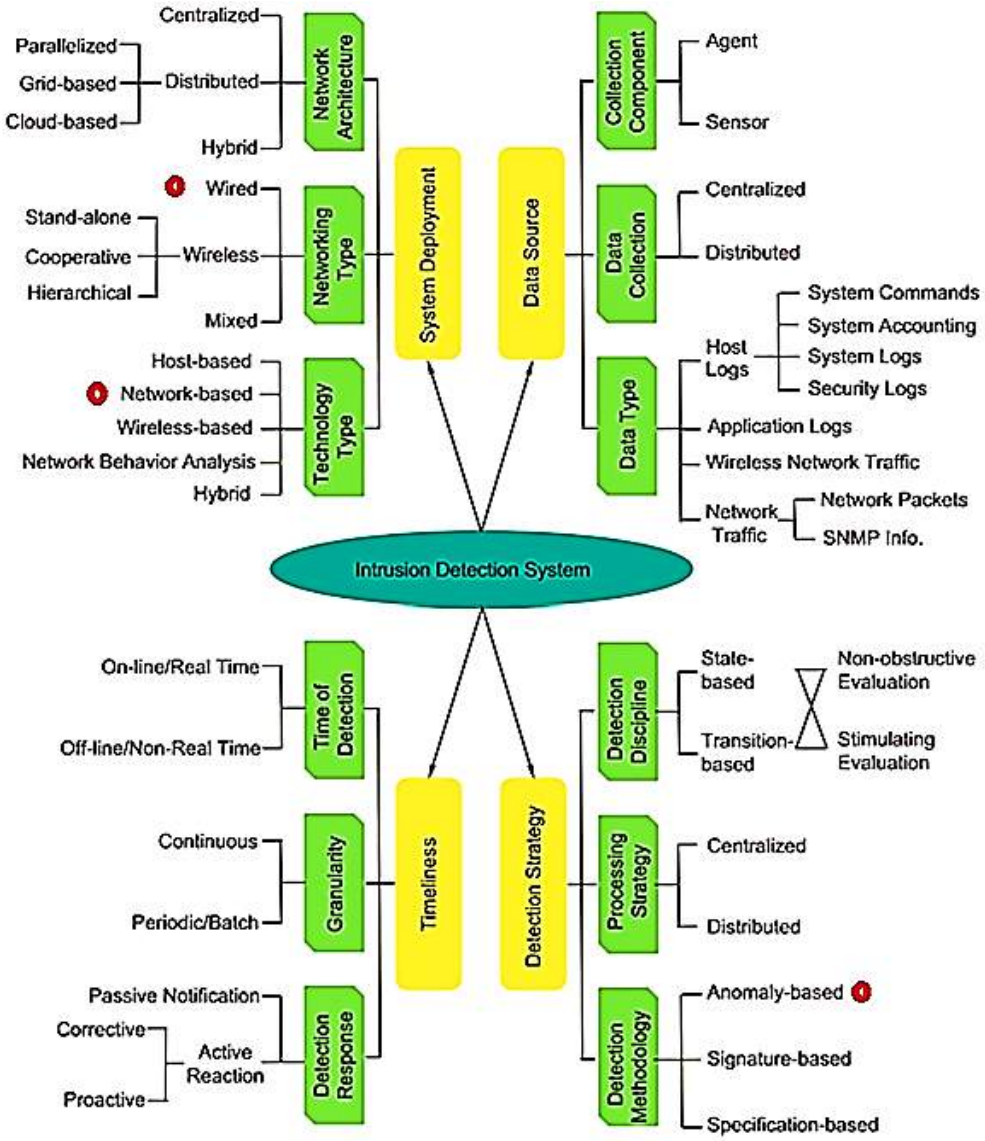


شکل ۱۲ - روشهای مورد استفاده برای تشخیص ناهنجاری شبکه و توسعه ی A – NIDs ها

جدول۲ - مزیتها و معایبِ سیستم های تشخیص نفوذ بر مبنای معیار منبع اطلاعات

رویکرد	مزایا	معایب
HIDS	۱-قابلیت نشان دادن سریع شکست یا موفقیت یک حمله ۲-نظارت سطح پایین ۳-تشخیص و واکنش بلادرنگ ۴-قابلیت عمل در محیطهای رمز شده ۵-اثربخشی و اقتصادی بودن	۱-امکان غیرفعال شدن سیستم در بخشی از حمله ۲-نیاز به انباره زیاد برای ذخیره اطلاعات ۳-سربار محاسباتی برای میزبان
NIDS	۱-قابلیت تشخیص حملاتی که سیستمهای مبتنی بر میزبان آنها را از دست می دهند چون ترافیک شبکه را در لایه انتقال نظارت می کنند. ۲-دشوار کردن حذف شواهد توسط مهاجم ۳-تشخیص و واکنش بلادرنگ ۴-قابلیت تشخیص حملات ناموفق و سوء قصدهای مخرب ۵-عدم تداخل با عملکرد معمولی شبکه	۱-عدم قابلیت عمل در محیطهای رمز شده مثل شبکه خصوصی مجازی [1] ۲-نقاط کور شبکه که قادر به دیدن آنها نیستند. ۳-عدم عملکرد صحیح در ترافیک سنگین

[1] VPN



شکل ۱۳ - اینفوگرافیک دسته بندی کلّی سیستم های تشخیص نفوذ با معیارهای مختلف ، برگرفته از [۲۶]

۸-۲-۲- معیارهای مقایسه و ارزیابی سیستم های تشخیص نفوذ

در مقایسه و ارزیابی عملکردِ سیستم های تشخیص نفوذ دو جنبه کلیدی وجود دارد. یکی کارائی فرایند تشخیص[1] و دیگری هزینه عملیات [2]. بنابراین به هر دو جنبه بدون دست کم گرفتن هر کدام از آنها باید توجه شود. در سیستم های تشخیص نفوذ ، داده های حمله یا آنومالی به عنوان داده مثبت (+) و داده های نرمال

[1] Performance
[2] Operation cost

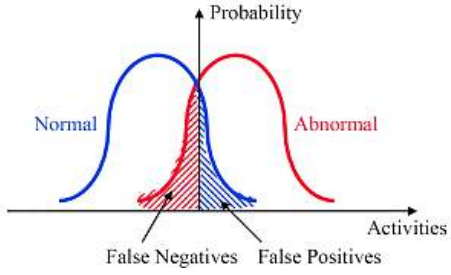
به عنوان داده منفی (-) قلمداد می شوند. به علاوه دسته بندی در این سیستمها می تواند درست و یا نادرست باشد و خطا در دسته بندی ترافیک شبکه و تشخیص رخ دهد. تکنیکهای داده کاوی و انتخاب ویژگی به منظور کاهش خطا می توانند استفاده شوند. به طور کلی با توجه به وجود دو کلاس تشخیص (درست و نادرست) ، چهار حالت بوجود می آید که مطابق با ماتریس درهم در جدول ۳ با اندازه گیری این حالتها و ارزیابی نتایج حاصل از آن می توان کارائی سیستم تشخیص نفوذ را ارتقا داده و پیش بینی های لازم را در جهت بهبود نرخ تشخیص صحیح سیستم انجام داد. این چهار حالت عبارتند از :

- مثبت واقعی : رویداد آنالیز شده به طور صحیح به عنوان نفوذ تشخیص داده شود.
- مثبت کاذب: رویداد آنالیز شده از منظر امنیتی بی خطر است اما به عنوان مخرب تشخیص داده شود. به عبارت دیگر رویداد واقعاً نرمال بوده باشد اما اشتباهاً توسط سیستم تشیخص به عنوان رفتار آنومالی (نفوذ) تشخیص داده شده باشد.
- منفی واقعی : رویداد آنالیز شده به طور صحیح به عنوان بی خطر و نرمال تشخیص داده شود.
- منفی کاذب : رویداد آنالیز شده مخرب است اما به عنوان بی خطر تشخیص داده شود. یعنی رویداد واقعاً نفوذ می باشد ولی سیستم به دلیل خطا آن را به اشتباه به عنوان نرمال تشخیص داده باشد. این حالت بسیار اهمیت دارد و بالا بودن مقدار آن در سیستم های تشخیص لطمات جبران ناپذیری را ممکن است به سیستم وارد سازد.

جدول ۳ – ماتریس درهمریختگی (حالتهای مختلف تشخیص داده ها)

پیش بینی	کلاس واقعی	حالت
نفوذ	نفوذ	True Positive
نفوذ	نرمال	False Positive
نرمال	نرمال	True Negative
نرمال	نفوذ	False Negative

شکل زیر تمایز این حالت ها را به وضوح نشان می دهد. ناحیه سمت چپ تشخیص به عنوان بی خطر و ناحیه سمت راست تشخیص به عنوان مخرب را نشان می دهد و منحنی سمت چپ رفتار نرمال و منحنی سمت راست رفتار غیر نرمال یا در واقع نفوذ را نشان می دهد.



شکل ۱۴ – هشدارهای واقعی و کاذب

وجود ناحیه قرمز رنگ هاشور خورده در سمت چپ نمودار بسیار برای سیستم خطرناک است و در هر سیستم تشخیص نفوذ می بایست سعی شود تا حجم نواحی هاشور خورده و مخصوصاً ناحیه قرمز هاشور خورده کاهش یابد. از بین این چهار حالت فوق دو حالت منفی کاذب و به خصوص حالت مثبت کاذب (نواحی هاشور خورده نمودار فوق) بسیار اهمیت دارند. روابط مربوط به معیارهای سنجش و ارزیابی دسته بندی در [پ-الف-۱۲] معرفی و تحلیل شده اند [۲۴].

۹-۲-۲- مبانی تجربی پژوهش

توأم با پیشرفت فناوری نفوذ و وابستگی زیرکانه حملات به فناوری و از سوی دیگر وجود انکار ناپذیر آسیب پذیریها (مانند روز صفر) در سیستمهای ارتباطی قاعدتاً می بایست پا را فراتر نهاده و بر طراحی و به کارگیری رویکردهای دفاعی متفابل هم از نظر پیشگیری و هم از نظر دفاع و مقابله ی به موقع با حملات ناشناخته چاره ای اندیشید. امّا توسعه رویکردهای امنیتی سازگار و منعطفی که بتوانند حداقل یک گام جلوتر از توسعه و پیشروی دانش نفوذ هکرها باشند نیز چالش بس بزرگی است. رفع این چالش نیازمند سطح دانش بسیار بالایی است که به نظر میرسد باید تلفیقی از هوش ذاتی و اجتماعی موجودات زنده در طبیعت و هوش مصنوعی ساخته و پرداخته بشر باشد. به عقیده محقّقان علم محاسبات نرم ، این هوش ذاتی در سیستم ایمنی زیستی نهفته است. **پرسش**: تفاوت بین ایمنی و امنیّت چیست ؟

بنا به استناد به نظر پروفسور زابرودسکی [۶۴] دانشمند برجستهٔ حوزه ایمنی شناسی در رابطه با پاسخ به پرسشی آگاهانه با مضمون فوق ، امنیّت ، درجه ای ویژه از مقاومت است که در اثر واکنش طبیعی سیستم دفاعی بدن جاندار (BIS) در برابر نفوذ عوامل آلوده مانند باکتری زا و میکروب زا (آنتی ژن ها [۱]) به بافت های سلّولی، در بدن شکل میگیرد. در حالیکه ایمنی یا مصونیّت ، متدی برای محافظت از یک ارگانسیم مانند بافت های سلّولی بدن می باشد. در واقع ، سیستم ایمنی بدن انسان با برقراری ایمنی درون سیستم به درجه ویژه ای از مقاومت می رسد که امنیّت نام دارد. این دیدگاه زیستی امنیّت و ایمنی به دنیای امنیّت فناوری اطلاعات - تشخیص نفوذ شبکه نیز قابلیّت انطباق و اقتباس دارد. در اثر شکل گیری این مقاومت، بدن با تولید آنتی بادیها از طریق فعالسازی گلوبولهای سفید خون (لنفوسیت[۲]های T و B) واکنش تطبیق پذیر از خود نشان داده و با آنتی ژن مقابله می کند. این فرایند هوشمندانه بوده و بسته به الگو و نوع آنتی ژنی که به بافت های بدن آسیب رسانده ، آنتی بادی هایی با الگوهایی منطبق با آن تولید و در بدن تکثیر می شوند تا به بهترین

[۱] به عوامل نفوذ غیر خودی از طُرق مختلف مانند پوست وارد بافتهای بدن شده و الگوهای ناشناخته ای دارند آنتی ژن می گویند. آنها به داخل سلولها نفوذ کرده و آنها را از بین می برند.

[۲] به زبان ساده ، به سلولهای پروتئینی که اقدام مقاومتی و دفاعی در برابر نفوذ انجام داده و از بدن محافظت می نمایند ، لنفوسیت گفته می شود.

شکل ممکن به آنتی ژن چسبیده و اثر آنرا از بین ببرند. در نتیجه هدف سیستم ایمنی بدن در تمامی موجودات زنده یکسان بوده و استمرار حیات جاندار به هر شکل ممکن می باشد. در واکنش به آنتی ژن ، لنفوسیتهای نوع T پس از شناسایی آنتی ژن مربوطه اقدام به فعالسازی سلّولهای B می نمایند تا آنتی بادیهایی با الگوهایی با حداکثر تطبیق پذیری با الگوی آنتی ژن ها را تکثیر نموده و با آنها مقابله کنند. هر آنتی بادی تولید شده در صورتی تکثیر می شود که در طول عمر خود با الگوی حداقل یک آنتی ژن تطبیق خورده باشد. به تدریج با رشد تاثیر آنتی بادی ها در تطبیق و از بین بردن آنتی ژنها ، آنتی بادیهای تکثیر شده که هر یک تعدادی آنتی ژن را شناسایی نموده اند تبدیل به آنتی بادی حافظه می شوند تا بعداً سیستم سریعتر بتواند از این آنتی بادیهای با تجربه در حافظه ی خود استفاده نماید.

AIS با الهام گرفتن از BIS مُدلسازی و مهندسی شده است. متدهای ایمنی مصنوعی هر یک به نوعی تقلیدی از واکنشی خاص از مجموعه ی واکنش های ایمنی زیستی و دفاعی BIS می باشند.

در [۳۳] از پارادایم ها و الگوهای سیستم ایمنی مصنوعی (AIS) به عنوان مکانیسمی موفّق در ارائه ی یک روش کشف نفوذ توزیع شده در محیط ماشین مجازی (VM) استفاده شده است. مهم ترین ضعف این مقاله آنست که مجموعه ی آزمایشات آن در محیط مجازی سازی تحت یک دادگان قدیمی نفوذ ($NSL-KDD$ بروز شده در ۲۰۰۹) به انجام رسیده و مهمتر از آن فقط از ۱۹ ویژگی توصیه شده در مراجع استفاده کرده و هیچ گونه الگوریتم انتخاب ویژگی به کار نرفته است. اصولاً ویژگیهای انتخاب شده بسته به هر آزمایش متفاوت اند و در هر شرایطی جواب نمی دهند پس نیاز است تا در هر آزمایش مستقل ، زیر مجموعه ویژگیهای مطلوب مجدداً بوسیله متدهای انتخاب ویژگی انتخاب گردند.

مهمترین مزیّت این مقاله اینست که به منظور آزمون ایده پیشنهادی بر مبنای AIS از ماشین مجازی مشابه محیط واقعی به جای شبیه سازی استفاده کرده است. ذکر این نکته نیز حائز اهمیّت است که در صورت امکان، پیاده سازی بهتر از شبیه سازی است و در موارد بسیاری نتیجه واقعی تری را ارائه می دهد. AIS از سیستم ایمنی بدن انسان (HIS) الگو برداری شده است. مراحل این الگوبرداری که از فاز الهام گیری تا مدلسازی و مهندسی مراحل مختلفی دارد در (۳-۱-۲) بیان گردید. در [۱۰] محقّق ضمن بحث از شباهت های بین امنیّت سیستم های رایانه ای با AIS سیستمی را با ویژگی دفاع پیش گسترانه [۱]در تشخیص آنومالی های جدید و ناشناخته و مقابله با تهدیدات پیشنهاد نموده است. مفهوم پیش گسترانه یا به معنی توان بالقوّه سیستم در پیش بینی احتمالی خطر می باشد. تئوری خطر که در (۲-۲-۳) ارائه شده بطور بالقوّه از خصوصیّت پیش گسترانه برخوردار است. مهمترین مزیّت این مقاله ، دستیابی به ویژگی خود میزان سازی[۲]و ارائه و اثبات روابطی به منظور عدم تَجانس میان تشخیص دهنده ها بوده که به پوشش بهتر فضا خیلی کمک می کند.

[۱] Proactive Defense

[۲] Self – Tuning

این ویژگی مهم منجر به مرزبندی صحیح فضای خودی – غیر خودی گردیده و همکاری ، ارتباط موثّر و مشارکت بین عوامل خودی سیستم ، تشخیص دهنده ها ، و خود پیکربندی سیستم را منجر می شود. این مشارکت بین عوامل سیستم با مکانیسمی به نام رأی گیری (پارامتر Vote) صورت میگیرد. ما در این پایان نامه از نتایج این پژوهش در ارائه و اثبات ایده هایی برای خط نخست دفاعی نهایت استفاده را نموده ایم.

در [۱۱] نیز محقّق از مدل ترکیبی گوسیُن (GMM) برای رفع چالش تولید تشخیص دهنده های بهینه برای پوشش بهتر فضای غیر خودی استفاده نموده است. یکی از چالشهای روش انتخاب منفی ، پوشش بهتر فضای غیر خودی با کمترین میزان تشخیص دهنده ها می باشد. در حقیقت ، در فاز بررسی بلوغ یک تشخیص دهنده به جای آنکه فواصل اقلیدسیِ هر تشخیص دهنده ی نابالغ از تمامی نمونه های خودی محاسبه گردد ، فواصل آن تا نزدیکترین مولّفه های گوسیَن در فضای خودی چک می گردد. بدین ترتیب از تعداد محاسبه ها و زمان لازم برای آزمون کاسته می شود. در [۴۹] نویسنده برای مسئله خوشه بندی با استفاده از متد شبکه ایمنی مصنوعی (aiNet) و بهبود آن ، دسته بندی را پیشنهاد نموده (S – aiNet) که در مقایسه با الگوریتمِ اصلی زمان کمتری به منظور جستجو نیاز دارد. استراتژی این الگوریتم بدین صورت است که دادگان آزمون را ابتدا به بلاک هایی تقسیم می کند، سپس آن بلاک ها را به شکل توزیع شده [1]خوشه بندی می نماید که در مقایسه با خوشه بندهای سنّتی هم سریعتر است و هم عملکرد دسته بندی بهتری را در دادگان حجیم نشان می دهد.

مقاله [۱۲] تمامی الگوریتمهایی که در طی سالهای اخیر جهت بهبودِ الگوریتم انتخاب منفی ارائه شده اند را در دو دسته بندی باینری مبنا (Binary NSA) و حقیقی مبنا (Real Valued NSA) مقایسه و ارزیابی نموده است. از جمله مهمترین مزیّتهای استفاده از الگوریتم انتخاب منفی حقیقی مبنا مانند V – Detectors ، قابلیت انعطاف پذیری و مقیاس پذیری بالای آن ، پوششِ بهتر فضای غیر خودی و کمتر بودن تعداد تشخیص دهنده های بالغ در نسخه های بهبود یافته ی آن در مقایسه با رویکرد Binary NSA عنوان شده است. همچنین از جمله نقاط ضعفِ رویکرد حقیقی مبنا نیز پیچیدگی زمانی و فضایی بالا برای تولید تشخیص دهنده های موثّر حافظه ، پیدایش حفره ها [2]و بروزِ همپوشانی بین تشخیص دهنده ها عنوان شده است. همچین از جمله نقاطِ ضعفِ استفاده از Binary NSA نیز محدودیت آن جهت کاربرد در مسائل دنیای واقعی بیان شده است. مطالعه این مقاله کمک بسیاری به ما نمود تا بتوانیم مبنای اصلی پژوهش را بر پایه ی Real valued NSA قرار داده و نسخه ای از بهترین الگوریتم های این رویکرد را در روش پیشنهادی خود به کار ببریم. در زیر به مرور مهمترین کارهای پژوهشی انجام شده در حوزه تشخیص نفوذ شبکه پرداخته ایم. همه آنها در کار خود از سیستم ایمنی مصنوعی به نحوی استفاده نموده اند.

[1] Distributed Clustering
[2] Holes

۱۰-۲-۲- مروری بر کارهای پژوهشی دارای رویکرد ایمنی مصنوعی

به طور کلّی سیستم ایمنی مصنوعی دارای چهار متد میباشد که از دهه ۹۰ میلادی تاکنون توسعه داده شده اند. [۲۹] این چهار متد عبارتند از: NSA و انتخاب تکثیری [1]، DCA الهام گرفته شده از DT ، تئوری شبکه ایمنی مصنوعی [2] از دیدگاه ما و بر اساس مطالعه عمیق و ارزیابیِ کُتُب و مقالات متعدّد در حوزه ایمنی مصنوعی ، [۶۲][۲۹] [۴] [۲۱] [۳۴-۳۰] [۳-۱] [۴۸] [۵۲] [۵۴] ، به طور کلّی پتانسیل مهمّ متدهای سیستم ایمنی مصنوعی در مسئله تشخیص نفوذ را می توان در موارد زیر خلاصه نمود :

- توانایی شناسایی حملات ناشناخته و مقابله با آسیب پذیری روز صفر .
- کنترل امنیّت شبکه از طریق برقراری و تضمین ایمنی (مصونیّت) درون سیستم.
- امکان خود یادگیری لزوماً بدون نیاز به آموختن ترافیک ناهنجار شبکه.

همچنین برخی از نقاط ضعف این سیستم موارد زیر می باشند :

- حساسیّت به پارامترهای ورودی.
- کارائی سیستم تشخیص مبتنی بر ایمنی مصنوعی وابستگی زیادی به حجم ترافیکِ نرمالِ شبکه دارد، بطوریکه هر چه میزان حجم ترافیک نرمال دریافتی شبکه بیشتر باشد کارائی سیستم بالا رفته و میزان تولید و تکثیر آنتی بادیها (تشخیص دهنده ها) ی موثّر حافظه سیستم در جهت تضمین مصونیّت سیستم بالاتر می رود. زیرا این سیستم ها در رویکرد انتخاب منفی خود ، فاز یادگیری را با الگوهای خودی آموزش می بینند. پس هر چه الگوهای خودی بیشتری موجود باشند یادگیری بهتر است.

از دیدگاه تئوری ، سیستم تشخیص مبتنی بر ایمنی مصنوعی پس از اوّلین راه اندازی در سطح شبکه جهت تولید و تکثیر آنتی بادیها (تشخیص دهنده های غیر خودی) ی موثر حافظه ، به پروسه زمانی نسبتاً طولانی نیاز دارد تا سیستم شبکه را مصون نگه داشته و به آن سطح قابل قبول از مقاومت برسد. علاوه بر موارد فوق، کاربرد متدهای ایمنی مصنوعی در تشخیص نفوذ با چالشهای جدّی زیر همراه بوده اند. در سالهای اخیر پژوهشگران در مقالات خود بیشتر بر رفع این چالشها متمرکز بوده اند.

- تصادفی بودن تولید تشخیص دهنده های بالغ.
- کشف و استخراج زیر مجموعه ویژگیهای متناسب و امکان نگاشت و تخصیص آنها به سیگنالهای ورودی [3]در الگوریتم سلولهای دندریت.
- زمان محاسباتی بالای مقایسه الگوها در نسخه استانداردِ الگویتم انتخاب منفی حقیقی مبنا (RNSA).

[1] CSA (Clonnal selection Algorithm)
[2] INT (Idiotypic Network Theory)
[3] Signal Representation and Signal Categorization

- تشخیص بلادرنگ / نیمه بلادرنگِ نفوذ و یادگیری در حین تشخیص و بالعکس.

با جستجو در موتور جستجوگر گوگل اسکولار و تحقیق در ژورنالهای متعدّد علمی متوجه شدیم که روند توسعه و تحقیقات در زمینه پژوهشیِ " کاربرد ایمنی مصنوعی و متدهای آن در تشخیص نفوذ " عمدتاً به چند سال اخیر باز میگردد. در زیر مواردی از مهمترین کارهای پژوهشی در این زمینه را با هدف ارائه دیدی جامع در خصوص موضوع پژوهش جاری مورد مطالعه و مرور قرار داده ایم. یک نکته در اینجا به ذهن می رسد و آن " شیوه ارزیابی مقالاتی که از سیستم ایمنی مصنوعی بهره برده اند " می باشد. برای رفع این ابهام لازم است نکاتی را یاد آوری نماییم. در پیوست ج ، خلاصه ای از پتانسیلها و معایب مربوط به طرح های پیشنهادی ایمنی مصنوعی در مقالات مختلف سالهای اخیر به دقّت بررسی شده اند.

۱-۱۰-۲-۲- شیوه ارزیابی

به دلیل آنکه متدهای ایمنی مصنوعی _چهار متد اشاره شده در بالا با تولید و تکثیر آنتی بادیها ، مدیریت حافظه و عمل تطبیق الگو در نهایت عملاً همان کار دسته بندی را انجام می دهند ، پس شیوه ارزیابی مقالات این حوزه تفاوت چندانی با مقالاتِ امنیتی با متدهای رایج در حوزه تشخیص نفوذ ندارد. در نتیجه تنها مسئله مهمّی که در بحث ارزیابی یک "سیستم تشخیص مبتنی بر ایمنی مصنوعی" وجود دارد ، کار با پارامترهای زیستی و حدود آستانه نرخ های تولید و تکثیر آنتی بادی ، شعاع پوشش آنتی بادیها ، نرخ تطبیق الگو و مفاهیمی از این قبیل می باشند که همگی نیاز به درک عمیقِ مفاهیم زیستی از سوی پژوهشگران دارد.

در [۳۵] از هشت متد ایمنی مصنوعی پیاده سازی شده در نرم افزار وِکا ، جهت آنالیز توالی فرایندهای برنامه های در حال اجرای سیستم عامل ویندوز و دسته بندی و کشف فعالیت مشکوک استفاده شده و یک HIDS را توسعه داده است. بدین منظور از ابزار API Monitor 1.5 جهت مانیتورینگ پردازشهای در حال اجرا کمک گرفته شده است. در آزمایشات این مقاله تمام متدهای انتخاب تکثیری مصنوعی وابسته به پارامتردهی کاربر بوده و تمامی آزمایشات و ارزیابی های صورت گرفته نیز بر اساس معیارهای معمول دقّت، خطاهای کاذِب انجام و رویکرد استخراج ویژگی PE روش آنالیز $n-gram$ برای استخراج ویژگیها به کار رفته است. نتایج آزمایش ها ، دو متد انتخاب تکثیریِ ARIS2ParallelARIS2 , را به عنوان بهترین متدها از نظر عملکرد دسته بندی و کشف فرایندهای مخرّب معرفی می کند.

در مجموعه گزارشات فنّی [۶۲] مرکز تحقیقاتی سیستم های هوشمند دانشگاه سوِین بورن استرالیا پروژه ای تحقیقاتی در سال ۲۰۰۵ اجرا گردید و نهایتاً به طراحی و توسعه متدهای سیستم ایمنی مصنوعی در نسخه ۳.۶ نرم افزار وکا منجر شد ، تمام این متدها را بهمراه پارامترهای ورودی ، نرخ های حدود آستانه و تکثیر آنتی بادیها و مفاهیم پایه زیستی آن معرّفی و به خوبی تشریح نموده و آزمایشاتی را با دسته بندهای ایمنی مصنوعی پیاده سازی شده انجام داده است. این گزارشات فنّی ، کمک موثّری در جهت فهم نحوه کار با این متدها و

ساختار الگوریتم های ایمنی مصنوعی به کاربر ارائه می دهند ولی به عقیده نگارنده ، متدهای ایمنی مصنوعی شبیه سازی شده در نرم افزار ماینینگ وکا ، بیش از حد وابسته به کاربر بوده بطوریکه تعداد پارامترهای آن زیاد اند و نیاز به بهینه سازی اساسی دارند. این تقطه ضعف عمده این پروژه می باشد.

کارهای پژوهشی [۲۹] [۳۴] [۳۱] [۱-۳] [۴۸] [۵۴] به مطالعه و ارزیابی الگوریتمها و کاربردهای سیستم ایمنی مصنوعی در تشخیص نفوذ پرداخته اند. به عنوان نمونه [۲۹] مدلهای مختلفِ پیاده سازی شده از متد انتخاب منفی را که به طور گسترده در تشخیص خطا و امنیت سیستم های شبکه مورد استفاده قرار میگیرند آنالیز نموده است. همچنین تحقیق مذکور یک بررسی موردی برای شناسایی پتانسیلهای الگوریتم انتخاب تکثیریِ ClonalG و الگوریتم شبکه ایمنی مصنوعی aiNet در دادگان آزمون سرطان انجام داده و توانایی های هر یک را به خوبی محک زده است.

این دو الگوریتم در نرم افزار وکا [1]نسخه ۳.۶ پیاده سازی شده اند. در [۳۴] ضمن بررسی و دسته بندی مطالعات انجام شده در زمینه سیستم های تشخیص نفوذ مبتنی بر ایمنی مصنوعی، چارچوبی را برای طراحی یک سیستم تشخیص نفوذ مبتنی بر ایمنی مصنوعی ارائه داده است. در این تحقیق، تفاوت میان دو رویکرد انتخاب منفی باینری و حقیقی مبنا از دیدِ تکاملی به دقّت بررسی شده و نویسنده با تمرکز بر مِتد Real Valued NSA نقش الگوریتم ژنتیک را در بهینه سازی و تولید پویای تشخیص دهنده های بالغ در کاهش خطا و افزایش نرخ تشخیص دسته بندی به خوبی ارزیابی نموده است.

مقاله [۳] با عنوان "به سمت ایجاد سِرورِ ایمنی مصنوعی جهت مقابله با حملات سایبری " چارچوبی را ارائه داده که دو واکنش ایمنی ذاتی و اِکتسابی در کنار هم "سِرور ایمنی مصنوعی" را تشکیل می دهند. این سرور اختصاصی می تواند جهت امنیت وب سرور مورد استفاده قرار گیرد.

نحوه کار این سرور به این صورت است که حملات سایبری در اوّلین گام دفاعی و بر مبنای میزان آسیب پذیر بودن توسط تابع ایمنی ذاتی ، تشخیص داده شده و قرنطینه می شوند. سپس این تابع فرایند جدیدی را در سرور ایجاد می کند که بر مبنای این فرایند، زیر سیستم یادگیری مبتنی بر ایمنی اکتسابی، سورس کد برنامه ها و پردازشهای موجود در قرنطینه را مورد ارزیابی و تحلیل کد قرار میدهد. امّا چالشی که وجود دارد آسیب پذیر بودن این وب سرورِ مجهّز به ایمنی مصنوعی در اوّلین فاز شناسایی حملات می باشد که در مقاله نشان داده شده که به مرور زمان این چالش رفع گردیده و نرخ شناسایی سِرور ایمنی نیز بهبود می یابد. مبنا و علّتِ زیستی بروز این چالش را می توان از شکل ۸ و توضیحات موجود در بخش ۲-۱-۷-۱ (انطباق) به خوبی درک نمود.

[1] http://www.cs.waikato.ac.nz/~ml/weka/

فصل سوّم

Provide the proposed Method

(روش شناسی تحقیق و تجزیه و تحلیل یافته ها)

Research Methodology

And

Analysis of Findings

ارائه روش پیشنهادی
بخش نخست
تئوری خطر

۳-۱-۱- مقدمه

این فصل به ارائه روش تشخیص نفوذ پیشنهادی اختصاص داشته و از دو بخش تشکیل شده است. ارئه ی ایده استفاده از دو خط دفاعی بر مبنای مصونیّت ایمنی ذاتی – تطبیق پذیر و با دو استراتژیِ دسته بندی و تشخیصِ کاملاً جداگانه موضوع این فصل میباشد. به منظورِ ارزیابیِ روش پیشنهادی قبل از همه نیازمند استخراجِ دادگانِ تست و یادگیری به صورت تصادفیِ یکنواخت از دادگان مربوطه هستیم. از اینرو در [پ-ب-۱] بخشی تحت عنوان آماده سازی دادگانِ یادگیری و تست آمده است.

امّا به عنوان بخش نخستِ فصل ، تئوری خطر و مکانیسمِ سلوّل دِندریت بیان شده و چگونگی کاربرد این مکانیسم به عنوانِ یک سِپَر دفاعی و در حقیقت به عنوان لایه ی نخستِ استراتژی دفاع در عمق در روش پیشنهادی بررسی شده است. همانگونه که در بخش مبانی نظری در فصل دوّم نیز اشاره گردید ، انتظار می رود تا روش پیشنهادی به طریقی بتواند استراتژی دفاع در عمق را در گامهای مختلف شناساییِ نفوذ خود پیاده نماید. در بخش دوّم ، خط دوّم دفاعی با استفاده از مکانیسم انتخابِ منفی حقیقی مبنا بهبود یافته با دو فاز یادگیریِ اوّلیه و ثانویّه شبیه سازی شده است. سپس امکانِ شبیه سازیِ یک حیات مصنوعی و طی روندِ چند سیکل اوّلیه از این حیات بررسی شده و در نهایت ایده ترکیبِ نتایجِ ارزیابی خطوط در سیکلهای متوالی این حیات با بهره برداری از مفاهیمِ مدیریت حافظه و مکانیسم رای گیری و مشارکتِ تشخیص دهنده های توزیع شده به منظور تخصیص برچسب احتمالی نهایی نمونه های تست ارزیابی گردید.

خط نخست دفاعی - تئوری خطر

الگوریتم سلّولهای دندریت مدلی انتزاعی از رفتار سلّولهای دندریت بر مبنای تئوری خطر می باشد. [۴۶][۳۶]

۲-۱-۳- الگوریتم سلّولهای دِندریت

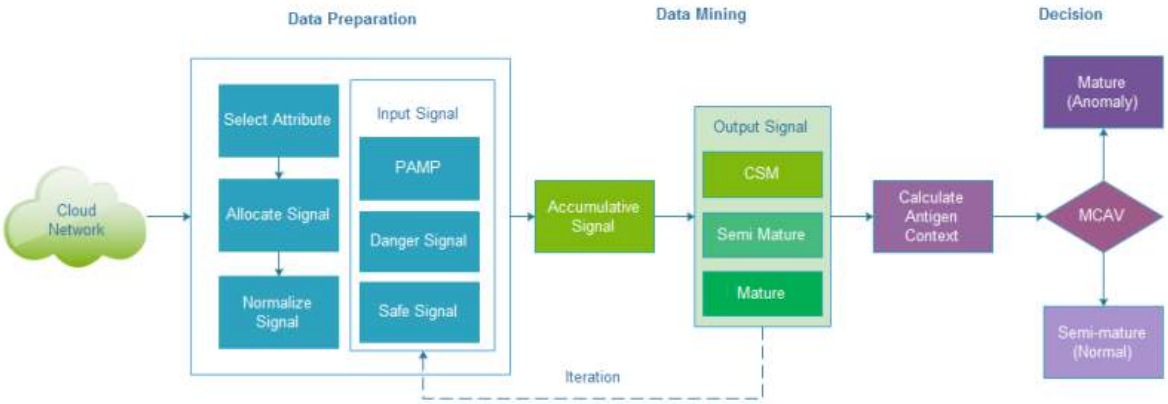
مکانیزمِ تئوری خطر بوسیله الگوریتم سلّولهای دندریت بر روی ترافیکِ شبکه اعمال می گردد. مطابق آنچه که در فصل دوّم این پژوهش بیان گردید ، از دیدگاه بیولوژیک [۱۵] وقتی آنتی ژنها وارد بافت های سلّولی بدن می شوند اغلب تخریب سلّولهای بافتهای بدن را در پی دارد ، در نتیجه این تخریب سلّولی ، سیگنالهایی به سلّولهای دَندریتِ مراقب در بافت آسیب دیده ارسال می شوند. سلّولهای دندریت با محصور کردن سلّولهای آسیب دیده شروع به نمونه برداری از آنتی ژنها نموده و کلیه سیگنالهای دریافتی از آنها را پردازش می نمایند. با پردازش سیگنالها نهایتاً هر سلول دندریت به حالت بالغ یا نیمه بالغ در آمده و مهاجرت می کند. اصطلاح مهاجرت سلّول ها به معنای اطلاع رسانی در مورد وضعیت بافت های آسیب دیده و نفوذِ رخ داده به لایه دفاعی پایین تر (مص.تط. با واکنشِ لنفوسیت های T و B) می باشد.

مکانیزم ایمنیِ تطبیق پذیر بوسیله یک سری سلّولهای پروتئینی به نام لنفوسیت های T و B انجام می شود. این پروتئین ها ، همواره حاوی مجموعه ای از آنتی بادیهای موثرِ حافظه هستند که قادر به تشخیص هر نوع الگوی آنتی ژنیک ناشناخته می باشند. لنفوسیتها ، نمونه ها و سیگنالهای محاسبه شده توسط سلولهای دندریتِ مهاجرت یافته بالغ یا نیمه بالغ را دریافت نموده و فعّال می شوند. به عبارت بهتر ، سلّولهای دندریت پس ازپردازش سیگنال و مهاجرت خود ، نمونه مربوطه را بهمراه اطلاعاتی راجع به میزان خطر ناک بودن آن نمونه (سیگنالهای خروجی پردازش شده) به زیر سیستم واکنش اکتسابی (لنفوسیت تی) می دهند و آنرا فعال می کنند. به محض فعّال شدن لنفوسیت ها ، آنها اقدام به تولید و تکثیر و بروز رسانی استخرِ آنتی بادیهای خود می نمایند تا حافظه خود را پالایش و تعدیل نمایند. اگر نمونه های آنتی ژنیکِ شناسایی شده بوسیله سلولهای دندریت ، توسط هر آنتی بادی بالغِ حافظه بایند گردد بدان معناست که احتمالاً حمله ای جدید رخ داده است. در صورتی که الگوی آنتی ژن توسط آنتی ژنهای خودی تشخیص داده شود به معنی آنست که آنتی ژن دریافت شده از نوع خودی بوده و با توجه به سیگنالهای سلول دندریتی که آن آنتی ژن را شناسایی نمونه و منجر به مهاجرت اش شده تصمیم میگیرد که آن آنتی ژن الگوی بی خطر دارد یا خیر.

در فصل دوّم با مرور ادبیّات نظری مفاهیم بیولوژیک به حد کافی بحث شد. بدن جانور بوسیله سلولهای دندریت و سازوکار مهاجرت این سلولها و بلوغ ، به نفوذ عوامل پاتوژنیک و آنتی ژنها واکنش نشان می دهد. این واکنش از نوع ذاتی بوده و نوعی دسته بندی بدون / نیمه نظارت شده است. علّت آن در بخش های بعد بیان شده است.

طی سالهای اخیر ، مقالاتی مانند [۵۵] [۳۶ و ۳۷] با هدف بهبود عملکرد مکانسیمِ سلّولِ دندریتِ استاندارد به ارائه ایده هایی پرداخته اند. برای نمونه در [۳۷] الگوریتمی بر تشخیص نفوذ مبتنی بر آبر ، ارائه شده که

مکانیسمِ سلّولهای دندریت را پیاده سازی نموده است. شکل زیر مدلی از این مکانیسم را در ابر نشان می دهد. ضمن اینکه در این مقاله رابطه ای برای محاسبه ی MCAV نیز پیشنهاد شده است. ما از این رابطه برای محاسبات خود در خط نخستِ دفاعی استفاده نمودیم. در [۴۶] که یک رساله دکتری می باشد ، پژوهشگر به بررسی ویژگیهای DCA از دو جنبه تئوری و تجربی پرداخته است. از جمله موارد مهمی که در این پایان نامه بر روی آنها کار شده ، فاز پیش پردازش و آنالیز ویژگیهای دادگان نفوذ یادگیری NSL – KDD برای تخصیص سیگنالهای ورودی میباشد. نتایج این تحقیق نشان می دهد که بسته به اینکه چه تکنیکی برای انتخاب ویژگی انتخاب گردد ممکن است زیر مجموعه ویژگیهای متفاوتی برای نِگاشت[1] به سیگنالهای ورودی کاندید شوند.



شکل ۱ – الگوریتمی برای مدلِ تشخیص نفوذِ ابری مبتنی بر مکانیسمِ سلولهای دندریت، برگرفته از [۳۷]

در پژوهش مذکور ، دو متد مبتنی بر همبستگی و بهره اطلاعات[2] به کار رفته که در نگاشت زیر مجموعه ویژگیهای انتخاب شده به سیگنالها دو نتیجه کاملاً متفاوت داشته است. بطوریکه در متد IG به ترتیب سه ویژگی dst_{byte}. src_{byte} count به دلیل دارا بودن بیشترین وابستگی[3] با ویژگی کلاس داده یادگیری، به ترتیب به عنوان مناسبترین ویژگیها به منظور نگاشت به سیگنالهای ورودی PAMP. Danger و Safe انتخاب شدند. در حالیکه در مورد متد دیگر ، سه ویژگی dct_host_{count} . count. $srv_{diff_{host_{rate}}}$ بدین منظور در نظر گرفته شدند. مطابق مواردی که قبلاً بیان گردید ، در DCA ، دو نوع سیگنال وجود دارد. ما در این قسمت این سه سیگنال ورودی را از جنبه بیولوژیکی آن تشریح می کنیم :

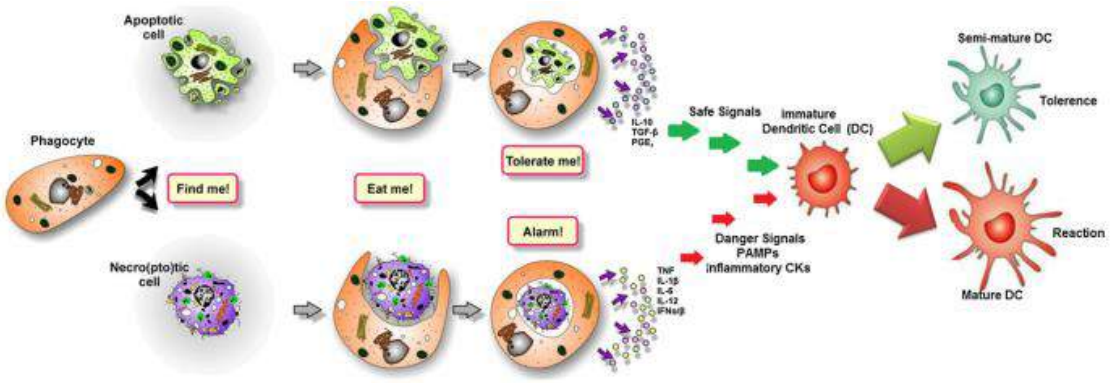
- سیگنالهای ورودی :

[1] Mapping
[2] IG and Correlation based method
[3] Dependency

این سیگنالها سه نوع اند ، امن ، خطر و سیگنالِ الگوهای مولکولی پاتوژنیک [1]. این سیگنالها از بافت های سلّولی آسیب دیده که آنتی ژن در آنها حضور دارد ، منتشر شده و به سمت نزدیکترین سلّولهای دندریت در آن بافت ارسال میشوند. سلول دندریت ضمن پردازش سیگنالهای منعکس شده از هر سه نوع، از الگوی آنتی ژنهای نفوذ یافته در بافت ها نیز اطلاع یافته و از آنها نمونه برداری می کند. مطابق تعاریف ارائه شده در مقاله [۱۵] ، PAMP مولکولهایی هستند که به گروهی از پاتوژنها وابستگی دارند.

این مولکولها گونه ای از میکروبها هستند که در صورتی که حضورشان برای سلّول دندریت با دریافت سیگنال اثبات شود ، این سلّول را به حالت فعّال/ بالغ می برد. سیگنالهای از نوع خطر نیز سیگنالهایی هستند که وقتی که سلّول خودیّ بدن در اثر نفوذ آنتی ژنهای غیر خودی ، میمیرد ، سیگنالهایی را منتشر می کند که از نوع خطر می باشند. در صورتی که سلّول دندریت این نوع سیگنالها را دریافت کند نیز بسته به پردازش آن ممکن است به وضعیت بالغ برود. امّا سطح اطمینان کمتری نسبت به سیگنالهای PAMP برای سلول دندریت از نظر بالغ شدن بوجود می آورند. محقّقان به منظور کشف رابطه ای که منجر به پردازش سیگنالهای ورودی و نهایتاً مهاجرت سلّول به یکی از دو زیر مجموعه بالغ یا نیمه بالغ می گردد را به شکل یک ماتریس تحت عنوان ماتریس وزنی پیشنهاد نموده اند. این ماتریس بر روی مقادیر سیگنالهای ورودی به سلول دندریت اعمال می شود و نهایتاً سه سیگنال خروجی بدست می آیند. سه سیگنال خروجی بدست آمده ، وضعیت و میزان بلوغ سلّول دندریت را تعیین خواهند نمود. در بخش های بعد ما عملکردِ این ماتریس را توضیح داده ایم.

سیگنال امن نیز سیگنالی است که در نتیجه مرگ طبیعی سلولهای خودی بافت های بدن [2]توسط سلّولهای دندریت دریافت می شوند. شکل زیر این دو نوع مرگ ، طبیعی و بر اثر تخریب سلولی را بهتر نشان می دهد.



شکل ۲ - مرگ طبیعی در برابر مرگ ایجاد شده بر اثرِ تخریبِ بافت سلّول خودیِ آسیب دیده و تاثیر آن بر بلوغِ سلّول دندریت

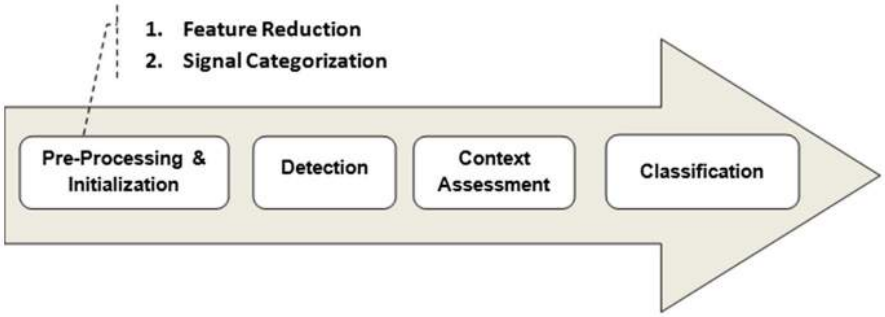
[1] PAMP
[2] Apoptosis

دریافت این نوع سیگنال (امن)، معرّفِ نُرمال بودن وضعیت سیستم بدن می باشد. اگر سیگنال دریافت شده در سلّول دندریت ، از نوع امن باشد ، سلّول دندریت مهاجرت ننموده و زیر سیستم دفاعی با واکنش تطبیق پذیر (لنفوسیتهای بی و تی) را نیز فعّال نمی کند. در نتیجه به نمونه برداری آنتی ژنها در بافت ها همچنان ادامه می دهد. در مقاله [۲۱] که به بررسی مقایسه ای عملکرد دو الگوریتم NS و DC پرداخته ، نویسنده سه سیگنال ورودی را از دو جنبه زیستی و واقعی در دنیای شبکه ، به خوبی تحلیل کرده و توابعی را بدین منظور ارائه داده است. (جدول ۱) این توابع بیانگر این نکته هستند که اگر از بین زیر مجموعه ویژگی ها زیر مجموعه A برای نگاشت به سیگنال PAMP مناسب تشخیص داده شود ، باید نشان دهنده "پیام های خطا در هر ثانیه" باشد.

جدول ۱ - توابع سیگنال در مکانیسمِ سلولهای دندریت

Signal	*Biological property*	*Computational example*
PAMP	Indicator of the microbes presence	Error message per second
DS	Indicator of tissue damage	Network packet per second
SS	Indicator of healthy tissue	Size of network packets

شکل زیر مکانیسمِ سلّولهای دندریت را نشان میدهد. مطابق این شکل در فاز اوّل پیش پردازش ، دو تابع "کاهش ابعاد" و "پردازش سیگنال" انجام می شود.



شکل ۳ - مکانیسمِ سلولهای دندریت، برگرفته از [۱۵]

بسته به اینکه کدام نوع سیگنال ورودی دریافت شود و میزان هر یک از این سه سیگنال چه قدر باشد وضعیت سلّول دندریت به یکی از سه حالت زیر (سیگنالهای خروجی) پردازش می شود. [۲۱][۱۵]

- سلولهای دندریت نابالغ [۲] (نارس) :

سلّولهای دندریت ابتدا در این وضعیت می باشند. به محض وقوع نفوذ یا ورود آنتی ژنهای بیگانه به بافت سلّولی بدن، سلّولهای خودی بدن تخریب شده و سیگنالهای ورودی از بافت های آسیب دیده منتشر می شوند. سلّولهای دندریت این سیگنالها را دریافت نموده و بر روی آنها پردازش سیگنال انجام می دهند تا وضعیت نهایی سلّول دندریت مشخص گردد.

به عبارت دیگر با پردازش سیگنالهای ورودی بسته به اینکه چه نوعی (امن ، خطر یا PAMP) باشند ، وضعیت سلّول دندریت ممکن است به یکی از سه وضعیت نارس، بالغ و نیمه بالغ تبدیل شود. اگر سلول دندریت به دو وضعیت بالغ / نیمه بالغ برود مهاجرت سلولی را در پی خواهد داشت.

- وضعیت بالغِ سلول دندریت[۳]:

اگر سلول دندریت سیگنالهای زیادی از نوع PAMP یا خطر از سمت بافت های آسیب دیده دریافت کند، بالغ شده و رنگ آن نیز تغییر می کند. این سبب می شود که سلول دندریت از بافت سلولی به غدّه لنفاوی[۴]مهاجرت کند و الگوهای آنتی ژنهای نمونه برداری شده را نیز با خود ببرد. پس از مهاجرت ، سلّول دندریت بالغ ، آنتی ژنهای نمونه برداری شده [۵]را در اختیار لنفوسیتهای T و B که دارای واکنش ایمنی تطبیق پذیر هستند قرار می دهد تا لنفوسیت B حافظه بوسیله آنتی بادی های بالغ و موثر خود ، الگوی آنتی ژن های نمونه برداری شده را شناسایی نماید.

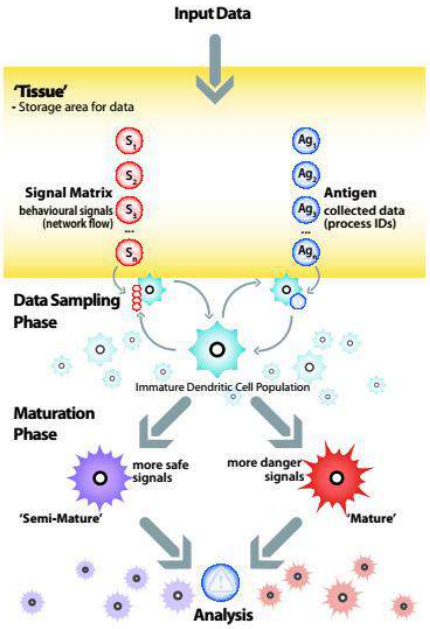
- وضعیت نیمه بالغِ [۶]سلول دندریت :

اگر سلول دندریت سیگنالهای از نوع امن را دریافت کند ، به وضعیت نیمه بالغ در آمده و لنفوسیت های از نوع T را نیز فعّال می کند تا سیستم را در حالت آماده باش نگه دارند. در این وضعیت لنفوسیتهای B فعّال نمی شوند. شکل زیر شماتیک زیستی DCA را بهتر نشان می دهد. سیگنالهای سمت چپ طی فرایند مرگ سلولی منتشر می شوند که در شکل قبلی توصیف گردید. علاوه بر این ، مقدار سیگنال تحریک[۷]یا CSM نیز با

---

[۱] Cumulative
[۲] iDCs
[۳] Mature DCs
[4] Lymph node
[5] Present
[6] Semi mature
[7] Co – stimulatory moclule

پردازش سیگنال [1]بدست می آید. سلول دندریت ، سیگنال CSM را به منظور تسهیل فرایند نمونه برداری آنتی ژنها محاسبه می کند. مقدار این سیگنال معیار بررسی مهاجرت سلول دندریت و توقف نمونه برداری آنتی ژن ها و سیگنالهاست.



شکل ۴- شماتیک DCA : نمونه برداری پیوسته داده (آنتی ژن) ورودی و سیگنالهای ورودی ، فرایند مهاجرت سلول دندریت و آنالیز

به منظور بررسی کامل مکانیسمِ الگوریتمِ سلّولهای دندریت و آشنایی با تمامی مراحل تولید سیگنال های ورودی و پردازش آنها، تشخیص تا دسته بندی ، ما این مراحل را از جنبه الگوریتمیک بررسی و در ادامه ، استراتژی پیشنهادی را برای خط نخستِ دفاعی ارائه نمودیم.

۳-۱-۳- آزمایشِ شبیه سازی و بهبود مکانیسمِ سلّول دِندریت

نسخه هایی مختلفی از الگوریتم سلّولهای دندریت در مقالات مختلف توسعه داده شده اند [۵۴][۲۱] [۴۶] [۳۶] که هر یک بسته به شرایط مسئله می تواند مزیّت ها و نقاط ضعفی بهمراه داشته باشد. برای نمونه در

[1] Signal categorization

[۵۴] نسخه قطعی[1]این الگوریتم به نام dDCA پیشنهاد شده است. این الگوریتم برای شبکه هایی مناسب است که حملات در آن با تاخیر بالا یا به ندرت رخ می دهند و برای شبکه هایی که نرخ وقوع حملات بالایی دارند نتایج قطعی ندارد.

در مقاله [۳۶] نویسنده از مفهوم ایمنی خودکار[2]در بدن ، الهام گرفته است تا عملکرد الگوریتم سلول های دندریتِ استاندارد را در دسته بندی باینری بهبود ببخشد. میزان این بهبود چشمگیر به نظر میرسد. از جمله استراتژیهای پیشنهادی در بهبود عملکرد میتوان به متدِ تزریقِ واکسن اشاره کرد که تاثیر بسزائی در کاهش خطای دسته بندی داشته است. واکسن زدن در سیستم های ایمنی مصنوعی به معنای بروز رسانی و تعدیل هر دو پایگاه امضای حملات با تزریق آنتی بادیها [3](پادزهر) و یا تزریق داده نرمال جدیدً به پایگاه پروفایلهای جاری سیستم می باشد که مقاومت سیستم (بدن) را برای تشخیص و از بین بردن عوامل آنتی ژنیک خارجی بالا می برد.

ما به منظور پیاده سازی خط نخست دفاعی بر مبنای تئوری خطر ، الگوریتم سلّولهای دندریت استاندارد پیشنهاد شده در مقاله [۱۵] را شبیه سازی نموده و ضمن مقایسه با تعدادی الگوریتم های خوشه بندِ فراابتکاری مبتنی برجمعیت نمونه اولیّه که در بخش قبل بررسی گردید (BHA. PSO. GA. K − means) ، این الگوریتم را به طور چشمگیری بهبود دادیم. DCA استاندارد در مقاله مذکور علیرغم اینکه نرخ های دسته بندی آن چندان مطلوب نبوده و خطاهای آن نیز چندان قابل چشم پوشی نیست ولی پیچیدگی سایر نسخه های این الگوریتم را نداشته و به نظر می رسد برای ارائه ایده های بعدی مناسب باشد.

از طرفی این الگوریتم در مسئله ما به صورت نظارت نشده/ بدون یادگیری قبلی عمل نموده و مسله را به صورت کاملاً توزیع شده حل می کند برخلاف خوشه بندهای فراابتکاری مذکور که استراتژیِ پردازش متمرکزگِرا و تصادفی می باشد ، بنابراین از این نظر گزینه مناسبی برای کاربرد آن در خط نخست دفاعی می باشد. مقاله مذکور ، به طور اجمالی به بررسی این الگوریتم ، از دو بعد تئوری (بیولوژیکی) و الگوریتمیک پرداخته و مکانیسمِ آنرا را به دقت بررسی نموده است.

بنا به استناد به مقاله مذکور ، ما در این قسمت خلاصه ای از مهمترین جزئیات این الگوریتم را بررسی نموده و سپس استراتژی پیشنهادی برای خط نخست دفاعی روش پیشنهادی را ارائه نمودیم.

۱-۳-۱-۳- خلاصه ای از جزئیات الگوریتمیکِ تئوری خطر

[1] Deterministic
[2] Auto immunity
[3] non self detectors
[4] Normal self detectors

تاریخچه پیدایش تئوری خطر و ارائه DCA در فصل دوّم بحث گردید. این الگوریتم با دو ورودی کار می کند: ورودی اوّل آنتی ژن ها (منظور نمونه های موجود در دادگان نفوذ آزمون) و ورودی دوّم که سیگنالها هستند. کار الگوریتم ، با ایجاد و توزیعِ سلّولهای دَندریت آغاز می گردد. نقش سلّولهای دندریت شناسایی و نمونه برداری از آنتی ژنها و محاسبه سیگنالهای خروجی آنهاست. در الگوریتم استاندارد ، به ازاء هر آنتی ژن در دادگان آزمون ، یک سلّول دندریت ایجاد می شود که به صورت تصادفی وظیفه نمونه برداری از الگوهای آنتی ژنیک موجود در دادگان آزمون [1] و همچنین محاسبه سیگنالها را بر عهده دارد. بدین صورت که هر سلّول دندریت در هر موقعیتی در فضای ابعاد مسئله ، بر حسب فاصله اقلیدسی از نمونه های آنتی ژنهای اطراف آن ، آنتی ژنها را نمونه برداری می کند. در مکانیسم سلّولِ دندریتِ استاندارد در [۱۵] هر سلول دندریت ممکن است آنتی ژنهای مختلف با الگوهای متفاوتی را نمونه بردای کند. تابع Get_Antigen() این کار را انجام می دهد. نحوه نمونه برداری یا Present کردن آنتی ژنها توسط سلولهای دندریت به صورت زیر در محیط متلب قابل شبیه سازی می باشد.

هر آنتی ژن در دادگان آزمون در فضای ابعاد مسئله به شکل تصادفی یکنواخت تکثیر [2] می شود. به عنوان نمونه اگر مجموعه دادگانِ آنتی ژن بصورت $\mathrm{Ag} = \{\mathrm{Ag_1}.\mathrm{Ag_2} \ldots\ldots \mathrm{Ag_n}\}$ باشند ، تکثیر تصادفی آنتی ژنها ، بُرداری را به صورت زیر ایجاد می کند. وظیفه سلّولهای دَندریت اینست که بسته به موقعیت خود در فضای ابعاد مسئله، از این آنتی ژنها نمونه برداری کنند. نکته مهمی که وجود دارد ، اگر n نوع آنتی ژن داشته باشیم حاصل تکثیر، بُرداری را به صورت زیر بدست می دهد که اندازه این بُردار چندین برابر تعداد آنتی ژنهای دادگان آزمون می باشد.

$$A_{Vector} = \{Ag_1.Ag_1.Ag_1.Ag_1.Ag_2.Ag_2.Ag_2.Ag_2.Ag_3.Ag_3.Ag_3.Ag_3.Ag_3.Ag_3.\ldots.Ag_n.Ag_n\}$$

یک مثال از نمونه برداری توسط یک سلّول دندریتِ فرضی $\mathrm{DC_i}$ می تواند به شکل زیر باشد :

$$DC_i = \{Ag_1.Ag_1.Ag_1.Ag_2.Ag_2.Ag_2.Ag_2.Ag_3\}$$

در این مثال ، سلّول دندریت مورد نظر به شکل تصادفی سه آنتی ژن از نوع اوّل ، چهار آنتی ژن از نوع دوّم و یک آنتی ژن از نوع سوّم را نمونه برداری می کند. این سه آنتی ژن نسبت به بقیه آنتی ژنها در فضای ابعاد مسئله به موقعیت سلّول دندریت مورد نظر نزدیکترند. علاوه بر آن سیگنالهای هر سه نوع آنتی ژن نیز دریافت و توسط سلّول دندریت پردازش می شوند.

---

[1] هر نمونه داده آزمون دارای n ویژگی

[2] Clone

در اینجا لازم است توضیح دهیم که برای هر نمونه آنتی ژنیک Ag_n با الگوی مشخص (نمونه های ترافیک شبکه آزمون)، مقادیر سه سیگنال ورودی امن ، خطر و PAMP بدست می آیند. ترکیب وزنیِ مقادیر سیگنالهای ورودی با اعمال رابطه وزنی ، مشخص می کنند که هر نمونه به چه میزان می تواند سلّول دندریت را بالغ/ نیمه بالغ سازد تا نهایتاً منجر به مهاجرت آن گردد. نحوه محاسبه سیگنالهای خروجی بالغ ، نیمه بالغ و تحریک (CSM) برای سلّول دندریت در بخش بعدی بیان شده است. شکل زیر DCA استاندارد را نشان میدهد. [۲۱] در بخشهای آینده خواهیم دید که چالش اصلی DCA استاندارد ، بدست آوردن و تخصیص سیگنالهای ورودی[1] می باشد.

Algorithm1. Pseudo code of Original DCA

Inputs : S = Input $Pre_{categorized}$Signals. Migration Threshold. Antigens Set.
Output: E = MCAV Values (Mature Context Antigen Value)
−Create an initial population of Dendridic Cells (DCs). D /∗ **Preprocessing/**
−Randomely select n DCs from DC population
For each selected DC **DO** /∗ **Detection_phase/**
−Get the antigen
−Store the antigen
−Get the signals /∗ **Signal_Assessment**
−Calculate interim output signals
−Update the cumulative output signals
If cumulative Csm > migration threshold **Then**
−Remove the DC population
−Assign the Cell − context to DC
If cumulative Semi ≤ cumulative Mat **Then** /∗ **Context_Assessment_phase/**
Cell context = 1
Else
Cell context = 0
End
− All DCs which collected the antigen and have a cell_context out for analysis
−Termination this DC and add a naive DC to the population
Else
− DC back to population
End
For each incoming data **Do** /∗ **Classification_phase/**
− Calculate the number of mature DC and semi − mature DC
If nb semi − mature DC > nb mature DC **Then**
Antigen = normal
MCAV = 0
Else
Antigen = abnormal
MCAV = 1
End
End

شکل ۵ – شبه کد مربوط به الگوریتمِ سلولهای دندریت استاندارد ، بر گرفته از [۲۱]

[1] Input Pre_categorized Signals

۴-۱-۳- تشریح سازوکار الگوریتم سلّولهای دندریت به بیان ساده

DCA دو ورودی می گیرد: سیگنال ها و آنتی ژنها. نمونه های ترافیک شبکه معادل با آنتی ژنها هستند. الگوی پپتیدهای (پُرزها) موجود بر روی سطح مولکولی آنتی ژنها متفاوت از یکدیگر می باشد. همین تفاوت ، الگوهای خاصی را ایجاد می کند که آنتی ژنهای غیر خودی را قادر می سازد تا به سلّولهای خودی در بافتهای خاصی نفوذ نموده و آنها را تخریب نمایند. تفاوت الگوی آنتی ژنها را می توان معادل با مقادیر مختلف ویژگیهای نمونه های مختلف ترافیک شبکه در نظر گرفت.

فرایند الگوریتم با فاز پیش پردازش آغاز می شود و جمعیتِ اولیّه سلّولهای دندریت ایجاد می شوند. در مقاله اصلی ، تعداد و موقعیّتِ سّلولهای دندریتِ تولید شده را به ترتیب برابر با تعداد و موقعیّتِ آنتی ژنها در فضای ابعاد مسئله در نظر گرفته است. بدین معنا که اگر n آنتی ژن داشته باشیم ، دقیقاً n سلّول دندریت ایجاد می شوند و موقعیت آنها دقیقاً متناظر با موقعیتهای آنتی ژنها می باشد.

در مرحله بعد ، هر سلّول دندریت اقدام به نمونه برداری با تابعِ Get_Antigen() می نماید، این نمونه برداری به شکل تصادفی می باشد و در الگوریتم استاندارد ، به فاصله یا معیارهای دیگر وابسته نیست. در بخش های بعد خواهیم دید که ما برای نمونه برداری و کشف و تخصیص سیگنالهای ورودی استراتژی خاصی پیشنهاد نموده ایم. همزمان با نمونه برداری تصادفی از آنتی ژنها ، سلّولهای دندریت سیگنالهای دریافتی از آنتی ژنهای نمونه برداری شده را نیز محاسبه می نمایند. بَرایندِ محاسبه سیگنالهای ورودی ، تولید سیگنالهای خروجی بالغ / نیمه بالغ و CSM می باشد. سیگنال CSM معرّف آنست که هر سلّول دندریت تا چه حد توانسته بالغ / نیمه بالغ گردد تا مهاجرت صورت گیرد.

همانطور که بیان شد هر سلّول دندریت هنگام نمونه برداری تصادفی از آنتی ژنها ، سیگنالها را نیز دریافت و برایند آنها را نیز محاسبه می نماید. ممکن است در طول اجرای نمونه برداری ، هر آنتی ژن چندین بار توسط چندین سلّول دندریت مورد نمونه برداری قرار بگیرد که بستگی به معیار خاصی ندارد زیرا نمونه برداری به صورت تصادفی صورت میگیرد. همچنین در فاز پردازشِ سیگنالهای خروجیِ بالغ ، نیمه بالغ و CSM ، حد آستانه ای را برای مهاجرت سلولهای دندریت در نظر گرفته می شود که بسته به تجربه کاربر تعیین می گردد.

۱-۴-۱-۳- منظور از مهاجرت چیست ؟

هر سلّول دندریت پس نمونه برداری به دو وضعیت ممکن است تغییر پیدا کند : بالغ یا نیمه بالغ گردد. وضعیت بالغ در صورتی رخ می دهد که آنتی ژنهای نمونه برداری شده دارای الگوی غیر خودی بوده باشند. بدین معنا که از دیدگاه بیولوژیک اگر بیشتر سیگنالهای دریافت شده از سلّولها ، سیگنالهایی از نوع خطر یا تحریک باشند نشان دهنده وضعیتی است که بیشتر سلّولها آسیب دیده اند و یا احتمال خطر وجود دارد. بنابراین این سیگنالها

منجر به بالغ شدنِ سریعِ سلّولِ دندریت شده و سلّول مهاجرت می نماید. سل. های دن.بغ. مهاجرت یافته ، آنتی ژنهای نمونه برداری شده خود را در اختیار لایه دفاعی پایین تر که واکنش تطبیق پذیر دارد قرار می دهند.

B cell ها و T cell ها لنفوسیت های (سلولهای پروتئینی بدن) موجود در لایه دفاعی زیرین با واکنش اکتسابی هستند که قابلیت های بالایی در یادگیری و شناسایی و کشف نفوذ دارند و معمولاً پس از آنکه سلّولهای دندریت بالغ می شوند ، نمونه های خود را به این لنفوسیتهای B می دهند تا فعال شوند. سپس این سلولها (B) ، اقدام به یادگیری از آنتی ژنهای نمونه برداری و دریافت شده نموده و آنتی بادیهای متناسب با الگوی این آنتی ژنها را تولید و تکثیر می کنند تا حافظه خود را نیز بروز و تعدیل کرده باشند. لنفوسیتهای B ، دو واکنش هیمورال و واکنش ایمنی سلّولی را به منظور از بین بردن کامل و مصون سازی سیستم در برابر نفوذهای بعدی به بدن انجام دهند. در واقع لنفوسیتها با مص.تط. خود نقش موثر و مهم تری در مصون سازی سیستم در برابر نفوذ ایفا می کنند. این موارد در فصل دوّم به طور مفصل بحث گردید.

مهاجرت سلّولهای دندریت چه بالغ چه نیمه بالغ ، تنها در صورتی ممکن است که مقدار سیگنال خروجی CSM هر سلّول دندریت بیشتر از حد آستانه مهاجرت سلّولی تعیین شده ، mt باشد. اگر از این مقدار کمتر بود ، بدین معناست که مقدار سیگنال خروجی CSM محاسبه شده ی سلّول به حد لازم برای مهاجرت نرسیده و باید نمونه برداری بیشتری انجام دهد و سیگنالهای آنتی ژنهای نمونه برداری شده ی زیادی را پردازش نماید. اگر این حد آستانه عبور کرد ، ضمن مهاجرت سلّولی ، فاز context_assessment برای سلّول دندریت اجرا می گردد. در این فاز ، به هر سلّول قبل از مهاجرت یک برچسب تخصیص داده می شود به نام context. بدین صورت که اگر تعداد آنتی ژنهای نمونه برداری شده ای که منجر به بلوغِ سلّول دندریت می شوند (بالغ کننده) بیشتر از تعداد آنتی ژنهای نمونه برداری شده ی نیمه بالغ کننده باشد ، برچسبِ سلّول دَندریت بصورت بالغ تعیین می شود و به دسته سل. های دن.بغ. مهاجرت می کند. بدین معنا که این سلّول ، تعداد بیشتری از آنتی ژنهایی را نمونه برداری نموده که حاوی سیگنالهای خطر بیشتری هستند که در نتیجه تخریب سلّولی بوجود آمده اند. پس از اتمام برچسب زنی و مهاجرت تمامی سلّولها، موقعی که دیگر هیچ سلّول دندریتی باقی نماند که مهاجرت نکرده باشد، فاز دسته بندی آغاز می شود. فاز دسته بندی ، برای هر آنتی ژن محاسبه می شود که توسط چه تعداد سل.دن.بغ. و چه تعداد سل.دن.نبغ. نمونه برداری present شده است؟

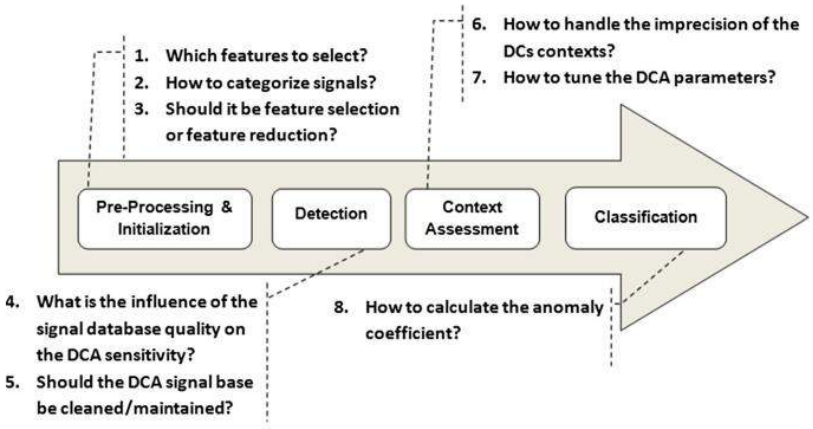
اگر آنتی ژن مذکور را سل. های دن.بغ. بیشتری نمونه برداری کرده باشند ، آنتی ژن برچسب آنومالی می خورد و MCAV یک می شود بدین معنا که احتمال زیادی دارد که نرمال نباشد. در غیر اینصورت $MCAV = 0$ شده و آنتی ژن برچسب نرمال می خورد.

دو مورد از نقاط ضعف مهم این الگوریتم ، نمونه برداری تصادفی آنتی ژنها توسط سلّولهای دندریت و همچنین روابط محاسبه و برچسب زنیِ گسسته برای MCAV می باشد. زیرا با توجه به اینکه مسئله ما پیوسته بوده و برچسب های نمونه ها همگی بصورت احتمالی در بازه [0,1] تعیین می گردند و همچنین ترافیک شبکه نیز

از قبل مورد نرمالیزه سازی $\min - \max$ قرار میگیرد، پس به نظر می رسد استفاده از روابطی که در نهایت منجر به تخصیص احتمالیِ وضعیّتِ نرمال/ آنومالیِ ترافیک در محاسبه مقادیر MCAV گردند ، مناسبتر باشند.

۲-۴-۱-۳- محاسبه سیگنالهای ورودی

به طور کلی تاکنون دو روش برای انتخاب و نگاشت مناسب ترین زیر مجموعه ویژگیها به سیگنالهای ورودیِ DCA ارائه گردیده است. اوّلی استفاده از تکنیک کاهش/ استخراج ویژگی و روش دوّم استفاده از دانش متخصصان امنیت شبکه. [۴۶] [۱۵] هر دو روش چالشها و نقاط قوّتی دارند از جمله اینکه چالش اصلی استفاده از روش کاهش ابعاد $\mathrm{NP} - \mathrm{Hard}$ بودن مسئله می باشد ، زیرا هر متد کاهش/ استخراج ویژگی با 2^n زیر مجموعه ویژگی رو به رو خواهد بود که حل آن ؛ یعنی جستجو و یافتن زیر مجموعه ویژگیهای مناسب؛ با افزایش بُعد مسئله سخت تر خواهد شد.



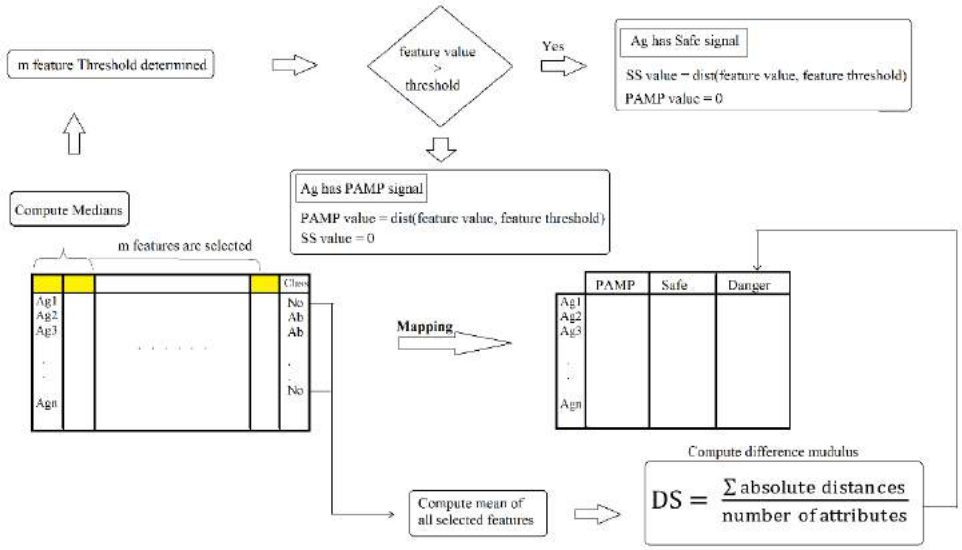
شکل ۶ – چالشهای الگوریتم دندریت استاندارد [۱۵]

روش دوم نیز نقات قوّت و چالشهای خود را دارد و بیشتر وابسته به بروز بودنِ دانشِ متخصصانِ آنالیز ترافیک شبکه می باشد. البته در حال حاضر این روش چندان دقّت و عملکرد دسته بندی بالایی ندارد و نمی توان صرفاً بدان وابسته بود.

به منظور کشفِ (استخراج) موثرترین زیر مجموعه ویژگیها جهت نگاشت آنها به سه سیگنال ورودی ، می توان از روشهای کاهش ابعاد مختلفی مانند PCA. CC. IG برای فاز پیش پردازش خودکارِ ترافیک ورودی استفاده کرد. این فاز $\mathrm{initialize} - \mathrm{DC}()$ می باشد. شکل زیر مهمترین چالشهای هر یک از فازهای الگوریتم سلّولهای دندریت را نشان می دهد.

۳-۴-۱-۳- فاز تشخیص

الگوریتم سلّولهای دندریت در فاز تشخیص ، مقادیر این سه سیگنال را برای تمام نمونه ها محاسبه می کند. این مقادیر در جدول سیگنالها با فرمت زیر ذخیره می شوند. سطر ها نشان دهنده آنتی ژن و سه ستون نیز معرّف نوع سیگنال ورودی می باشند.



شکل ۷ - پایگاه سیگنالها

قبل از محاسبه سیگنالهای ورودی ، باید بدانیم که کدام ویژگیهای دادگان را برای نگاشت به سیگنالهای ورودی انتخاب کنیم. پس از انتخاب و نگاشت مناسبترین ویژگیها به سه سیگنال ورودی با استفاده از یکی از دو روش انتخاب ویژگی یا استفاده از دانش متخصصان ، نوبت به فرایند پردازشِ سیگنالهای امن ، خطر و PAMP فرا می رسد که برای این سه سیگنال فرایندی متفاوت وجود دارد. سیگنال PAMP نسبت به سیگنال خطر نماینده بهتری برای آنومالی بودن می باشد. [۱۵] از طرفی دو سیگنال امن و PAMP روی یک ویژگیِ مناسب به صورت زیر بدست می آیند. به بیان بهتر سه ویژگی به سه سیگنال نگاشت نمی شود بلکه یک زیر مجموعه ویژگی برای محاسبه دو سیگنال امن و PAMP و زیر مجموعه ویژگی دیگر نیز صرفاً برای محاسبه سیگنال خطر با دو روش اشاره شده در بالا بدست می آیند.

۳-۴-۱-۴- تخصیص سیگنالها

۳-۴-۱-۴-۱- فرایند محاسبه (تخصیص) سیگنالهای SS و PAMP

- انتخاب ویژگی مناسب با استفاده از یک روش انتخاب ویژگی (مثلاً IG ویژگیهای با آنتروپی بالا را پیدا می کند). این ویژگی از نوع عددی پیوسته و غیر سمبلیک می باشد. (یعنی به جز ویژگیهای ۲ و ۳ و ۴ در دادگان نفوذ)
- محاسبه میانه [1] برای تمام مقادیر ویژگی انتخاب شده به عنوان "حد آستانه" برای هر دو کلاس آنومالی و نرمال.
- اگر مقدار ویژگی انتخاب شده بیشتر از حد آستانه تعیین شده باشد ، بیانگر آنست که آن الگوی ترافیکی (آنتی ژن) ، سیگنال "امن" دارد و قدرمطلق تفاضل مقدار ویژگی تا "میانه" به عنوان اندازه سیگنال امن در نظر گرفته شده و اندازه سیگنال PAMP نیز صفر می شود. در غیر اینصورت اگر مقدار ویژگی کمتر از حد آستانه (میانه) بود، سیگنال آنتی ژن (ترافیک شبکه) از نوع PAMP بوده و اندازه آن برابر قدرمطلق تفاضل مقدار ویژگی از مقدار حدآستانه بوده و این بار سیگنال امن ، صفر می شود.

۲-۴-۴-۱-۳- فرایند محاسبه (تخصیص) سیگنالِ خطر

برای تولید اندازه سیگنالهای خطر نیز فرایند مشابهی اجرا میگردد. به منظور محاسبه این سیگنال برچسب کلاسِ دادگان نفوذ باید از قبل مرتّب و مشخص شده باشند. ترکیبی از ویژگیهایی که به نظر می رسد مقادیر آنها بتوانند به عنوان سیگنال خطر استفاده شوند در این مورد به کار می روند.

۱- میانگین تک تک ویژگیهای انتخاب شده و کاندید برای سیگنال خطر ، به طور جداگانه برای تمامی داده هایی که برچسب کلاس آنها "نرمال" میباشد ، محاسبه می گردد.

۲- قدرمطلق تفاضل تمام مقادیر ویژگیها از میانگین همان ویژگی ، یک به یک برای داده های ترافیک نرمال محاسبه می شود.

۳- برای هر داده ، میانگین مقادیر قدر مطلق تفاضل ویژگیهای انتخاب شده ، طبق رابطه زیر محاسبه می شود:

$$DS = \frac{\sum \text{absolute distances}}{\text{number of attributes}} \qquad (۳-۱)$$

۴- تکرار این فرایند برای تمام داده ها.

در محاسبه سیگنالهای ورودی دادگان آزمون چالشهای زیر وجود دارند :

[1] Median

- عدم اطلاع از برچسب دادگان آزمون.
- از آنجایی که برای محاسبه سیگنال خطر ، فقط نمونه های با برچسب کلاس نرمال مورد نیاز هستند بنابراین نیاز است که ویژگی کلاس دادگان $pre-Ordered$ و از قبل مرتّب شده باشند. این یک مشکل اساسی است زیرا هدف از دسته بندی هم همین است و برچسب ها از قبل مشخص نیست.

- **یک ایده موثّر**

به منظور رفع چالشهای فوق ، از دادگان نفوذ یادگیری استفاده کردیم. این دادگان دارای برچسب بوده و می توان به راحتی بر اساس ویژگی کلاس آنرا مرتّب و سیگنالهای خطر را محاسبه نمود. امّا با این وجود مشکل اساسی هنوز باقی می ماند و آن چگونگی تخصیص سیگنالها به نمونه های دادگان آزمون می باشد. برای رفع این مشکل نیز به صورت زیر عمل نمودیم :

ابتدا فاصله اقلیدسی هر نمونه داده آزمون از تمام نمونه های نرمالِ دادگان یادگیری اندازه گیری می شوند و نمونه هایی که نزدیکترین فاصله را با نمونه آزمون دارند به عنوان الگوی کاندید برای محاسبه سیگنالها انتخاب می شوند. به این ترتیب مشکل عدم وجود داده مرتّب شده بر اساس ویژگی کلاس برچسب در دادگان آزمون و همچنین عدم مشخص بودن برچسب این دادگان جهت محاسبه سیگنالهای ورودی حل خواهد شد. سودوکد الگوریتم پیشنهادی برای ایجاد دادگان آزمون دارای برچسب کلاس احتمالی (همزادِ دادگان آزمون واقعی) به صورت زیر می باشد. با استفاده از این الگوریتم ، دادگان آزمون را بر اساس برچسب های احتمالی مرتّب سازی نمودیم تا سیگنال خطر را برای نمونه ها محاسبه نماییم.

Alrogithm 2 . Psudo code of proposed idea for Danger signals assignment

Input: $Antien_{Set}$. $Self_{Normal_{Set}}$. $Distance_{Threshold}$
Output: $PreOrdered_{Antigen\_set}$
For all $Ag \in Antigen_{Set}$
    $Candidate_{dist}$ = Min Euclidian distance Ag from $Self_{Normal_{Set}}$
    If $Candidate_{dist}$ be normal, then
        Set Ag's class label to normal
        Add $Candidate_{dist}$ to $Dist_{Buffer}$
    End if
End for
If No. of candidates in $Dist_{Buffer}$ be very close to size of $Antien_{Set}$, then
    Sort $Dist_{Buffer}$ by ascending order
    Select first **N** number from buffer have distance lesser than $Distance_{Threshold}$
    Set class labels of **N** respective antigens to normal
End if
$PreOrdered_{Antigen\_set}$ = Sort $Antien_{Set}$ on class label

شکل ۸ – شبه کد مربوط به الگوریتمِ ایده پیشنهادی برای رفع مشکل مرتّب نبودن دادگان آزمون در محاسبه سیگنال خطر

۳-۴-۴-۱-۳- ماتریس ضرایب وزنی

برای محاسبه سیگنالهای خروجی (پردازش سیگنال) نیاز به روابط و ضرایب وزنی می باشد. از دیدگاه زیستی، پردازش سیگنالهای ورودی آنتی ژنها توسط سلّول دندریتی صورت میگیرد که الگوی آن آنتی ژن را نمونه برداری کرده باشند. رابطه پیشنهادی برای محاسبه سیگنالهای خروجی به شکل زیر می باشد. این رابطه سه بار تکرار می شود تا مقادیر سیگنالهای خروجی بدست آید. [۱۵] C ماتریس ضرایب وزنی می باشد.

$$C = \frac{[(W_{PAMP} \times \sum_i PAMP_i) + (W_{SS} \times \sum_i SS_i) + (W_{DS} \times \sum_i DS_i)]}{(W_{PAMP} + W_{SS} + W_{DS})} \times \frac{1 + Inf}{2} \qquad (۳-۲)$$

که Inf سیگنالِ چهارم (التهاب) می باشد که ۱ فرض شده و وزنها نیز بوسیله متخصّص تعیین می شوند.

۵-۴-۱-۳- پردازش سیگنالهای ورودی [1]

پس از تعیین برچسب احتمالی نمونه ها در دادگانِ آزمون بر اساس ایده پیشنهادی و محاسبه مقادیر هر سه سیگنال ورودی ، نوبت به پردازش این سیگنالها می رسد. در این فاز مقادیر سیگنالهای خروجی با در نظر گرفتن ماتریس ضرایب وزنی بدست می آیند. گرینسمیت [2]در رساله دکترای خود برای اوّلین بار پردازش سیگنالها را مُدلسازی نمود و در الگوریتم سلولهای دندریت پیشنهادی خود به کار برد. [۴۵] وزنهایی که او برای پردازش سیگنال از آنها استفاده کرده به صورت ماتریس $\begin{vmatrix} \text{Signals} & \text{PAMP} & \text{SS} & \text{DS} \\ \text{CSM} & 2 & 1 & 2 \\ \text{smDC} & 0 & 0 & 2 \\ \text{mDC} & 2 & 1 & -2 \end{vmatrix}$ می باشد.

همچنین در مقالات جدیدتری مانند [۱۵] نویسنده از وزنهای $\begin{vmatrix} \text{Signals} & \text{PAMP} & \text{SS} & \text{DS} \\ \text{CSM} & 2 & 1 & 2 \\ \text{smDC} & 0 & 0 & 1 \\ \text{mDC} & 2 & 1 & -1.5 \end{vmatrix}$ برای پردازش سیگنالها استفاده کرده است. در مجموع در بیشتر پژوهشها توصیه شده که مقادیر وزنها توسط متخصّصان خِبره تعیین گردند. تعیین وزنها اهمیت زیادی در خروجی دسته بندیِ DCA دارد بطوریکه پس از وزندهی و پردازش سیگنالها ، در نهایت این مقدار سیگنال CSM است که مشخص خواهد کرد که آنتی ژنِ نمونه برداری شده ، بالغ کننده یا نیمه بالغ کننده می باشد. حال اگر مقادیر وزنها اشتباه تخصیص داده شوند منجر به مهاجرت اشتباه سلّولهای دندریت شده و خطای دسته بندی بالاتر می رود. پس برای پردازش صحیح سیگنالها و کارآمدیِ کیفیت دسته بندی دو مورد زیر باید به طور صحیح انجام گیرد : ۱) تخصیص صحیح سیگنالهای ورودی و محاسبه دقیق میزان بالغ یا نیمه بالغ کنندگی برای هر آنتی ژن ۲) تعیین صحیح وزنها. همانطور که در بخش های قبل نیز بیان گردید سه سیگنال خروجی به ترتیب سیگنالهای بالغ (وضعیت خطر)، نیمه بالغ (وضعیت آماده باش/ نرمال) و سیگنال CSM می باشند. سیگنال CSM به دلیل اینکه از ترکیب وزنی

[1] Signal Processing equation
[2] Grinsmit

هر سه سیگنال ورودی امن ، خطر و PAMP بدست می آید می تواند معیار مناسبی برای بررسی شرایط مهاجرت سلّول دندریت باشد.

جهت بررسی مقادیر سیگنالهای CSM مقدار حد آستانه لازم برای مهاجرت[1] باید از قبل مشخص باشد. این حد آستانه معمولاً در تمامی فازهای اجرای DCA ثابت تعیین می شود. در قسمت بعد ، ایده ای را با طرح این سؤال مطرح و بررسی نمودیم :

سؤال : آیا می توان مقدار حد آستانه مهاجرت را صرفاً به زیر مجموعه ویژگیهای کشف و نگاشت شده به سیگنالهای ورودی وابسته نمود و امکان تغییر پویای این مقدار حدآستانه را در هر فاز تکرار DCA فراهم نمود؟ استفاده از حد آستانه مهاجرتِ پویا چه تاثیری در خروجی عملکردِ دسته بندی این الگوریتم دارد؟

۴-۱-۳-۱-۵- یک ایده – پویا نمودنِ حد آستانه ی مهاجرتِ سلّولهای دِندریت

در الگوریتم .1 پارامتر migration threshold به عنوان حد آستانه مهاجرت سل. های دن.بغ / دن.نبغ. به صورت پیشفرض ثابت بوده و به عنوان شرط لازم جهت مهاجرت سلّولها در هر فاز تکرار الگوریتم با مقدار CSM بدست آمده مقایسه می گردد. مقدار این پارامتر در بازه [0,1] به صورت احتمالی توسط کاربر تعیین و به ورودی الگوریتم داده می شود. به نظر میرسد مقدار این پارامتر تا حد زیادی وابسته به سیگنالهای بدست آمده در هر فاز تکرار باشد. بطوریکه در الگوریتم .1 مشاهده می گردد در هر فاز تکرار ، مهاجرت سلّولهای دَندریت بر اساس شرط تجاوزِ مقادیر سیگنال خروجی CSM از حد آستانه تعیین شده صورت می گیرد.

حال با توجه به اینکه مقدار سیگنال CSM با در نظر گرفتن ضرایب وزنی سیگنالها، وابسته به سه سیگنال ورودی بوده و هر سلّول مقدار CSM متفاوتی دارد، بنابراین ایده ای که به ذهن می رسد لزوم پویایی حد آستانه مهاجرت سلّولها و وابسته نمودن آن ، به میانگین / میانه مقادیر CSM سلّولها در هر فاز تکرار است. زیرا در هر فاز مقدار سیگنالها برای سلّولهایی که تاکنون مهاجرت نیافته اند مرتباً بروز شده و به نظر میرسد مقدار این سیگنال خروجی (CSM) تاثیر مستقیمی بر تعیین حدّ آستانه مهاجرت داشته باشد. از اینرو ما رابطه زیر را برای تعیین پویای حد آستانه مهاجرت در هر فاز تکرار پیشنهاد و در الگوریتم پیشنهادی برای خط نخستِ دفاعی از این ایده استفاده نمودیم. این مورد در بخش ارائه استراتژی پیشنهادی بررسی و ارزیابی شده است.

$$mt = \left((\text{median}(Ppamp) * 2) + \text{median}(Pss) + (\text{median}(Pds) * 2)\right) \quad (۳-۳)$$

$$mt = \left((\text{mean}(Ppamp) * 2) + \text{mean}(Pss) + (\text{mean}(Pds) * 2)\right) \quad (۴-۳)$$

; Ppamp . Pss. Pds are Probabilities of Three Input Signals for immature DCs

[1] Migration Threshold

۲-۵-۴-۱-۳- آزمایش

نه فقط DCA بلکه تقریباً تمامی الگوریتم های فرا ابتکاری و متدهای یادگیری ماشین ، بدون اعمال فاز پیش پردازش و تکنیک کاهش ابعاد نتیجه مطلوبی را ارائه نمی دهند. امّا با این وجود ، به دلیل آنکه استراتژیِ پردازشِ متدهای ایمنی مصنوعی مانندDCA ، متفاوت با بقیه تکنیکها بوده و ذاتاً به شکلِ توزیع شده و غیر متمرکز می باشد در نتیجه بهترین گزینه برای مسئله تشخیص نفوذ و دسته بندی ترافیک شبکه هستند. از جمله مزیّتهای الگوریتم های دسته بندی مبتنی بر ایمنی مصنوعی ، توانایی مقابله با نفوذ از طریق اتّکا بر خود-یادگیری و تعدیل پروفایلهای خود ساخته می باشد بدون نیاز به کوچکترین فیدبکی از محیط [۱]. از جمله معایب آنها نیز مصرف زمانی و حافظه پردازشی بالای آنهاست. [۳۸]

از اینرو این الگوریتم در بدترین حالت خود ، با پارامترهای پیشفرض و بدون اعمال فاز پیش پردازش / کاهش ابعاد ، عملکردی به مراتب بهتر از الگوریتم های سنّتی و متاهیوریستیکِ رایج مبتنی بر جمعیّت نمونه اوّلیه مانند $k-means$ ، GA. ACO. Bee. BBO. BHA و PSO ارائه می دهد.

به منظور آشنایی با عملکردِ کاهش ابعاد با الگوریتم جستجو و انتخاب ویژگی دَه – پا ، CFA ، در [پ-الف-۴] ما ضمن معرّفیِ این روش ، آزمایشی را در دادگان نفوذِ $NSL-KDD$ جهت ارزیابی عملکردِ دَه-پا در مقایسه با $NSGA-II$ به انجام رساندیم.

دلیل مهمی که در تفاوت میان دسته بندهای مبتنی بر ایمنی مصنوعی با سایر دسته بندهای رایج و فراابتکاری مطرح می شود در شیوه متفاوت حلّ مسئله می باشد.

- سیستم های ایمنی مصنوعی با رویکرد کاملاً توزیع شده و غیر متمرکز مسئله دسته بندی را حل می کنند. موفقیّت این تکنیکها بیشتر بستگی به حجم حافظه پردازشی کافی برای اجرای متد ایمنی مصنوعی با توجه به تکرارهای زیاد و همچنین میزان سازی صحیح حدود آستانه مهاجرت سلّولهای دندریت و تولید و تکثیر آنتی بادیها و بطورکلّی بهینه سازی پارامترهای ورودی دارد.
- تکنیکهای دسته بندی سنّتی و مبتنی بر یادگیری ماشین ، مسئله تشخیص و دسته بندی نفوذ را اغلب با رویکردی کاملاً متمرکز حل میکنند. در اصل شیوه حل مسئله با رویکرد این تکنیکها بیشتر بهینه سازی از طریق آزمون و خطای پاسخ های بدست آمده در جهت دستیابی به بهترین موقعیّتهای مرکزیتهای مطلوب می باشد.

[۱] این سیستم ها فقط وابسته به پایگاه امضاء ها نبوده و بیشتر بر خود ایمنی از طریق آزمون های درونی سیستم با تولید تش.د. های بغ.م. تاکید و تکیه دارند. برخی از تشخیص دهنده های بالغ بسته به موقعیّتی که در فضای ابعاد مسئله دارند به شکل توزیع شده و کاملاً نامتمرکز توانایی بالقوّه ای را شناسایی حملات ناشناخته دارند.

یک نکته

تکنیکهای دسته بندی رایج مبتنی بر یادگیری ماشین یا فراابتکاری برای حل مسئله خاص "دسته بندی باینری ترافیک شبکه " به منظور کشف ناهنجاری راهکار مناسبی نیستند به این دلیل که توزیع دادگانِ ترافیکِ شبکه به گونه ای است که اغلب ممکن است چندین خوشه نرمال / آنومالی داشته باشیم بطوریکه خوشه های آنومالی هر یک خود متعلّق به یک زیر مجموعه حمله باشند.

بنابراین در این مورد خاص بیشتر متدهای یادگیری ماشین و الگوریتم های فراابتکاری قادر نیستند مسئله دسته بندی باینری ترافیک شبکه را حل نمایند.[پ-الف-۷] در ادامه آزمایشات خواهیم دید که در مورد مسئله دسته بندی باینری ترافیک شبکه ، متدهای دسته بندی مبتنی بر ایمنی مصنوعی نتایج رضایت بخشی را ارائه نموده و از عملکرد بهتری در شناسایی حملات ؛ خصوصاً حملات ناشناخته برخوردار هستند . همچنین متدهای مذکور توان مقابله با حملات روز صفر [1]را نیز دارند.

۳-۵-۴-۱-۳- آسیب پذیری روز صفر[2]

به سوء استفاده از یک آسیب پذیری امنیّتی و درست در همان روز و قبل از کشف و اِعلانِ عمومی آن آسیب پذیری ، حمله روز صفر گفته می شود. معمولاً بیشتر سیستم های تشخیص نفوذ که صرفاً مبتنی بر پایگاه امضاءهای حملات عمل می کنند [3]، در برابر این نوع حمله دچار مشکل شده و در این شرایط fail می شوند. سیستم های تشخیص نفوذ مبتنی بر ایمنی مصنوعی توان مقابله با این حملات را دارند و این ، از جمله مزیّتهای این متدها به شمار می رود.

ما برتریِ DCA را با عملکرد دو الگوریتم مِتاهیوریستیک BHA و PSO برای مسئله خوشه بندی باینری مقایسه نمودیم. نتایج بدست آمده در آزمایشات زیر و همچنین نتایجِ آزمایشات [پ-الف-۲] این برتری را به خوبی اثبات میکند.

یک دلیل مهم برای اثبات این مسئله ، تفاوت در شیوه محاسبات برای حل مسئله است بطوریکه در اکثر متدهای یادگیری ماشین و الگوریتم های فراابتکاری مبتنی بر جمعیّت نمونه اوّلیه ، در مسائل دسته بندی باینری معمولاً موقعیّت دو مرکزیّت نرمال و آنومالی به شکل کاملاً تصادفی و یکنواخت انتخاب شده و الگوریتم دسته بندی برای دستیابی به بهترین موقعیّت های دو مرکزیّت (دو خوشه) مرتباً موقعیّت ها را بر اساس استراتژی خاص خود بهینه سازی و بهبود می دهند. در اغلب این الگوریتمها ، انتخاب تابع فیتنس / تابع هزینه مناسب ، بسیار

[1] Zero Day attacks
[2] Zero Day Exxploit period
[3] Misuse detection techniquies

موثر بوده و نتایج رضایت بخشی را به دنبال دارد. در حالت کلّی موقعیّتِ مرکزیّتهای بدست آمده پیوسته در حال بهبود بوده بطوریکه الگوریتم پس از چندین فاز تکرار بر اساس مقدار بهینه تابع فیتنس می تواند بهترین دو مرکزیت بدست آمده را به عنوان سولوشن مسئله ارائه دهد. توجه شود که در کلیه مسائل دسته بندی نظارت شده /نشده ، بهترین موقعیتهای n مرکزیّت بدست آمده ، فقط یک "راه حل" مطلوب از بین تمام راه حل های ممکن می باشد نه n راه حل. در واقع مسئله مسئله ای سخت (NP-Hard) خواهد بود.

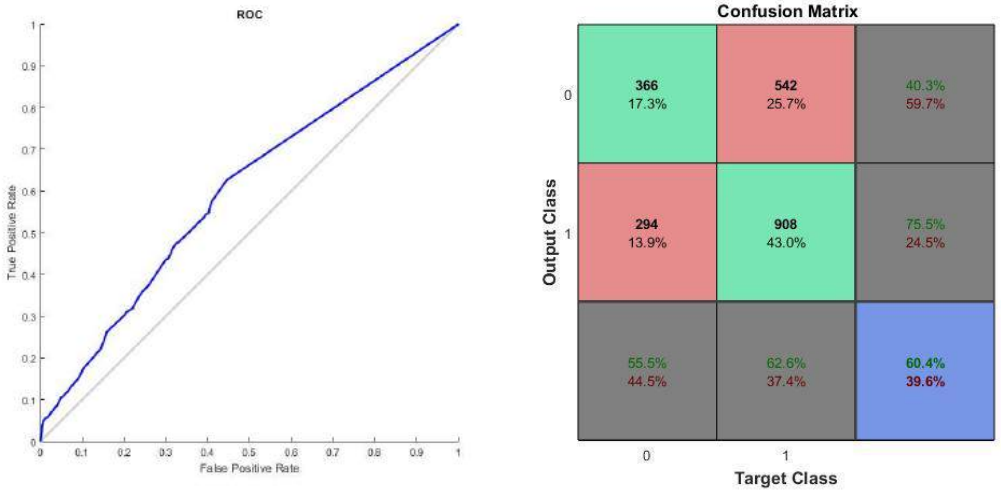
ما الگوریتم1. را در محیط نرم افزار متلب شبیه سازی نمودیم و با دادگان نفوذ UNSW – NB15 و با ۴۳ ویژگی (بدون کلاس) و بدون اعمالِ فاز پیش پردازش و کاهش ویژگی ، آزمون کردیم. نتایج زیر بدست آمد :

جدول ۲ - تخصیص پارامترهای ورودی DCA - تغییر پارامترهای ورودی الگوریتم سلولهای دندریت استاندارد در ۸ بار اجرا

Run	***Steps***	***Selected_Features***	***mt***	***at***	***runtime (s)***	***Min_Clone_Ag***	***Max_Clone_Ag***
*run*1	8	*all*	50	50	26.4219	5	10
*run*2	8	*all*	40	40	26.4219	5	15
*run*3	1	*all*	40	50	17.1413	5	25
*run*4	1	*Best Features* : 8.9.30.33.41.42 *Ref*. [16]	50	50	15.8608	5	10
*run*5	1	*Best Features* : 8.9.30.33.41.42 *Ref*. [16]	35	25	15.3479	10	40
*run*6	8	*middle ranked Features* : 4.5.6.10.11.15.18.19.20.21.22.23.24.25.26.31.32.40 *Ref*. [16]	50	50	26.7069	5	10
*run*7	1	*middle ranked Features* : 4.5.6.10.11.15.18.19.20.21.22.23.24.25.26.31.32.40 *Ref*. [16]	35	35	16.1472	10	40
*run*8	1	*By CFA* [14] : 1.2.3. 4.5.6.8.9.10.11.14.18.19.20.21.22.23.24.30.32.33.41.42	40	40	19.0889	3	10

جدول ۳ – عملکرد الگوریتم سلولهای دندریت استاندارد در ۸ بار اجرا

$Parameter$	DR	FPR	FNR	$Accuracy$	$misclassification_{rate}$	$Correct\_Prediction\_normal$	$Correct\_Prediction\_attack$
$run1$	51.4	46.5	48.6	52.0	48.0	33.4	70.8
$run2$	48.2	34.8	51.8	53.5	46.5	36.4	75.2
$run3$	46.3	33.9	53.7	52.5	47.5	35.9	75.0
$run4$	51.0	34.1	49.0	55.6	44.4	38.0	76.7
$run5$	39.4	24.5	60.6	50.7	49.3	36.2	77.9
$run6$	54.3	42.3	45.7	55.4	44.6	36.5	73.9
$run7$	45.9	27.9	54.1	54.1	45.9	37.7	78.3
$run8$	62.6	44.5	37.4	60.4	39.6	40.3	75.5
$Best$	$run8$						



نمودار ۱ – ماتریس درهم ریختگی و منحنیِ ROC مربوط به اجرای هشتم و کاربرد انتخاب ویژگی CFA

ملاحظه می شود که نتایج دسته بندی مناسب نبوده و نیازمند بهبود می باشند. با توجه به اینکه مسئله ما مسئله دسته بندی نظارت نشده ترافیک شبکه و بدون یادگیری قبلی می باشد ، جهت کاهش خطاهای مثبت و مخصوصاً منفی اشتباه و همچنین ارتقای نرخ های عملکردی و تشخیص ، استفاده از ایده های ترکیبی (مثل ترکیب دسته بندها) نتایج خوبی را ارائه می دهند. به عنوان نمونه در [۲] که به بررسی جامع کاربردهای سیستم های ایمنی مصنوعی در امنیّت کامپیوتر پرداخته ، استفاده از ایده ترکیب hybrid متدهای ایمنی مصنوعی با تکنیکهای یادگیری ماشین نتایج رضایت بخشی را ارائه داده است.

همچنین در [۳۴] نیز به استفاده از ایده ترکیب توصیه شده و یکی از مشکلات اساسی سیستم های تشخیص مبتنی بر ایمنی مصنوعی ، زمان تشخیص بالا بیان شده است بطوریکه تاکید شده که از پلتفرم زیر ساخت ابری؛ سرویس IaaS؛ جهت تامین منابع لازم برای پردازش و حافظه و جبرانِ بلادرنگ بودن زمان تشخیص استفاده گردد ضمن اینکه سیستم های تشخیص مبتنی بر ایمنی مصنوعی در محیط ابری توسعه داده شوند.

همانطور که بیان گردید به دلیل آنکه شیوه ارزیابی دسته بندهای مبتنی بر سیستم ایمنی مصنوعی با تکنیکهای یادگیری ماشین متفاوت بوده و بر مبنای "برقراری ایمنی درونی سیستم بواسطه تولید و تعدیل الگوهای نرمال (خودی) " اقدام به کشف رفتار ناهنجار (غیرخودی) می نمایند، در نتیجه ما ضمن ارائه ایده ترکیبی بیشتر بر روی بهینه سازی الگوریتمهای ایمنی مصنوعی و توابع استفاده شده در آنها متمرکز شدیم.

۵-۱-۳- ارائه استراتژی پیشنهادی برای الگوریتم DC

در این بخش به بیان و ارزیابی ایده هایی می پردازیم که می توانند DCA استاندارد را کارامدتر نمایند. دستیابی به این ایده ها و آزمون و ارزیابی کارآمدی آنها در جهت بهبود DCA به مرور و با کسب تجربه کار با این الگوریتم و پارامترهای آن در محیط شبیه ساز متلب حاصل شده است.

۱-۵-۱-۳- یک ایده – سازوکار نمونه برداری [1] از بُردار تکثیر شده از نمونه های ترافیک شبکه؛[2]

در این پژوهش ، موقعیت [3]های تمامی سلّولهای دندریت ، متناظر با موقعیت های نمونه های ترافیک شبکه در فضای ابعاد مسئله است. البته می توان این موقعیتها را به صورت تصادفی نیز در نظر گرفت اما این خود استراتژی الگوریتم استاندارد می باشد. با توجه به اینکه نمونه برداری از آنتی ژنها در DCA استاندارد؛ تابع Get_Antigen() ؛ سازوکار مشخصی نداشته و صرفاً به صورت تصادفی نمونه برداری را انجام می دهد، ما سازوکاری را پیشنهاد نمودیم که عیناً مطابق کارکرد زیستی این سلّولها عمل می نماید.

روابط پیشنهادی برای نمونه برداری بدین صورت است که ابتدا فواصل اقلیدسی هر نوع نمونه در دادگان آزمون با سایر نمونه ها در همان دادگان محاسبه شده و سپس میانگین / میانه / ماکزیمم این فاصله تعیین و رابطه پیشنهادی زیر روی آن اعمال می شود.

[1] Sampleing
[2] Cloned network traffic vectors
[3] Position

$$\text{Present\_raduis}_{DC(j)} = \left[\frac{\text{Clone}_{\text{Vector}(j)}}{\max(\text{Clones}_{\text{Vector}})}\right] \times \max\big(\text{dist}(DC(j).i)\big) \quad (۳-۵)$$

$$\text{Present\_raduis}_{DC(j)} = \left[\frac{\text{Clone}_{\text{Vector}(j)}}{\max(\text{Clones}_{\text{Vector}})}\right] \times \text{mean}\big(\text{dist}(DC(j).i)\big) \quad (۳-۶)$$

$$\text{Present\_raduis}_{DC(j)} = \left[\frac{\text{Clone}_{\text{Vector}(j)}}{\max(\text{Clones}_{\text{Vector}})}\right] \times \text{median}\big(\text{dist}(DC(j).i)\big) \quad (۳-۷)$$

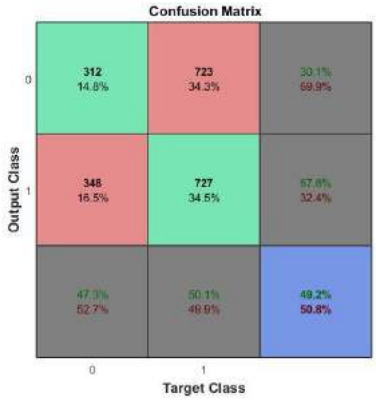
به منظور تعیین صحیح این که کدام یک از سه رابطه پیشنهادی بالا برای نمونه برداری سلّولهای دندریت مناسب می باشند آزمایشی را به صورت زیر انجام دادیم. خروجی آزمایش نشان می دهد که معیار میانه می تواند محدوده پوشش واقع بینانه تری را برای سلّول دندریت در جهت نمونه برداری آنتی ژنها ارائه دهد در نتیجه رابطه (۳-۷) در الگوریتم پیشنهادی خط نخستِ دفاعی به کار گرفته شد. در این سه آزمایش ، پارامترهای ورودی به شکل زیر تعیین گردیدند.

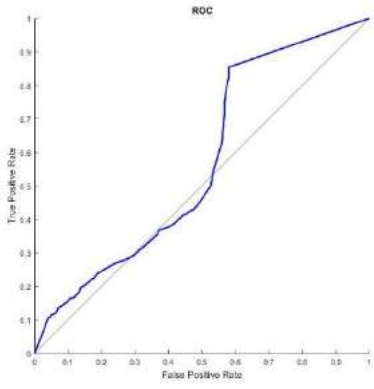
جدول ۴ - تخصیص پارامترهای ورودی

Parameter :	*Steps*	*Selected_Features*	*mt*	*at*	*Min_Clone_Ag*	*Max_Clone_Ag*
*run*1	1	*all*	40	50	5	10

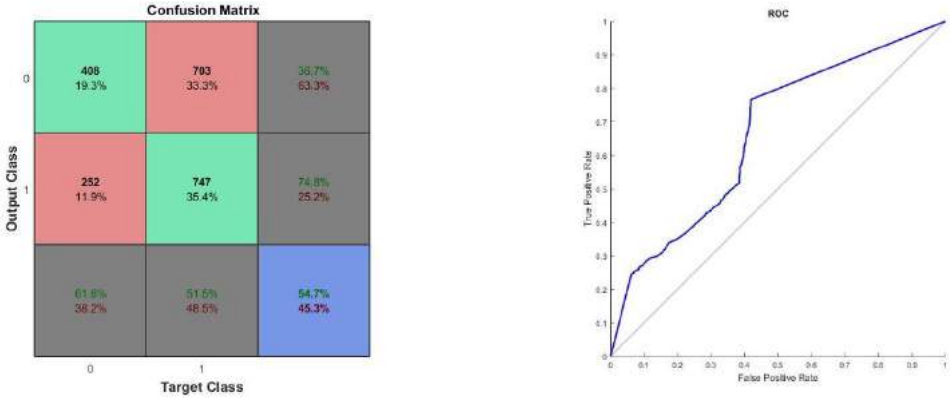
جدول ۵ - خروجی های سه آزمایش

Test	*runtime*	$Antigen\_presentation_{formule}$	*iterations*
1	31.8561	1	42
2	16.7096	2	42
3	18.3378	3	42

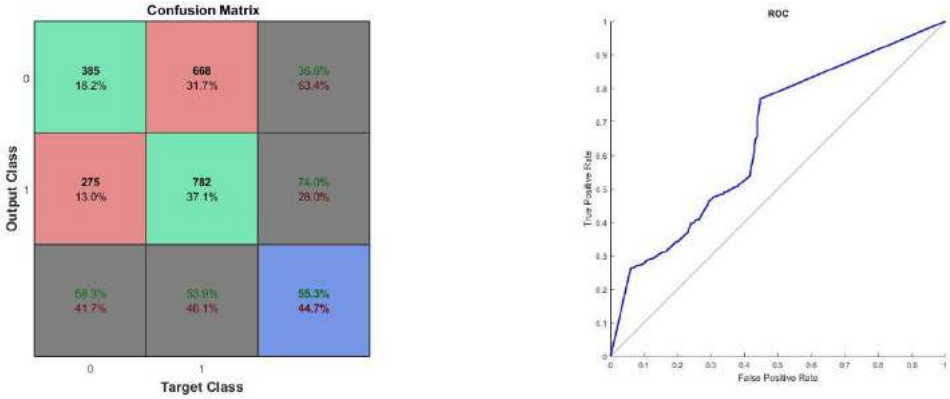




الف) استراتژی شعاع پوشش سلول دندریت با اعمال رابطه ۳-۵ جهت نمون برداری آنتی ژنها



ب) استراتژی شعاع پوشش سلول دندریت با اعمال رابطه ۳-۶ جهت نمون برداری آنتی ژنها



ج) استراتژی شعاع پوشش سلول دندریت با اعمال رابطه ۳-۷ جهت نمون برداری آنتی ژنها

نمودار ۲ - خروجی های حاصل از سه آزمایش با اعمال روابط $(7-3).(6-3).(5-3)$ به عنوان استراتژی شعاع تحت پوشش سلولهای دندریت

با بررسی این سه نمودار و آزمون های بیشتر به این نتیجه رسیدیم که "میانه" می تواند در ارزیابی شعاع پوشش سلّولهای دندریت معیار استاندارد تری بوده و نتایج رضایت بخشی را ارائه دهد. بدین ترتیب با معیار میانه ، این بار آزمون دیگری با کاربرد الگوریتم انتخاب ویژگی $CFA-ANN$ با پارامترهای ورودی زیر انجام گردید. ما به نتایج زیر دست یافتیم. در این آزمایش نرخ حدود آستانه مهاجرت مطابق با بحث مطرح شده در بخش قبل به صورت متغیر بوده و همچنین ضرایب وزنی در این آزمایش تعیین گردید.

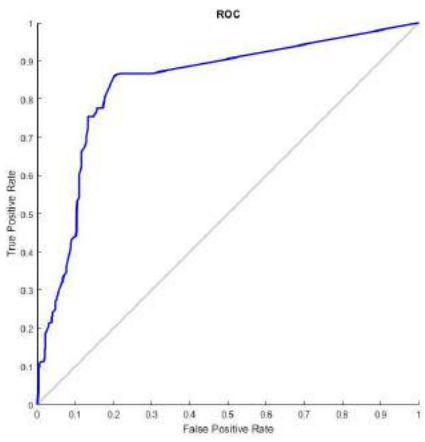
$$\begin{vmatrix} \text{Signals} & \text{PAMP} & \text{SS} & \text{DS} \\ \text{CSM} & 2 & 1 & 2 \\ \text{smDC} & 0 & 2 & 0 \\ \text{mDC} & 2 & 1 & -2 \end{vmatrix}$$

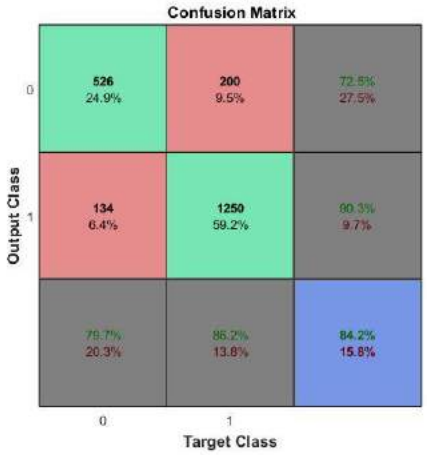
جدول ۶ - پارامتردهی پیشنهادی پس از کسب تجربه کار با DCA

Parameter	Step	Selected_Features	mapping & input Signal	mt	at	Min_Clone_Ag	Max_Clone_Ag
$run1$	1	$By\ CFA:$ 1.2.3. 4.5.6.8.9.10.11.14.18.19.20.21.22.23.24.30.32.33.41.42	IG	$Variable\ (mean)$	50	5	15

جدول ۷ - خروجی های آزمایش DCA پیشنهادی

$\boldsymbol{Test}$	$\boldsymbol{runtime}$	$\boldsymbol{Antigen\_presentation_{formule}}$	$\boldsymbol{iterations}$
1	17.5560	3	18





نمودار ۳- منحنیِ ROC ، ماتریس درهم ریختگی آزمایش DCA پیشنهادی با اعمال رابطه $(3-4)$و پارامترهای جدول ۷

دسته بندی ، در نهایت با چک کردن وضعیّت تمامی سلّولهای مهاجرت یافته صورت میگیرد. بدین معنی که در فاز تشخیص ، به عنوان مثال برای هر آنتی ژن چک می شود که در طول اجرای الگوریتم ، به چه میزان سل. دن.بغ / دن.نبغ. این آنتی ژن را نمونه برداری نموده اند. اگر تعداد سل. های دن.بغ. بیشتر بود ، به معنی آنست که به احتمال زیاد ، آنتی ژن مذکور الگویی با رفتار ناهنجار دارد یعنی شناسایی آنومالی. این احتمال وقتی بیشتر می شود که اختلاف تعداد سل.های دن.بغ. و سل. های دَن.نبغ. ای که این آنتی ژن را نمونه برداری کرده اند زیاد باشد.

بعضی کارهای پژوهشی مانند [۵۶] ایده جایگزینی فاز دسته بندی DCA را با یک تکنیک یادگیری ماشینِ سنّتی پیشنهاد کرده اند و به این نتیجه رسیده اند که استفاده از یک تکنیک سنّتی برای فاز دسته بندی برای دادگانی با نویز بالا و دسته بندیِ بلادرنگ ، مناسب بوده ولی برای پردازش دادگان مصنوعی (مانند مجموعه

دادگان هوش مصنوعی UCI) مناسب نیست. در ادامه با بررسی محدودیتها و نقاط ضعف این الگوریتم ، به ارائه ایده هایی نو جهت بهبودِ مکانیسمِ DCA می پردازیم.

۶-۱-۳- محدودیتهای DCA

DCA علیرغم اینکه گزینه مناسبی برای کاربرد در مسئله دسته بندی باینری خصوصاً در مسائل تشخیص نفوذ شبکه می باشد ، امّا ضعف هایی نیز دارد [۱۵] که به طور خلاصه در زیر بیان می کنیم:

۱- فاز پیش پردازش این الگوریتم که وظیفه نگاشت ویژگیهای مناسب و محاسبه سیگنالهای ورودی را بر عهده دارد، نقطه ضعف اصلی این الگوریتم دسته بندی می باشد. مطابق آزمایشات انجام شده در بخش های قبل در خصوص نحوه نگاشت سیگنالهای ورودی ، مشاهده شد که دو متد رایج برای اینکار وجود دارند :

- کاهش ابعاد
- تجربیات و دانش متخصصان امنیّت

در هر دو حالت مشکلاتی وجود دارد که بحث گردید. به طور خلاصه مشکل کاهش ابعاد اینست که چه ویژگیهایی را مطلوب دانسته و چه ویژگیهایی حذف گردند. این به دلیل آنست که مسئله انتخاب ویژگی اصولاً یک مسئله بهینه سازی سخت از نوع $NP-hard$ می باشد. [۳۹] به عبارت دیگر ، با افزایش ابعاد مسئله و تعداد ویژگیها متد دسته بند یادگیری به کاررفته در انتخاب ویژگی سخت تر می تواند بهینه ترین زیر مجموعه ویژگی را انتخاب کند. با توجه به اینکه کاهش ابعاد مسئله در اکثر موارد به صورت راپر صورت می پذیرد (مشابه آزمایشات ما) ، در نتیجه دسته بندِ یادگیری ممکن است ویژگیهای مناسب را بنا به استراتژی خود مطلوب ندانسته و حذف نماید. حال ممکن است این ویژگیها که از نظر دسته بندی یادگیری غیر مرتبط تشخیص داده شده و حذف شده اند، موثرترین ویژگیها برای نگاشت به سیگنالهای ورودی بوده باشند که در این صورت DCA خروجی مطلوبی نخواهد داشت. در نتیجه پیشنهاد می شود که متد کاهش ابعاد از نوع فیلتر باشد.

بدین منظور استفاده از یک خوشه بند به جای دو متد رایج (بهره اطلاعات و دانش متخصصان) جهت کشف و تخصیصِ سیگنالهای ورودیِ DCA در [پ-الف-۸] روابطی پیشنهاد گردیده است.

۱-۶-۱-۳- یک ایده و یک نتیجه ی منفی - استفاده از خوشه بند سیاه چاله برای تخصیص [1] احتمالات سیگنالهای ورودی

[1] Input Signals Probability Assignment (Mapping most appropriated features to three input signals)

در ابتدای کار پژوهشی پایان نامه ، ایده ای برای محاسبه سیگنالهای ورودی DCA از طریق خوشه بندی به ذهن ما رسید به این صورت که از یک متد خوشه بندِ موثّر مانند حفره سیاه یا ازدحام ذرّات برای "انتخاب ، نگاشت و محاسبه سیگنالهای ورودی" استفاده شود. دلیل انتخاب الگوریتم سیاه چاله ، تعداد ارجاعات بالای مقاله مربوط به این الگوریتم [۱۳] در طول سالیان گذشته است.

در نتیجه الگوریتم بهینه سازی حفره سیاه یا Black Hole بدین منظور استفاده و در طول آزمایشات با سایر متدهای خوشه بندِ فراابتکاری مشابه مبتنی بر جمعیت نمونه اولیّه نیز ارزیابی مقایسه ای گردید.[پ-الف-۲] استراتژی این ایده بدین صورت می باشد که ابتدا خوشه بندِ حفره سیاه بر روی نمونه های دادگان آزمون اعمال شده و دادگان دسته بندی باینری می شوند ، سپس مراکز خوشه ها تعیین و میانگین/میانه فواصلِ تمام نمونه ها از مراکز خوشه ها به عنوان مَرز مشخص هر خوشه تعیین می گردد.

در نهایت بر اساس فاصله هر نمونه آنتی ژنیک از مرکز/ مرز تعیین شده ، تمام خوشه ها و بسته به اینکه نواحی تحت پوشش خوشه ها چه اشتراکی با هم داشته باشند ، روابط خاصی ایجاد می شوند که این روابط احتمالات سه سیگنال امن ، خطر و PAMP را برای هر یک از نمونه ها تعیین می کنند. ایده پیشنهادی در ابتدا کارآمد به نظر میرسید ولی با بررسی های به عمل آمده و نتایج بدست آمده از آزمایشات [پ-الف-۲] مشاهده شد که هیچ کدام از الگوریتم های فرا ابتکاری مبتنی بر جمعیت نمونه اولیّه به دلیل آنکه مسئله تشخیص ناهنجاری شبکه را به صورت متمرکز حل می کنند نرخ های تشخیص بسیار پایینی داشته و گزینه مناسبی برای حل مسئله دسته بندی باینری ترافیک شبکه نمی باشند. هر چند در صورتی که مسئله چند کلاسه می بود ممکن بود همین متدِ حفره سیاه یا متدهای مشابه آن بتوانند نرخ های قابل قبولی ارائه داده و در تخصیص سیگنالها مفید باشند. امّا به دلیل آنکه عملکردِ خوشه بند تاثیر زیادی در محاسبه مراکز دقیق خوشه ها و مرزهای احتمالی و نواحی تحت پوشش آنها و در نهایت محاسبه و تخصیص سیگنالها می گذارد، در نتیجه این متدها نمی توانند گزینه مناسبی برای تخصیص احتمالات سیگنالهای ورودی باشند. (شرح ایده در [پ-الف-۸] تشریح شده است)

در اینجا نکته ای لازم است ذکر شود که در رابطه با بحث "تخصیص احتمالات سیگنالهای ورودی" می باشد. تخصیص احتمالات سیگنالهای ورودی با سه فرایند زیر انجام می شود :

۱- انتخابِ زیر مجموعه ویژگیهای موثّر
۲- نگاشت صحیح ویژگیهای انتخاب شده به سیگنالهای مربوطه [۱]
۳- پردازش سیگنالها به منظور بدست آوردنِ سیگنالهای خروجی و مشخص شدن سلّولهای مهاجرت یافته

[۱] اینکه کدام ویژگی/زیر مجموعه ویژگی باید به کدام سیگنال نگاشت گردد.

همانگونه که می دانیم برای هر نمونه ترافیک شبکه (آنتی ژن) در دادگان آزمون می بایست سه سیگنال امن ، خطر و PAMP تخصیص گردد. دو متد رایج برای این کار استفاده از انتخاب ویژگی یا دانش متخصّصان می باشد. متد انتخاب ویژگی کاری که انجام می دهد ، جستجو و انتخاب بهینه ترین زیر مجموعه ویژگی و سپس نگاشت آنها به سیگنالهای مرتبط است تا در نهایت طبق استراتژی [پ-الف-۸] سیگنالها محاسبه و تخصیص داده شوند. همچنین دانش متخصصان نیز در یافتن بهترین زیر مجموعه ویژگیها به منظور نگاشت به سه سیگنال می باشد. در [پ-الف-۹] در خصوص نحوه استفاده از دانش متخصصان در تخصیص سیگنالهای ورودی روابطی پیشنهاد شده است.

همانطور که اشاره شد در ابتدای کار پایان نامه هدف از ارائه این ایده ، صرفاً استفاده از یک خوشه بند مناسب، مشخصاً حفره سیاه بود که طبق استراتژی پیشنهادی [پ-الف-۸] بتواند ضمن انجام دسته بندی ترافیک شبکه ، سیگنالهای ورودی را به شیوه ای نوین محاسبه و تخصیص دهد که با دو روش رایج اشاره شده در بالا تفاوت اساسی دارد. ایده پیشنهادی فوق برای تخصیص احتمالات سیگنالهای ورودی ، نیازمند به تحقیقات بیشتر و آزمون و ارزیابی با سایر خوشه بندهای جدید می باشد مثل خوشه بندی هایی که مسئله دسته بندی را به شیوه غیر متمرکز و به صورت توزیع شده حل می کنند و یا مثلاً خوشه بندهای جریانی .[1]

به بیان بهتر به نظر می رسد در صورتی که این ایده و روابط ارائه شده در آن (روابط ۸-۱ ، ۸-۲ و ۸-۳ در [پ-الف-۸]) در خوشه بندهای جدید غیر متمرکزگِرا به کار برده شوند نتایجِ اعمال آن در تخصیص سیگنالها و نهایتاً خروجی مطلوب دسته بندی DCA امیدوار کننده خواهد بود.

۲- مشکل دیگر الگوریتم سلولهای دندریت ، وزنهای استفاده شده در ماتریسِ پردازش سیگنالهاست. این وزن ها درصورتی که توسط کاربر تعیین گردند بسته به تغییر شرایط مسئله ضرورتاً باید انعطاف پذیری نشان داده و تغییر یابند که عدم تغییر آنها عملکرد الگوریتم را تنزّل می دهد. به عبارت دیگر، بحث تخصیص ضرایب وزنی بهینه باید با شرایط مسئله دسته بندی تطبیق داشته باشد. حل این مشکل می تواند با کاربرد سیگنال چهارم [2]که از مبنای زیستی آن کمکی برای سه سیگنال می باشد در فصل دوّم مفصل بحث شده است.[۱۵]

۳- ضعف دیگر ، نیاز الگوریتم به وجود ویژگیِ کلاسِ نمونه های ترافیک شبکه و لزوم از پیش مرتب بودن[3] آنهاست. با توجه به اینکه این مشکل تاکنون حل نشده است ایده ای را بدین منظور ارائه نمودیم که می توان این مسئله را با تعیین فاصله اقلیدسیِ نزدیکترین نمونه یادگیری به آنتی ژن مورد نظر و تخصیص احتمالی برچسبِ نمونه یادگیری به آنتی ژن مذکور حل نمود. البته می توان مجموعه یادگیری را صرفاً یکی از زیر

[1] Stream Data Clustering
[2] Inflamation
[3] Pre – Ordered

مجموعه امضاء ها یا پروفایل ها (به ترتیب آنومالی یا نرمال) در نظر گرفت و دادگان آزمون را بر مبنای تخصیص برچسب کلاسِ نزدیکترین نمونه ها در زیر مجموعه یادگیری به آنتی ژن ها ، مرتّب سازی نمود.

۷-۱-۳- ایده – ارائه یک استراتژی موثّر برای نمونه برداری غیر تصادفی سلّولهای دِندریت

همانگونه که در بخش ۳-۱-۴ بیان شد ، نمونه برداری سلّولهای دِندریت در الگوریتم استاندارد به صورت تصادفی می باشد که این حالت تصادفی از جمله ضعف های الگوریتم اصلی می باشد. این ضعف موجب می گردد که نتایج خوشه بندی DCA در یک آزمایش مشخص، هر بار نتایج دور از انتظاری را ارائه دهد که منجر به بُروز عدم قَطعیّت [1]به مفهوم عدم ثُبات در دسته بندی ، در خروجی الگوریتم نیز می گردد.

البته می دانیم که خروجی حاصل از خوشه بندی همیشه دارای خطا بوده و بستگی به پارامترهای زیادی دارد همچون نرخ های حدود آستانه ، نوع و نحوه توزیع دادگان در فضای ابعاد مسئله (نوع و تعدادِ حملات نسبت به داده های نرمال) ، استفاده از متد کاهش ابعاد در فاز پیش پردازش و غیره. ما با ارائه روابط زیر ، تابع استانداردِ نمونه برداری ، Get_Antigens() ، را از حالت تصادفی خارج نموده و استراتژی شُعاع تحت پوشش را به صورت زیر پیشنهاد نمودیم :

$$R(i) = \left[\frac{\text{Vector(i)}}{\max(\text{Vector})}\right] * \text{mean}\big(\text{dist(all from i)}\big) \qquad (۳-۸)$$

$$R(i) = \left[\frac{\text{Vector(i)}}{\max(\text{Vector})}\right] * \text{median}\big(\text{dist(all from i)}\big) \qquad (۳-۹)$$

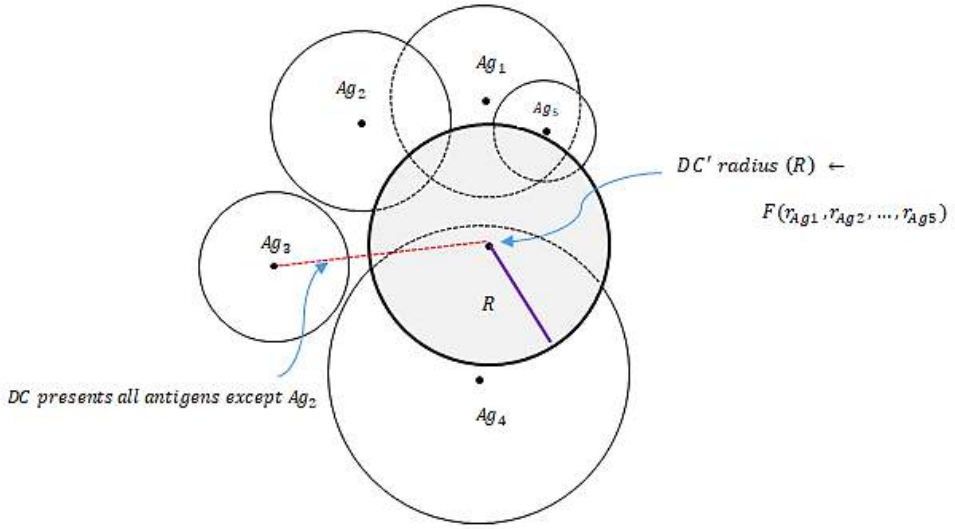
; dist is based on Eucledian distance ; Vector(i) is clones of sample i

این رابطه مقداری را نتیجه می دهد که می توان به عنوان شعاع دایره ای در نظر گرفت که فضایی را به مرکزیّت سلّول دِندریت ایجاد و دور تا دور آنرا محصور می کند. بدین ترتیب ، نمونه برداری هر سلّول از آنتی ژنها بستگی به فضای تحت پوششِ آن سلّول خواهد داشت و وسعت این فضا نیز بستگی به اندازه شُعاع دارد. تعیین شعاعِ فضای تحت پوشش سلّول بستگی به دو عامل دارد : یکی میانه یا میانگین فاصله سلّول از سایر آنتی ژنها و دیگری کَسرِ تعداد تکثیر های آنتی ژنی که سلول در موقعیّت آن قرار گرفته به حداکثر تعداد تکثیرهای کُل آنتی ژنها. همچنین حداکثر مقداری که شُعاع می تواند اختیار کند نیز برابر با میانگین / میانه فاصله سلّول از سایر آنتی ژنها خواهد بود. شکل زیر این مفهوم را بهتر نشان می دهد.

[1] Uncertaintability

در این شکل سلّول دِندریت در موقعیّتِ آنتی ژنی قرار گرفته که می خواهد سیگنالهای آنرا محاسبه (نمونه برداری) نماید . این سلّول فضایی را محصور ساخته است.

همانطور که بیان گردید فضای تحت پوشش به کسرِ تعداد تکثیرهای آن آنتی ژن به کل تکثیرهای کل آنتی ژنها در دادگان و در عین حال به میانه فواصل آن آنتی ژن از سایر آنتی ژنها بستگی دارد. هر چه تعداد تکثیر های آنتی ژن در مقایسه با کل تکثیرها بیشتر باشد کسر مربوطه به سمت ۱ بیشتر میل خواهد نمود که در حالت ایده آل مقدار شُعاع نیز به سمت میانه/ میانگین فواصل میل خواهد کرد. (روابط ۳-۸ و ۳-۹)



شکل ۹ –نمونه برداری آنتی ژنها توسط سلّول دندریت با روابط ۳-۸ و ۳-۹

بدین ترتیب پس از محاسبه شعاع های تمامی نمونه های آنتی ژنیک ، مشاهده می شود که نواحی همپوشانی به منزله فضای نمونه برداری سلّول می باشند. فرض کنید تمام نمونه های آنتی ژنیک موجود در شکل ۹ در یک بافت سلّولی واقع شده باشند. به بیان بهتر دو ویژگیِ پروتکل-سرویس در تمامی آنها یکسان بوده باشد. حال اگر سلّول DC_i که در موقعیّت آنتی ژنِ Ag_i قرار گرفته است دایره ای را با شُعاع مشخصِ R دور تا دور خود ایجاد کند در حقیقت در بافت سلّولی مشخص شده نواحی همپوشان و اشتراکی را با نمونه های دیگر مجاور خود ایجاد نموده و این نواحی همپوشان را می توان به عنوان نمونه برداری برای آن DC_i در نظر گرفت. در اینجا نکته ای لازم است ذکر گردد. سلّولهای دِندریت به صورت پیشفرض به تعداد آنتی ژنها تولید شده و هر سلّول در موقعیّتی متناظر با موقعیّت آنتی ژنی قرار میگیرد که آنرا Present می کند. به عبارت دیگر در شکل بالا DC_i اقدام به Present کردن آنتی ژنی (Ag) نموده که در موقعیّت آن قرار گرفته است. در الگوریتم استاندارد به استنادِ مرجع [۱۵] سلّول به محض آنکه در آن موقعیّت قرار گرفت اقدام به نمونه

برداری از نزدیکترین آنتی ژنهای مستقر در بافت های سلّولی نموده و سیگنالهای دریافت شده از آنها را پردازش می کند تا در نهایت وضعیّتِ مهاجرت DC_i مشخص گردد.

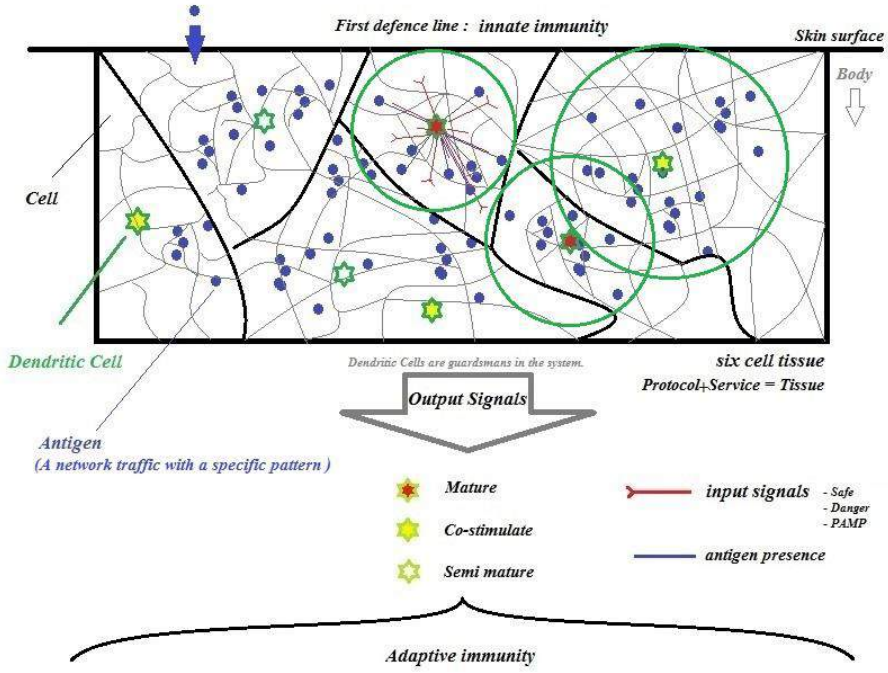
۱-۷-۱-۳- نحوه مُحاسبه بردار تکثیر آنتی ژنها

در ۳-۱-۳-۱ خلاصه ای از نحوه مُحاسبه بُردارِ تکثیرِ آنتی ژنها بیان گردید. فرایند تکثیر آنتی ژنها مستقیماً بستگی به تعداد آنتی ژنهای موجود در هر بافت سلّولی دارد. بطوریکه در الگوریتم ۴ (شکل ۱۲) مشاهده می گردد ابتدا بافت های سلّولی بر مبنای دو ویژگی پروتکل-سرویس ایجاد می شوند. سپس آنتی ژنهای هر بافت بر اساس رابطه زیر می توانند تکثیر گردند تا در نهایت به صورت تصادفی یکنواخت ، بوسیله نزدیکترین سلّولهای دندریت و با اعمال روابط ۳-۸ و ۳-۹ نمونه برداری گردند.

$$\begin{aligned}\mathrm{AgVector(i)} = [\mathrm{ceil(Unifrnd(minimumCloneRange.\,maximumCloneRange))} \\ \times \mathrm{size_{of}tissues(i)}] \; + \; \mathrm{size_{of}tissues(i)}\end{aligned}$$

(۳-۱۰)

در رابطه بالا ، میزان تکثیر آنتی ژنها در بافت سلّولی i - ام به تعداد آنتی ژنهای موجود در آن بافت $(\mathrm{size_{of}tissues}(i))$ وابسته بوده و همیشه ضریبی از این تعداد به آن اضافه می گردند.



شکل ۱۰ – شماتیکِ بافت های سلّولی ، نفوذ آنتی ژنها و نقش سلّولهای دِندریت در بدن

به عنوان مثال اگر اندازه بافت برابر با ۵۰ آنتی ژن بوده و بازه ی حداقل و حداکثرِ میزان تکثیر [1]برای آنتی ژنها نیز توسط کاربر به ترتیب برابر ۵ و ۱۵ تعیین گردند ، بدین ترتیب طبق رابطه بالا ابتدا مقداری در بازه ۵ تا ۱۵ به صورت تصادفی یکنواخت تعیین شده و سقف این مقدار به عنوان ضریب در اندازه آن بافت اعمال می گردد و حاصل آن نیز به اندازه بافت سلّولی اضافه می گردد. پس از تعیین میزان تکثیر آنتی ژنهای هر بافت نوبت به نمونه برداری می رسد که با روابطِ ۳-۸ و ۳-۹ بدست می آید.

شکل بالا شماتیکی فرضی از بافت های سلّولی زیستی را به صورت فرضی نشان می دهد. در این شکل آنتی ژنهایی که به سلّولهای بافت ها نفوذ کرده اند با رنگ آبی مشخص شده اند. این آنتی ژنها از سطح پوست (بالا) وارد بدن شده و از لایه های دفاعی عبور می کنند. سلّولهای دِندریت در نقش پاسدارانی هستند که وظیفه نمونه برداری از آنتی ژنها و پردازش سیگنالها (فلش های قرمز رنگ) را برعهده دارند. به محض تخریب هر سلول یا در صورت سالم بودن آن نیز ، سیگنالی به نزدیک ترین سلّول دندریت مستقر در آن بافت ارسال می گردد. در DCA فرض شده که تعداد سلّولهای دندریت به تعداد آنتی ژنها بوده و هر یک از این سلّولها در موقعیّتهای متناظر با آنتی ژنها قرار میگیرند. به نظر میرسد دلیل این فرض ، امکان شبیه سازی مفهومِ نمونه برداری (Present) می باشد.

۸-۱-۳- الگوریتم پیشنهادی برای خط نخست دفاعی

هدف اصلی ما از همان ابتدای آغاز به کار پژوهش پایان نامه ، شبیه سازی واکنش های ایمنی ذاتی و تطبیق پذیر و آزمون و ارزیابی با دادگان نفوذ در شبیه ساز متلب بوده است. برای اینکار در فصل دوّم مروری به پژوهش های نظری و تجربی در این حوزه انداختیم و با پیش زمینه زیستی آن آشنا شدیم.

از ابتدای این فصل نیز سعی کردیم به نوعی مکانیسمِ DCA را به عنوان محصول تئوری خطر بهبود داده و با آن کار کنیم. مطابق پیش زمینه نظری تحقیق بیان شده در فصل دوّم ، میتوان سیستم ایمنی ذاتی را معادل با یک دسته بند نظارت نشده / نیمه نظارت شده در نظر گرفت. به عبارت دیگر واکنش سلّولهای دِندریت در مقابل تهاجم و نفوذ خارجی به صورت ذاتی و سریع بوده و بدون دانش قبلی و یا اندکی دانش می باشد. این مسئله منجر به این شد که ایده ای به ذهن ما برسد و آن ، بهبودِ مکانیسمِ DCA استاندارد به عنوان یک دسته بند نیمه نظارت شده برای خط نخست دفاعی در روش تشخیص نفوذ پیشنهادی بود. در واقع به نظر میرسد که نقش این سلّولها ، با اندکی یادگیری قبلی و یا بدون آن صرفاً کمکی برای لایه دفاعی پایینتر بدن (مصونیّت اکتسابی) باشد.

[1] [minimumCloneRange . maximumCloneRange]

بدین ترتیب وجود دو خط دفاعی برای ایجاد سیستم تشخیص نفوذ مبتنی بر ایمنی مصنوعی ، ضروری به نظر میرسد. ما این خطوط را شبیه به مکانیسمِ حفاظتی در یک قلعه فرضی تصوّر نمودیم که عملکرد آن کاملاً مبتنی بر خصوصیات اصل دفاع در عمق بوده و به صورت مثالی مبسوط در فصل دوّم بیان گردید.

در نتیجه برای پیاده سازی خط نخست دفاعی از الگوریتم سلولهای دندریت کمک گرفتیم و با ایده پردازی با الهام از مفهوم تئوری خطر و انجام آزمایشهای متعدّد سعی نمودیم آنرا به طریقی بهبود دهیم. شبه کد زیر مربوط به متد پیشنهادی ما برای کشف حدآستانه مهاجرت و تخصیص احتمالات سیگنالهای ورودی می باشد.

Algorithm. 3 − Pseudo code of the Proposed Signal_Categorization method

Inputs: $\text{Antigen}_{\text{Set}}.\ \text{SelfNormal}_{\text{set}}.\ \text{Features}_{\text{Subset}}.\ \text{Weight}_{\text{Coefficients}}.\ \text{Approperated\_Features}$
Outputs: $\text{Migration}_{\text{Threshold}}.\ \text{Probabilities of InputSignal}(\text{Pr}_{\text{ss}}.\ \text{Pr}_{\text{pamp}}.\ \text{Pr}_{\text{ds}})$
$\text{Sorted}_{\text{testSet}}$ Computed by Proposed Alg.[2]

/∗ **Question** :
Which one of these methods you want to select for Calculation of Pre − Categorized Input Signals ?
/∗ **1) Feature_Selection (IG)**
2) By knowledge of experts
3) Using a Clustering method . for e. q Black Hole Alg.

1) $[\text{Probabilities of InputSignals}(\text{Pr}_{\text{SS}}.\ \text{Pr}_{\text{PAMP}}.\ \text{Pr}_{\text{DS}})] = \text{Information\_Gain}\left(\text{Features}_{\text{Subset}}.\ \text{Similar}_{\text{Test}_{\text{Set}}}.\ \{\text{Median or Mean}\}?\right)$

/ The non − nominal features are candidate for signals SS. PAMP. DS
/ The non nominal feature with a highest information gain value is selected for to be Safe. Pamp signal. So other featuers with highest information gain are selected as a subset for generating the Danger signal

2) $[\text{Probabilities of InputSignals}(\text{Pr}_{\text{SS}}.\ \text{Pr}_{\text{PAMP}}.\ \text{Pr}_{\text{DS}})] = \text{knwoledge}_{\text{of}}\text{Experts}(\{\text{Approperated}_{\text{Features}}\}.\ \text{Similar}_{\text{Test}_{\text{Set}}}.\ \{\text{Median or Mean}\}?)$

3) $[\text{Probabilities of InputSignals}(\text{Pr}_{\text{SS}}.\ \text{Pr}_{\text{PAMP}}.\ \text{Pr}_{\text{DS}})] = \text{ByUsingClusterin}(\text{Features}_{\text{Subset}}.\ \text{Similar}_{\text{Test}_{\text{Set}}})$

/ The Migration Threshold by CSM Weight Coefficients. Using formula. [3 − 3 . 3 − 4]. Median or Mean ?

$$\text{mt} = \begin{pmatrix} \left(\{\text{median or mean}\}\left(\text{InputSignals}.\ \text{Pr}_{\text{pamp}} \times \text{Weight}_{\text{Coe}(1)}\right)\right) + \\ \left(\{\text{median or mean}\}\left(\text{InputSignals}.\ \text{Pr}_{\text{ss}} \times \text{Weight}_{\text{Coe}(2)}\right)\right) + \\ \left(\{\text{median or mean}\}\left(\text{InputSignals}.\ \text{Pr}_{\text{ds}} \times \text{Weight}_{\text{Coe}(3)}\right)\right) \end{pmatrix}$$

شکل۱۱-شبه کد مربوط به الگوریتم پیشنهادی برای تخصیص احتمالات سیگنالهای ورودی و تعیین حدآستانه مهاجرت پویا

ما در الگوریتم پیشنهادی-۴ ، ایده های پیشنهادی ۳-۲ تا ۳-۱۰ را به کار گرفتیم تا در نهایت بخش هاشور خورده ی شکل ۵ مربوط به الگوریتم-۱ را بهبود دهیم. این بخشها مربوط به سه قسمت مهم مکانیسمِ DCA می باشند که به طور خلاصه تحت عنوان تابعِ نمونه برداری مبتنی بر مفهوم تئوری خطر (DT) می شناسیم. به عبارت

بهتر ، این بخشها به ترتیب فاز تشخیص شامل نمونه برداری از آنتی ژنها با بهبودِ تابع ()Get_Antigens ، فاز پردازشِ سیگنالهای خروجی با بهبودِ تابعِ()Get_Antigens ، فاز تعیین وضعیّتِ مهاجرتِ سلّول (Context Assessment) را شامل می شوند.

Algorithm. 4 − Pseudo code of Propsoed DT Method (Sampeling)

Inputs : $S = \text{AntigenSet.trainSet.Features}_{\text{Subset}}.\text{Weight}_{\text{Matrix}}.\text{minSize}_{\text{of}}\text{CloneAg.maxSize}_{\text{of}}\text{CloneAg.}$
Probabilities of InputSignal. $\text{Migration}_{\text{Threshold}}$
Output: $E = \text{MigratedCellsSet.UnmigratedCellsSet.Cell}_{\text{context}}.\text{Vector}$

−Create tissues by dividing the Antigen Set based on Protocol-Service attributes to tissues (Alg.6) **/* Create tissues**
− Product Antigen Vector for Sampling by Using the Proposed Strategy of. [3-10] **/* Create Vector**
/∗ Detection_phase
For each selected DC DO
Get the antigen:
−Determine Clone Rates of Antigen as Uniformly Random.
−Compute Radius of Sampling for each Dendritic Cell by using the one of these formulas. [3-8, 3-9]
/* Presenting Antigen by DC
−Presenting Ag by using the one of these formulas. [3-5 to 3-7]
−Get the Probabilities of Input signals **/∗ Signal_Assessment**
−Calculate interim output signals by Formula. [3 − 2]
−Update the cumulative output signals
IF cumulative Csm > $\text{Migration}_{\text{Threshold}}$ then
−Remove the DC population
−Assign the Cell − context to DC
IF cumulative Semi ≤ cumulative Mat then **/∗ Context_Assessment_phase/**
$\text{Cell}_{\text{context}} = 1$
Add DC to $\text{MigratedCellsSet.Mat}_{\text{Cells}}$
Else
$\text{Cell}_{\text{context}} = 0$
Add DC to $\text{MigratedCellsSet.Semi}_{\text{Cells}}$
End If
− All DCs which collected the antigen and have a cell_context out for analysis
− Termination this DC and add a naive DC to the population
Else
− Add DC to $\text{UnmigratedCells}_{\text{Set}}$
Else If
− DC back to population
End For

شکل ۱۲ – سودوکُد مربوط به الگوریتمِ پیشنهادی برای تابع نمونه برداری

Algorithm . 5 − Proposed Algorithm for Initial Defence Line

Inputs : $\text{AntigenSet.trainSet.Features}_{\text{Subset}}.\text{Weight}_{\text{Matrix}}.\text{minSize}_{\text{of}}\text{CloneAg.maxSize}_{\text{of}}\text{CloneAg}$
Outputs : MigratedCellsSet. MCAV (Mature Context Antigen Value)

/∗ Preprocessing

Both of Probabilities of InputSignal $\left(\text{Pr}_{\text{ss}}.\ \text{Pr}_{\text{pamp}}.\ \text{Pr}_{\text{ds}}\right)$ and MT are computed by Proposed Signal Categorization Alg. [3]

While UnmigratedCellsSet is not empty

/∗ Detection & $\textbf{Context}_{\textbf{Assessment}}$**and** $\textbf{Signal}_{\textbf{Assessment}}$ **Phases**

Presenting Antigens by DCs and migrating them using the Proposed Function of DT (Alg. [4])

/∗ Calculating the Migration Threshold (Variable step by step). Median or Mean ?

$\text{Migration}_{\text{Threshold}} = \{\text{median or mean}\}(\text{CSM Values of all DCs})$

End while

/∗ **Classification Phase. MCAV Calculation**
/∗ **Using new approach for MCAV Calculation Ref** . [**37**]
Foreach MigratedCellsSet. $\mathrm{Semi}_{\mathrm{Cells}}$

$$\mathrm{Semi}_{\mathrm{DCs}}(\mathrm{Ag_i}) = \frac{[\mathrm{Semi}_{\mathrm{Cells}}(\mathrm{Ag_i}) - \min(\mathrm{Semi}_{\mathrm{Cells}})]}{[\max(\mathrm{Semi}_{\mathrm{Cells}}) - \min(\mathrm{Semi}_{\mathrm{Cells}})]}$$

/∗ **means that how much DendriticCells each Antigen has been able totransfer to SemiMatured state** ?
End foreach
Foreach MigratedCellsSet. $\mathrm{Mat}_{\mathrm{Cells}}$

$$\mathrm{Mat}_{\mathrm{DCs}}(\mathrm{Ag_i}) = \frac{[\mathrm{Mat}_{\mathrm{Cells}}\ (\mathrm{Ag_i}) - \min(\mathrm{Mat}_{\mathrm{Cells}}\ )]}{[\max(\mathrm{Mat}_{\mathrm{Cells}}\ ) - \min(\mathrm{Mat}_{\mathrm{Cells}}\ )]}$$

/∗ **means that how much DendriticCells each Antigen has been able to transfer to Matured state** ?
End forech
For each member of MigratedCellsSet
IF context(i) == 1 then

$$\mathrm{MCAV(i)} = \mathrm{mean}\left(\left[\frac{\mathrm{Mat}_{\mathrm{DCs}}(\mathrm{i})}{\mathrm{Mat}_{\mathrm{DCs}}(\mathrm{i}) + \mathrm{Semi}_{\mathrm{DCs}}(\mathrm{i})}\right].\left[\frac{\mathrm{migratedCellsSet(i).\,Present}}{\mathrm{migratedCellsSet(i).\,Clone}}\right]\right)$$

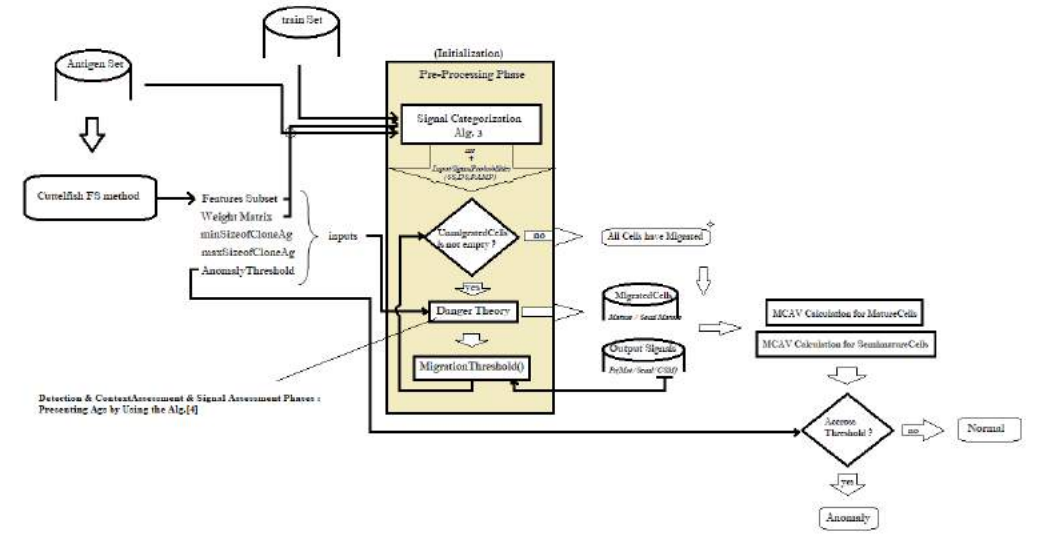
Else

$$\mathrm{MCAV(i)} = \mathrm{mean}\left(\left[\frac{\mathrm{Semi}_{\mathrm{DCs}}(\mathrm{i})}{\mathrm{Mat}_{\mathrm{DCs}}(\mathrm{i}) + \mathrm{Semi}_{\mathrm{DCs}}(\mathrm{i})}\right].\left[\frac{(\mathrm{migratedCellsSet(i).\,Clone} - \mathrm{migratedCellsSet(i).\,Present})}{\mathrm{migratedCellsSet(i).\,Clone}}\right]\right)$$

Endif
End foreach

شکل ۱۳ – شبه کد مربوط به الگوریتمِ پیشنهادی برای خط نخست دفاعی

به منظور فهم بهتر ساختار الگوریتم پیشنهادی ، در شکل زیر اینفوگرافیک آنرا ترسیم نموده ایم.



The Proposed DCA for Initial Defense Line

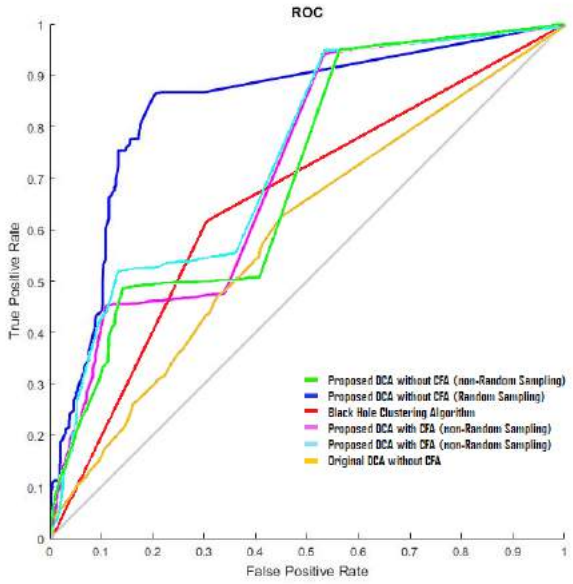
شکل ۱۴ – اینفوگرافیکِ مکانیسمِ الگوریتمِ پیشنهادی برای خط نخست دفاعی

۱-۸-۱-۳- آزمایش ارزیابی عملکرد خط نخست دفاعی

ما DCA پیشنهادی را در مقایسه با الگوریتمِ سیاه چاله ارزیابی مقایسه ای نمودیم. برای محاسبه و تخصیص سیگنالهای ورودی از استراتژی انتخاب ویژگی بهره اطلاعات یا IG استفاده شد. نتایج در منحنیِ ROC و جدول زیر نشان می دهند که الگوریتم پیشنهادی با و بدون اعمال انتخاب ویژگیِ دَه – پا در مقایسه با الگوریتم حفره سیاه (منحنی قرمز) نرخ تشخیص بسیار بالایی را تحت شرایط بدون یادگیری قبلی ارائه می دهند. این در حالیست که نرخ تشخیص الگوریتم حفره سیاه (65.7 درصد) می باشد.

از طرفی مشاهده می شود که در منحنی آبی رنگ که استراتژی خاصی جهت نمونه برداری آنتی ژنها استفاده نشده و صرفاً تصادفی بوده است، نرخ تشخیص اندکی پایینتر (86.2 درصد) از سه رویکردی است که در آنها از استراتژی رابطه ۳-۹ استفاده شده است. در این نمودار ، دلیل اینکه منحنی های مربوط به الگوریتم پیشنهادی در مقایسه با منحنی های دیگر بطور پیوسته نوسان زیادی داشته و صاف نمی باشند مربوط به خروجیِ برچسب پیوسته دادگان آزمون است که در بازه [0,1] می باشد. اما الگوریتم حفره سیاه ، مسئله را به صورت گسسته حل نموده و برچسب می زند. همچنین مطابق نتایج جدول ۹ نیز مشاهده می شود که میزان تشخیص حملات ناشناخته با الگوریتم پیشنهادی پس از افزایش دَه برابری حجم دادگان آزمون ، از نظر کاهش

نرخ خطای منفی اشتباه بهبود چشمگیری داشته است. (۹۸,۷۶) البته لازم به یادآوری است که یک نقطه ضعف متدهای ایمنی مصنوعی خطای بالای مثبت اشتباه آنها می باشد. نتیجه ای که در خصوص خطای مثبت مربوط به تمام خروجی های الگوریتم پیشنهادی در تمامی منحنی ها مشاهده می شود.



نمودار ۴ - منحنی ROC الگوریتم پیشنهادی برای خط دفاعی اوّل

جدول ۸ - نتایج دسته بندی الگوریتم پیشنهادی خط نخست دفاعی

Method (Color)	***Dr***	***FPR***	***FNR***	***TP***	***FP***	***FN***	***Accuracy***	***Error***	***F – measure***	***Precision Rate***	***Attack corrected prediction rate***	***normal corrected prediction rate***	***Attack incorrected prediction rate***	***normal incorrected prediction rate***
Green	95.0	56.4	5.0	65.3	17.6	3.4	79.0	21.0	85.74	78.77	78.7	80.0	21.3	20.0
Blue	86.2	20.3	13.8	59.2	6.4	9.6	84.2	15.8	88.17	90.24	90.3	72.5	9.7	27.6
Red	65.7	37.4	34.3	45.14	25.70	10.72	64.7	35.3	64.70	63.72	45.4	79.4	54.6	20.6
Violet	94.5	57.0	5.5	64.9	17.8	3.8	78.4	21.6	85.74	78.48	78.5	78.0	21.5	22.0
light Blue	95.5	55.2	4.5	65.6	17.3	3.1	79.7	20.3	84.54	79.13	79.2	82.0	20.8	18.0
Orange	62.6	44.5	37.4	43.0	13.9	25.7	60.4	39.6	68.47	75.57	75.5	40.3	24.5	59.7

ما این بار این میزان را برای دادگانی با اندازه دَه برابر به حدود ۱۵ درصد یعنی ۳۰۰۰ رکورد حمله ناشناخته افزایش داده و مجدداً با الگوریتم پیشنهادی اقدام به آزمون آن نمودیم. در دادگانِ آزمون قبلی که اندازه آن ۲۱۱۰ رکورد بود ، ۴,۰۷۵ درصد از این دادگان یعنی ۸۶ رکورد را حملات ناشناخته تشکیل داده بود که از

این میزان ، ۱۸ حمله ناشناخته جزو خطای منفی کاذب بودند. به این ترتیب میزان شناسایی حمله ناشناخته در منحنی (Violet) برابر با ۷۹,۰۷ درصد محاسبه گردید که نرخ چندان مطلوبی به نظر نمیرسد. نتایج جدول زیر عملکرد دسته بندی الگوریتم پیشنهادی خط نخست را از حیث میزان تشخیص حملات ناشناخته در دادگان جدید (افزایش یافته) نشان می دهد.

جدول ۹ – نتایج دسته بندی الگوریتم پیشنهادی با افزایش اندازه دادگان آزمون UNSW – NB15 به میزان ده برابر

Rumtime	***Dr***	***FP***	***FN***	***number of False Negative Records***	***Accuracy***	***Error***	***Total Rate of Unknown Attacks (TP)***	***non – Detected Unknown Attacks (FN)***
41.007	94.7	16.6	5.3	77	79.8	20.2	98.76	37 record

مطابق اطلاعات جدول فوق ، از مجموع ۷۷ حمله ای که به اشتباه جزو پروفایلهای نرمال سیستم شناخته شده اند (خطای منفی اشتباه) ، ۳۷ نمونه جزو حملات ناشناخته می باشند. معمولاً با افزایش اندازه مجموعه دادگان آزمون ، ضمن افزایش پیچیدگی زمانی ، میزان خطای ناشی از تشخیص حملات ناشناخته بالا می رود امّا در این وضعیت ما شاهد افزایش نرخ تشخیص این نوع حملات از ۷۹,۰۷ به ۹۸,۷۶ درصد بودیم.

۲-۸-۱-۳- تفسیر نتایج بدست آمده

لازم است توضیحی راجع به نتایج بدست آمده از آزمایش الگوریتم سلولهای دندریت پیشنهادی ارائه شود. مطابق نمودار ، آزمون ها شش بار با و بدون اعمالِ کاهشِ ویژگی دَه-پا انجام شدند. منحنی نارنجی به DCA استاندارد پیشنهاد شده در مقاله [۱۵] مربوط می باشد. ما در این آزمایش ، DCA پیشنهادی را بر مبنای آنکه که نمونه برداری آنتی ژنها بوسیله سلّولهای دندریت بصورت تصادفی انجام گردد یا بر اساس استراتژی پیشنهادی در الگوریتم ۵ و همچنین بخش ۳-۱-۷ ، صورت گیرد آزمون کردیم. منحنی آبی پررنگ مربوط به وضعیتی است که نمونه برداری در الگوریتم پیشنهادی به صورت تصادفی صورت گرفته است. همچنین منحنی های سبز ، بنفش و آبی کم رنگ نیز به ترتیب سه وضعیتی هستند که در همه آنها استراتژی مربوط به نمونه برداری غیر تصادفی در الگوریتم پیشنهادی به کار رفته و آزمونها بر اساس آن انجام شده است. در نتیجه مطابق اطلاعات نمودار و جدول فوق مشاهده می شود که بالاترین نرخ تشخیص و در عین حال کمترین میزان خطای منفی اشتباه مربوط به الگوریتم پیشنهادی مجهّز به استراتژی نمونه برداری و اعمال متد کاهش ویژگی ده پا در آن می باشد. همچنین منحنی آبی پر رنگ نیز که مربوط به آزمون الگوریتم پیشنهادی

با نمونه برداری تصادفی آنتی ژنهاست، نیز نرخ خطای کمتر و دقّت دسته بندی بالاتری را نشان می دهد. از نتایج این چنین بر می آید که باید بیشتر بر روی استراتژی نمونه برداری پیشنهادی و همچنین پارامترهای ورودی همچون ماتریس ضرایب وزنی و حدود آستانه mt بیشتر کار شود. ضمناً نکته ای که قابل تامل است نتیجه اعمال الگوریتم کاهش ویژگی ده پا و تاثیر آن بر خوشه بندی مشخص است. منحنی سبز و آبی پر رنگ هر دو به ترتیب نتایج آزمون الگوریتم پیشنهادی را بدون اعمال کاهش ویژگی نشان می دهند. پس از اعمال الگوریتم ده پا ، منحنی آبی کم رنگ بهبود را نشان می دهد. دلیل کاهش عملکرد دسته بندی در منحنی بنفش پس از اعمال الگوریتم ده پا نسبت به منحنی سبز ، تغییر حد آستانه mt و حدود کمینه و بیشینه تکثیر آنتی ژن است که مقادیر پیشفرض آن در آزمایشات قبل به ترتیب 5 و 15 می باشد. ما در آزمایش منحنی بنفش ، مقدار بیشینه تکثیر آنتی ژن را افزایش داده و تغییراتی در ماتریس ضرایب وزنی داده بودیم.

از طرفی تفسیر مهمی که از نتیجه منحنی فوق بدست می آید ، عملکرد دسته بندی بسیار بهتر الگوریتم پیشنهادی نسبت به BHA می باشد . همانطور که در بخشهای ۳-۱-۴ تا ۳-۱-۶ نیز توضیح دادیم علّت آنرا می توان در نحوه حل مسئله به صورت غیر متمرکز و توزیع شدگی ذاتی الگوریتم سلولهای دندریت بهبود یافته دانست. نتایج این بخش را می توان در ترکیب Hybridizasion با یک متد ایمنی مصنوعی نظارت شده مانند انتخاب منفی به نحو چشمگیری بهبود بخشید. زیرا از جنبه زیستی نیز عملکرد سلولهای دندریت در شناسایی رفتار ناهنجار عوامل آنتی ژنیک نفوذی به بدن به صورت ذاتی و سریع بوده و بدون یادگیری قبلی می باشد. از این نظر آنها با بالغ/ نیمه بالغ شدن شان کمک زیادی در فعال سازی سیستم ایمنی تطبیق پذیر (Adaptive immune response) و تولید و تکثیر آنتی بادیهای موثر ایفا می نمایند.

در نتیجه می توان این دو الگوریتم، سلولهای دندریت و انتخاب منفی با دو رویکرد" ایمنی ذاتی و تطبیق پذیر" را مکمّل یکدیگر دانست. از این رو به نظر میرسد این دو متد در کنار یکدیگر نوعی حیات مصنوعی را ایجاد می نمایند که هدف آن کنترل امنیت از طریق برقراری و تضمین نسبی ایمنی سیستم می باشد.

۳-۸-۱-۳- ارزیابی با معیار استعداد تشخیص نفوذ CID

طبق مواردی که در تحلیل روابط در فصل اوّل مطرح گردید ، معیاری به نام " استعداد تشخیص " وجود دارد که اوّلین بار در [۲۷] پیشنهاد شد. منحنی های نمودار ۴ همدیگر را در نقاطی قطع می کنند در نتیجه پاسخ به این سوال که اگر یک IDS از TPR بالاتر و FPR پایینتری برخوردار باشد در مقایسه با IDS دیگری که هر دو مقدار TPR و FPR آن به یک نسبت کاهش یا افزایش یافته باشند ، چه عملکردی دارد ؟

پاسخ به این سوال با مثالی در فصل اوّل در بخش مربوطه بحث گردید و در آنجا گفته شد که در چنین مواقعی که هر دو معیار تشخیص و خطا افزایش یا هر دو کاهش داشته باشند و منحنی ها همدیگر را قطع کرده باشند،

نمی توان به درستی تعیین کرد که کدامیک عملکرد دسته بندی بهتری دارد. بنابراین ما نتیجه بدست آمده در این آزمایش را با معیار استعداد تشخیص نفوذ آزمون نموده و از رابطه ۳-۹ استفاده نمودیم. مقدار پارامتر مبنا ، B یا (احتمال تشخیص نفوذ) به طور پیشفرض 10^{-5} تعیین گردید.

جدول ۱۰ - مقایسه استعداد متدهای تشخیص نفوذ

Method (by Color)	**Capability of Intrusion Detection**
Green	3.5753
Blue	10.4493
Red	2.3315
Violet	3.5311
light Blue	3.6443
Orange	1.1769

متدهای "آبی "و "آبی کم رنگ " به ترتیب با استعدادِ 3.6443 و 10.4493 با استعدادترین متد ها هستند. همچنین همانطور که مشاهده می شود الگوریتم حفره سیاه با رنگ قرمز ، نمره استعداد آن برابر 2.3315 بسیار پایین بوده و نسبت به DCA استاندارد با رنگ نارنجی ، تقریباً در یک سطح قرار گرفته اند.

همینطور رقابت میان متدهای پیشنهادی با رنگ های "بنفش" و "آبی کم رنگ" و "سبز" و تاثیر کاربرد انتخاب ویژگی دَه- پا در منحنی light Blue در مقایسه با منحنی سبز (الگوریتم پیشنهادی بدون انتخاب ویژگی دَه- پا) را مشاهده می کنید. در واقع اعمال متد انتخاب ویژگی در شرایط یکسان باعث افزایش نرخ استعداد بالقوّه تشخیص از 3.5753 به 3.6443 شده است.

ارائه روش پیشنهادی
بخش دوّم
انتخابِ منفی

۳-۲-۱- تحلیل تجِربی روش هیبریدیِ پیشنهادی

در این بخش به تست خُطوط دفاعی در دادگانِ تست مربوطه می پردازیم. نمونه های ترافیک شبکه ی تست در طول روند تست ها نوعی حیات مصنوعی را تجربه خواهند کرد. این حیات مصنوعی نسخه شبیه سازی شده حیاتی است که آنتی ژنها پس از نفوذ به بافت های سلولی بدن در سیستم ایمنی زیستی تجربه نموده و بسته به میزان آسیب پذیر بودن/خطر آنها در از بین بردن سلّولها ، سرنوشت متفاوتی خواهند داشت. سیستم در طول این حیات به سبب فرایندهای آزمون و خطا با عوامل خودی (تعدیل فضاها با تولید و تکثیر تشخیص دهنده های بالغ به منظور پوشش حفره ها و تعیین مرز بندی خودی-غیرخودی با اجرای فازهای یادگیری الگوریتم انتخاب منفی حقیقی مبنا) تجارب بسیاری را کسب نموده و واکنشهای ایمنیِ (ذاتی- تطبیق پذیر) متناسب را از خود نشان می دهد. فرایندی که در نتیجه ی تولید و تکثیر آنتی بادیهای بالغ و مرگ سیستماتیک آنتی بادیهای نابالغ به وقوع می پیوندد. در این پژوهش ما این فرایند را اصطلاحاً " مانُور[۱] " نام گذاری نموده ایم. به عقیده ما فاز های یادگیری اوّلیه و ثانویه مُشابه مانورهای واقعی بوده و مصونیّت سیستم را بالا می برند که این موضوع در دراز مدّت منجر به بلوغ و تضمین نسبیِ ایمنی سیستم در برابر نفوذ می

[۱] کاملاً مشابه مانُورهای نظامی می باشد که درآن نیروهای نظامی با عوامل خودی در نقش دشمن فرضی آزمون و خطا نموده و کسب تجربه می نمایند تا در نبرد واقعی و با تکیه بر تجارب کسب شده بتوانند کارائی لازم را از خود نشان دهند.

گردد. منظور از ایمنی سیستم یا اصطلاحاً "خود-ایمنی" داشتن نوعی پتانسیل بالقوّه در شناسایی نفوذ و عوامل ناشناخته می باشد.

در نتیجه آندسته از تش.د. های بالغی که تبدیل به حافظه شده اند خود ایمن هستند. در واقع سیستم با کسب تجربه حاصل از فازهای یادگیری اولیّه و ثانویه خود و در نتیجه ی آن تعدیل، حافظه سازی و تولید و تکثیر تش.د. های بالغ و مرگ برنامه ریزی شده ی موارد نابالغ می تواند خود را از تهدید تشخیص اشتباه مصون نگه دارد. این تهدید اساسی برای سیستم است زیرا سیستم هیبریدی پیشنهادی به مرور با سیکلهای اجرای متوالی رشد می کند و در صورتی که میزان خطای اشتباه در شناسایی حملات در سیکلهای نخست اجرا بالا باشد منجر به حافظه سازی ، تولید و تکثیرِ اشتباه تش.د. ها شده و سیستم تشخیص ، در برابر نفوذ بسیار دیر مَصون خواهد شد. از دیدگاه ایمنی شناسی ، مصونیّت BIS در برابر آنتی ژن ها یک پروسه دراز مدّت بوده و در صورتی که سیستم به میزان کافی لَنفوسیتهای B حافظه نداشته باشد زمان زیادی طول می کشد تا BIS روندی را در جهت افزایش حافظه طی نماید تا به مصونیّت محافظتی [1]برسد. (شکل ۸ از ۲-۱-۷-۱)

پس از اینکه سیستم به مصونیّت رسید به دلیل وجود تش.د. های بغ.م.حظ ، سیستم توانایی آنرا دارد که در زمان بسیار کمتری اقدام به شناسایی نفوذ نماید. نیاز به زمان طولانی برای افزایش حافظه از جمله نقاط ضعف این سیستم هاست. مطابق آزمایشات به انجام رسیده در این بخش سیستم هیبریدی پیشنهادی قادراست با ایده فازهای یادگیری اوّلیه و ثانویه و تعدیل پیوسته در فضاهای خودی-غیر خودی ، زمان دستیابی به مصونیّت را کاهش دهد.

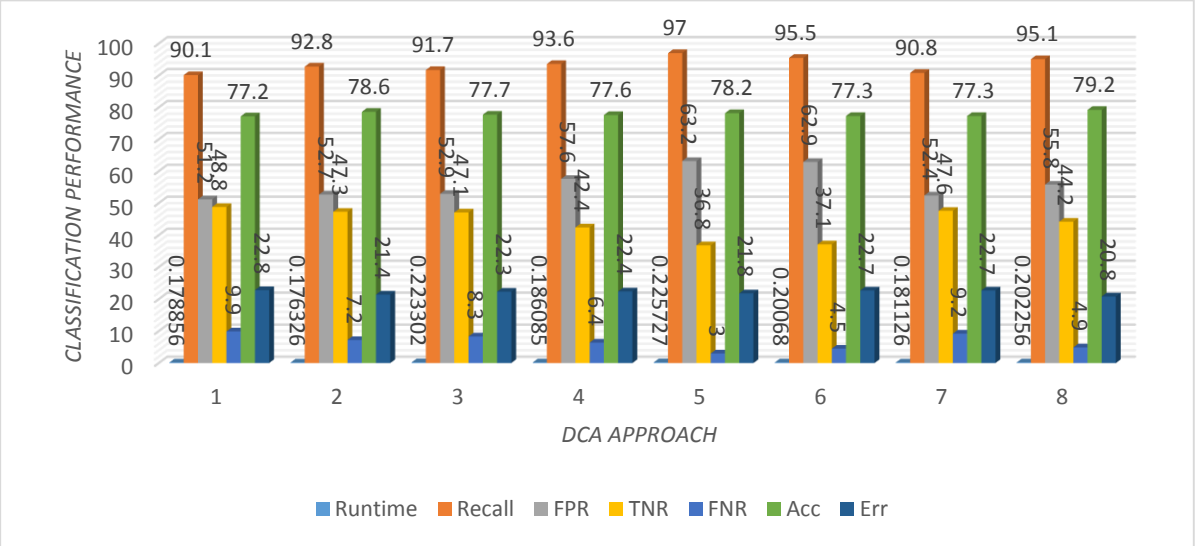
۲-۲-۳- شرح آزمایش

ما از دادگان نفوذ UNSW – NB15 چهار بار به صورت تصادفی یکنواخت نمونه برداری نمودیم و در مجموع چهار زیر مجموعه به عنوان جریان ترافیک آزمون استخراج گردید. (چهار زیر مجموعه در اندازه ۲۱۱۰ رکورد ترافیک شبکه) تمام آزمون های نهایی سیستم هیبریدی پیشنهادی با این چهار دادگان آزمون انجام شده اند.

ابتدا الگوریتم خط نخست با هشت رویکرد متفاوت در پارامترهای ورودی و هر رویکرد بیست بار آزمایش گردید و با بررسی این آزمونها به بهینه ترین پیکربندی پارامترهای ورودی دست یافتیم که شامل حدود آستانه مهاجرت (MT) ، بازه تکثیر سلّولها [2]، معیار مد نظر در تعیین شعاع نمونه برداری سلّولهای دندریت [3]و انتخاب زیر مجموعه ویژگیهای مناسب [4]می باشند. (جدول ۲۱ در [پ-ب-۴])

[1] Protective immunity
[2] min – max of ClonePresentedAg
[3] Radius of Presenting DCs
[4] Optimum attribute Subset

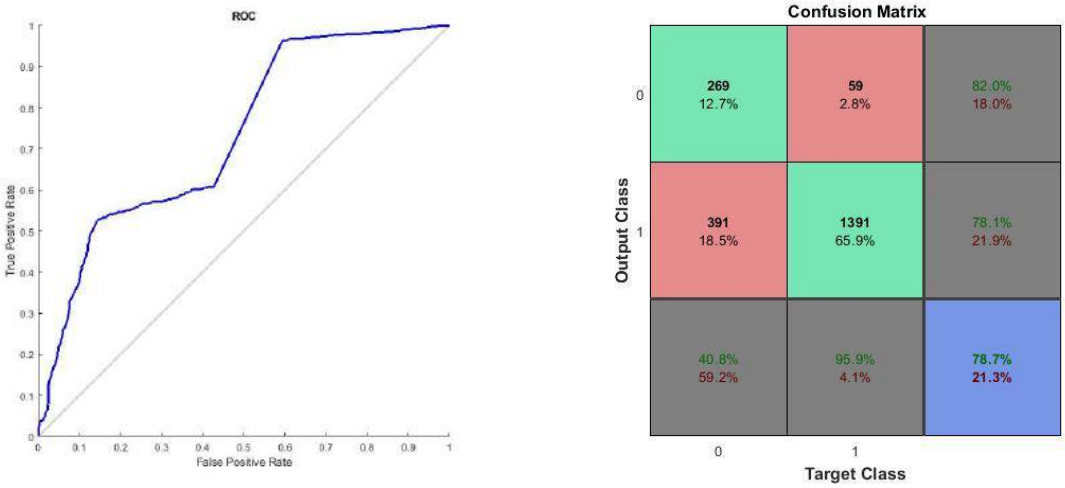
در خط نخست دفاعی به عنوان عامل پیشگیرانه از نفوذ همواره پیکربندی مناسب می تواند نقش موثری در کاهش خطای منفی کاذب و افزایش نرخ تشخیص نفوذ در آزمون و نمونه برداری اوّلیه از ترافیک شبکه داشته باشد. لازم به یاد آوری است که سیستم دفاعی DCA مشابه سیستم دفاعی سربازان یک قلعه با الهام از مص.ذت. سیستم ایمنی بدن انسان ، همواره نرخ مثبت کاذب اشتباه آن بالاست. به دلیل آنکه در سیستم ایمنی بیولوژیک نیز سلّولهای دندریت مشابه سربازان قلعه درصد بالایی از عوامل مخرّب را شناسایی می نمایند امّا ممکن است آنتی ژنهای خودی را به اشتباه به جای آنتی ژنهای غیر خودی درگیر کنند و آنها را تحویل لنفوسیتها دهند. به عبارت بهتر ، سربازان قلعه که مثال مناسبی در این خصوص می باشند همواره ممکن است تعداد زیادی از عوامل خودی را به اشتباه به جای نفوذی قرنطینه نموده و آنها را جهت بررسی بیشتر به قسمت امنیّتی پشتیبانی قلعه تحویل دهند. در نتیجه بالا بودن نرخ مثبت اشتباه در خط نخست طبیعی بوده و سیستم ایمنی این کمبود را با استفاده از الگوریتم انتخاب منفی بهبود یافته در خط دوّم دفاعی جبران خواهد نمود.



نمودار ۵ - ارزیابی عملکرد دسته بندی هشت رویکرد DCA پیشنهادی

در بین رویکردهایی که برای حدآستانه مهاجرت از معیار میانه استفاده نموده اند رویکرد دوّم فیتنس بالاتری دارد. در مجموع این پنج رویکرد (۱ تا ۴ و ۷) فیتنس به مراتب بهتری نسبت به سه رویکرد ۵ و ۶ و ۸ دارند. در پنج رویکرد اشاره شده از سه معیار مختلف میانه ، میانگین و بیشینه برای شعاع نمونه برداری استفاده شده است. بدین ترتیب و با تحلیل دقیق نتایج این آزمایش می توان استنباط نمود که معیار موثّر برای حد آستانه مهاجرت ، میانه می باشد.

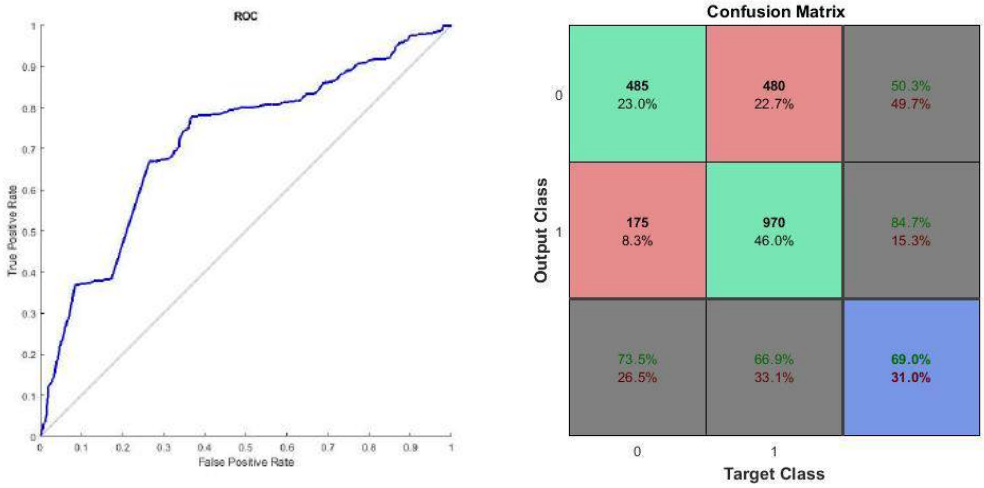
ما برای تعیین بهینه ترین معیار برای نمونه برداری از ایده ی میانگین گیری استفاده نمودیم بدین صورت که ابتدا میزان نمونه برداری آنتی ژنها را با هر سه معیار فوق مشخص نموده و سپس میانگین هر سه را میگیریم تا نتیجه نهایی حاصل گردد. در نهایت بهترین پیکربندی برای DCA پیشنهادی مطابق ردیف ۹ جدول ۲۱ بدست می آید.



نمودار ۶ - نتیجه آزمون خط دفاعی اوّل (DCA پیشنهادی) در دادگان آزمون اوّل
(تخصیص و تولید سیگنالها با متد بهره اطلاعات)

مطابق نتایج بدست آمده از نمودار و جدول ۲۱ در هر رویکرد سه معیار مختلف برای حدآستانه مهاجرت سلّولها و شعاع نمونه برداری آنتی ژنها اعمال شده است. در تمام آزمایشات فوق از متد بهره اطلاعات (IG) برای تولید سیگنالها استفاده شده است. به عنوان نمونه ما آزمایشی دیگر را در همین دادگان آزمون نخست بر اساس رویکرد ۹ جدول ۲۱ در [پ-ب-۴] به صورت زیر انجام دادیم.

نتایج نشان می دهند که نرخ تشخیص پایین بوده و در حدود ۶۶,۹ درصد و نرخ خطای منفی اشتباه آن بالای ۳۳ درصد میباشد. بنابراین استفاده از تجربیات دانش متخصصان در تخصیص بهترین زیر مجموعه ویژگی (ها) به سیگنالهای ورودی متناسب نمی تواند نرخ تشخیص چندان بالایی داشته باشد زیرا همچنان که در جدول فوق ملاحظه می گردد میانگین نرخ تشخیص تمامی هشت رویکرد فوق در بدترین حالت بالای ۹۰ درصد و میانگین نرخ منفی کاذب نیز به ده درصد نمی رسد. در میان نرخ های کاذب ، منفی کاذب از اهمیّتِ به مراتب بالاتری برخوردار است.



نمودار ۷- نتیجه آزمون خط دفاعی اوّل (DCA پیشنهادی) در دادگان آزمون اوّل
(تخصیص و تولید سیگنالها با استفاده از روش تجربه متخصصان)

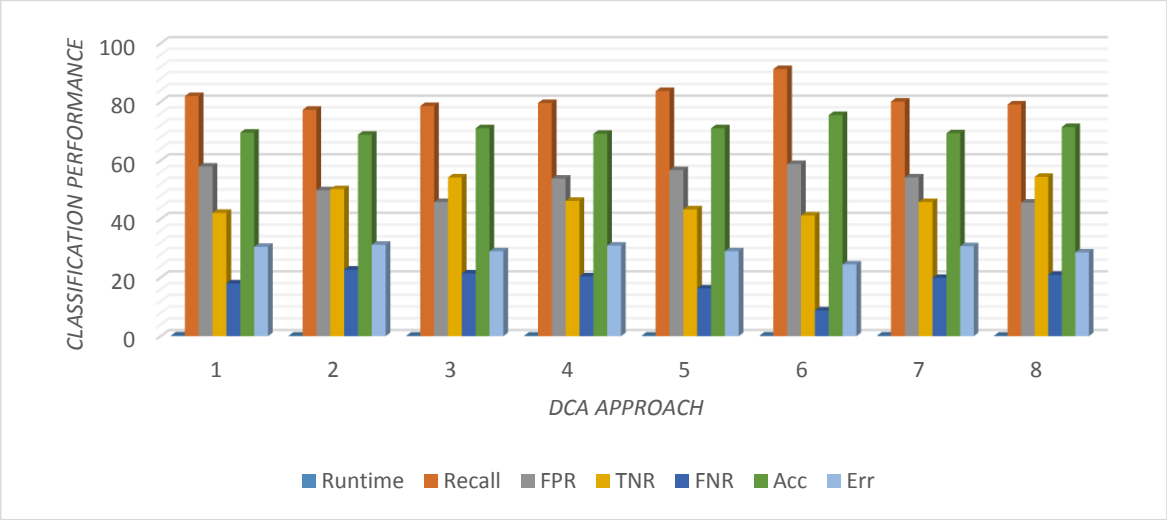
ما در مجموعه آزمایشات نهایی بخش بعد از متد بهره اطلاعات استفاده نموده ایم زیرا استفاده از روش ارزیابی متخصصان در تحلیل ترافیک شبکه و یافتن بهترین زیر مجموعه ویژگیها و تخصیص آنها به سیگنالهای متناسب کار بس دشواری است و همچنان که در نمودار فوق مشاهده می گردد با خطای بالایی نیز همراه است.

ما در آزمایشات نهایی در بخش بعدی این ادّعا را به اثبات رسانده و ناکافی بودن دانش متخصّصان را در تخصیص سیگنالهای ورودی بیشتر توضیح داده ایم. علاوه بر این طبق آزمایشات انجام شده در [پ-الف-۹] در این روش ویژگیهای ۱۵ ، ۱۶ ، ۱۷ ، ۱۸ ، ۲۳ و ۲۴ از دادگان مربوطه به ترتیب به سه سیگنالِ امن، خطر و PAMP نگاشت شده اند.

از طرفی چون هدف اصلی در هر IDS شناسایی صحیح عوامل نفوذ بوده و شناسایی ترافیک نرمال در واقع یک هدف جانبی و غیر مستقیم می باشد بنابراین DR از اهمیت به مراتب بالاتری نسبت به TNR برخوردار می باشد. مطابق تحلیل نظری و تجربی که در ۳-۲-۵ ارائه شده مشاهده می شود که اگر از تجربیات متخصصان به عنوان استراتژی تخصیص سیگنالهای ورودی استفاده گردد این دو نرخ حد یکسان و نسبتاً پایینی خواهند داشت و نرخ های پایین منجر به رخداد وضعیتهای عدم قطعیت زیاد و طولانی تر شدن سیکلهای اجرای سیستم هیبریدی پیشنهادی خواهند شد. زیرا این دو نرخ برای الگوریتم انتخاب منفی استفاده شده در خط دوّم همواره بالای ۹۰ درصد می باشند. بدین جهت از نظر تئوری نیز خوشه بندِ به کار رفته در خط نخست (BHA) حداقل باید یکی از این دو نرخ آن (مخصوصاً DR) بالاتر باشد تا احتمال وجود وضعیتهای عدم قطعیت در سیستم نهایی کاهش یابد و بیشتر نمونه ها Sure گردند در صورتیکه این چنین نبود.

۱-۲-۲-۳- بررسی تاثیرِ کاربردِ الگوریتمِ انتخاب ویژگی دَه-پا بر کیفیت دسته بندی خط نخست دفاعی

استفاده از متدهای انتخاب ویژگی دقّت دسته بندی را در اغلب متدها بالا میبرد. در این بخش عملکرد الگوریتم انتخاب ویژگی دَه – پا را در ترکیب با خط نخست دفاعی بررسی نمودیم. بدین منظور آزمایشی انجام شد که نتایج آنرا در جدول ۲۲ در [پ-ب-۴] می توان بدقّت مشاهده نمود. همان هشت رویکرد با پیکربندی مشابه جدول ۲۱ پیوست به کار رفت. نتایج قابل توجّه است.



نمودار ۸ - ارزیابی عملکرد دسته بندی هشت رویکرد DCA پیشنهادی (با اعمال الگوریتم انتخاب ویژگی ده-پا)

ارزیابی نتایج بدست آمده حاکی از آنست که تاثیر کاربرد الگوریتم انتخاب ویژگی ده-پا عملکرد دسته بندی را کاهش می دهد و صرفاً تعداد تکرارهای اجرا را کاهش می دهد که در سرعت اجرا تاثیر گذار است. با مشاهده نتایج بدست آمده مشاهده می گردد که نرخ فیتنس نهایی الگوریتم خط نخست با اعمال انتخاب ویژگی دَه-پا درتمام هشت آزمایش از ۷۴,۸۴۰۸ برای آزمایش اصلی (ردیف نهم جدول) که در آن از انتخاب ویژگی دَه-پا استفاده نشده است کمتر است. در نتیجه مطابق آزمایشات انجام شده به نظر میرسد که علّتی داشته باشد که چرا کاهش ابعاد مسئله نرخ های دسته بندی و دقّت را در مورد متدهای ایمنی مصنوعی همچون سایر متدهای رایجِ دسته بند یادگیری ماشین ارتقا نمی دهد.

- **علّت چیست ؟**

به دلیل آنکه متدهای ایمنی مصنوعی حقیقی مبنا خصوصاً الگوریتم انتخاب منفی و الگوریتم سلولهای دندریت از معیار فاصله به منظور شناسایی عوامل نفوذ استفاده می کنند بنابراین کاهش ابعاد مسئله بر کاهش دقّت در تعیین فواصل مستقیماً تاثیر منفی می گذارد. زیرا اگر فرمول محاسبه فاصله اقلیدسی را در نظر بگیریم تمام مقادیر ویژگیها جمع زده شده و از آنها رادیکال گرفته می شود. پس اگر یک یا چند ویژگی بوسیله متد کاهش ابعاد (صحیح یا غلط فرقی نمیکند) حذف گردد در اندازه فاصله نهایی مستقیماً تاثیر می گذارد هر چند مقدار آن کم و نزدیک به صفر باشد و این تاثیر (کاهش) کمتر از جَذر مقدار حذف شده است. (جزئیّات بیشتر در [پ-الف-۷])

به مثال زیر توجه کنید.

مثال – فرض کنید مقادیر ویژگیها به صورت زیر باشند. اگر ویژگی چهارّم با مقدار ۴ را حذف کنیم تاثیر حذف آن در فاصله اقلیدسی محاسبه شده ی نهایی تغییر محسوسی است. امّا اگر ویژگی با مقدار صفر حذف گردد تاثیری در اندازه فاصله ی اقلیدسی نهایی نخواهد داشت.

جدول ۱۱ - مثالی در مورد تاثیر اِعمال انتخاب ویژگی بر متدهای ایمنی مصنوعی مبتنی بر استراتژی فاصله حقیقی مبنا

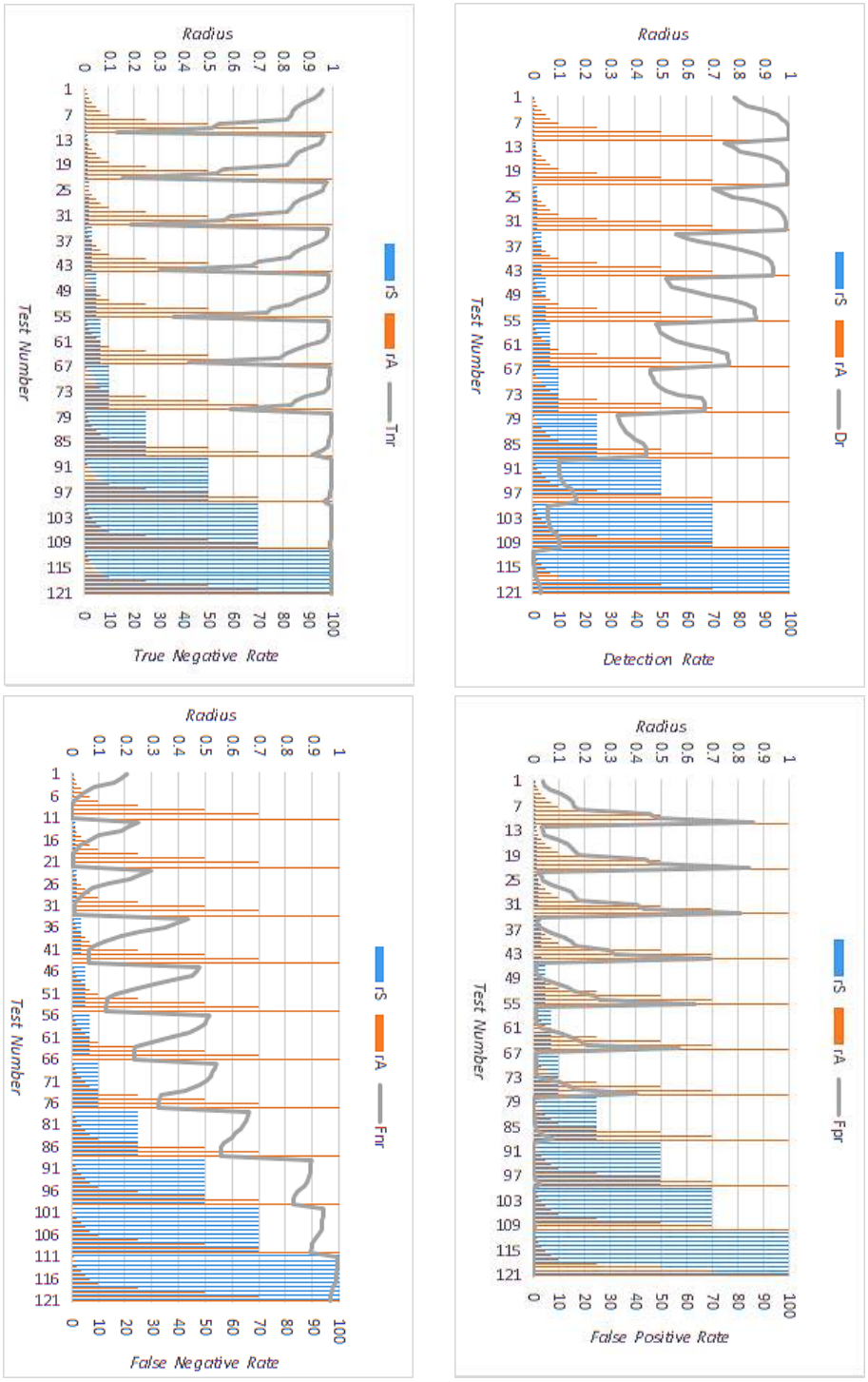
Features										Eucledian Distance
2	3.5	1.25	4	0	0.15	0.36	0.45	0.68	1.92	38.3159
After Delete										
2	3.5	1.25	4	Deleted	0.15	0.36	0.45	0.68	1.92	38.3159
2	3.5	1.25	Deleted	0	0.15	0.36	0.45	0.68	1.92	22.3159
2	3.5	1.25	4	0	Deleted	0.36	0.45	0.68	1.92	38.2934

لازم به یادآوری است که ویژگیهای ۲ تا ۴ در دادگانِ مربوطه که غیر عددی نیز هستند برای اجرای الگوریتم سلّولهای دندریت ضروری هستند زیرا مطابق بخش [۷-۲-۳] ، DCA پیشنهادی از این ویژگیها در استراتژی تقسیم بافتهای سلّولی استفاده می کند.

۳-۲-۳- آزمون الگوریتمِ انتخابِ منفیِ بهبود یافته با اوّلین زیر مجموعه آزمون استخراج شده

برای آزمون نهایی سیستم پیشنهادی نیاز به تعیین مقادیر مناسب شعاع مجموعه های خودی می باشد. تخصیص بهترین اندازه برای شعاع آنتی ژنهای خودی-نرمال و خودی-غیر نرمال می تواند نتایج مطلوبتری را در دسته بندی ارائه دهد. ما برای حصول بهینه ترین مقادیر برای شعاع خودی و کسب تجربه لازم ، آزمایشات زیر را با مقادیر پیشنهادی برای شعاع در بازه در $[0.01.0.1] = \{0.01.0.015.0.02.0.035.0.05.0.07.0.1.0.25.0.5.0.7.1\}$ در زیر مجموعه دادگان

آزمون نخست (دارای ۲۱۱۰ رکورد ترافیک) به انجام رسانیدم. این مجموعه آزمون ها، با و بدون اعمال انتخاب ویژگی انجام شده اند.



نمودار ۹ - تعیین شعاع خودی مناسب بدون اعمال الگوریتم انتخاب ویژگی

در آزمایشات انجام شده به منظور ارزیابی آزمون هایی که بهترین نتایج دسته بندی را با توجه به مقادیر شعاع خودی ارائه دهند نیاز به تعریف یک تابع فیتنس ضروری به نظر میرسید. بدین منظور رابطه ی زیر را پیشنهاد نمودیم. در رابطه پیشنهادی شِش معیار دسته بندی پایه باضافه زمان اجرا ، نقش اساسی داشته و در مجموع هفت پارامتر بدان اعمال شده اند.

$$\text{Fitness}_{\text{NSA}} = \frac{(((\text{Dr}-\text{FPr})+(\text{TNr}-\text{FNr}))*(\text{Acc}-\text{Err}))}{\text{Runtime}} \qquad (۳-۱۱)$$

با توجه به نمودارهای فوق ، بهترین دسته بندی مربوط به آزمون های ۶ تا ۹ و ۱۶ تا ۲۰ می باشد. در بین این آزمون ها نیز شعاع خودی-نرمال ۰,۰۱ و شعاع خودی-آنومالی ۰,۱ بهترین نرخ های دسته بندی را نتیجه می دهند. همچنین بدترین دسته بندی مربوط به آزمون های ۹۰ تا ۹۸ و ۱۰۹ تا ۱۱۱ بوده که بازه شعاع خودی-نرمال و شعاع خودی-آنومالی در آنها به ترتیب در بازه های ۰,۵ تا ۰,۷ و ۰,۰۱ تا ۱ می باشد.

بنابراین می توان نتیجه گرفت که بازه مناسب برای شعاع خودی-نرمال و شعاع خودی-آنومالی به ترتیب ۰,۰۱ تا ۰,۰۲ و ۰,۰۵ تا ۰,۲۵ هست. همچنین تاثیر الگوریتم انتخاب ویژگی نیز بررسی گردید و طبق جدول زیر مشاهده شد که کاربرد متد انتخاب ویژگی صرفاً در کاهش زمان محاسباتی موثر بوده و بر مبنای تابع فیتنس نمی تواند نتایج دسته بندی رضایت بخش تری را ارائه دهد.

جدول ۱۲ - نتایج آزمایش تعیین شعاع خودی مناسب

	DR	FPR	TNR	FNR	ACC	ERR	Time	rS	rA	rS Good Rates	rA Good Rates	Feature Selection	Fitness
Best Results	98.6	16.2	83.8	1.4	93.9	6.1	15.031518	0.01	0.1	0.01-0.02	0.05-0.25	-	962.60
	97.7	22	78	2.3	91.5	8.5	5.051869	0.015	0.02	0.01-0.07	0.015-0.1	CFA	2487.43
	95.4	24.4	75.6	4.6	89.2	10.8	3.259724	0.01	0.015	0.01-0.035 1	0.01-0.05 0.01-0.2	Ref.[16]	3415.25

۱-۳-۲-۳- تفسیر

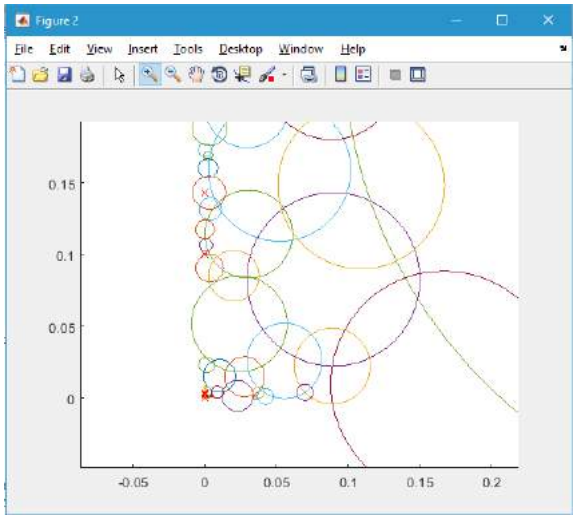
به دلیل آنکه کاهش زمان در رابطه فیتنس پیشنهادی (۳-۱۱) رابطه ی عکس دارد کاهش زمان اجرا می تواند مقدار زیادی نتیجه این رابطه را ارتقا دهد. مطابق اطلاعات جدول ، اعمالِ متد انتخاب ویژگی نقش بسزایی در کاهش زمان آزمون ها داشته و منجر به این شده که بازه شعاع مناسب برای خودی نیز تغییر محسوسی پیدا کند. علّت این تغییر آنست که با کاهش برخی از ویژگیها ، بُعد مسئله کاهش چشمگیری پیدا نموده و این مسئله سرعت اجرا را افزایش می دهد در عوض تاثیر کمی نیز در تنزّل کیفیت دسته بندی خواهد داشت. زیرا

با کاهش ویژگیها ، تعیین دقیق فاصله اقلیدسی بین نمونه های خودی و غیر خودی در فضای کاهش یافته مشکل تر می گردد.

به عبارت دیگر ، نیاز به تغییراتی در شعاع خودی احساس می شود تا بتواند فواصل را به اندازه همان وضعیت قبل از اعمال متد انتخاب ویژگی بُرده و بتواند نرخ های دسته بندی را نسبتاً ثابت نگه دارد. بنابراین در کُل می توان این گونه نتیجه گرفت که متد انتخاب ویژگی علیرغم افزایش سرعت اجرا به دلیل تاثیر گذاری مستقیم بر روی معیار فاصله اقلیدسی تاثیر چندانی در بهبود کیفیّت دسته بندی ندارد بنابراین به نظر نمی رسد که کاربرد انتخاب ویژگی برای بهبود تشخیص در خطوط دفاعی چندان موثّر باشد. دلیل این امر آنست که امروزه حجم منابع پردازشی و حافظه ای بسیار بالا رفته است که این، نیاز به افزایش سرعت را برطرف خواهد نمود.

۴-۲-۳- حالات ممکن در شناسایی نمونه ترافیک مشکوک شبکه (آنتی ژن) در فاز آزمون خط دوّم دفاعی

نمونه ترافیکِ شبکه پس از برچسب زنی اوّلیه بوسیله DCA جهت شناسایی و بررسی دقیق تر وارد خط دوّم دفاعی می شود. این لایه به نسخه بهبود یافته ای از مکانیسمِ الگوریتمِ انتخاب منفیِ حقیقی مبنا (RNSA) مجهّز شده است. در این راستا برای بهبود مکانیسمِ این متد در ادامه پژوهش، ترکیبی موفّق با الهام از ایده های موجود در مقالات [۸-۶] پیشنهاد گردید. شکل زیر باز نمایشی دو بعدی از فضای ابعاد مسئله است (مشخصاً ویژگیهای ۵ و ۶ از دادگان نفوذ). در این شکل نمونه های آزمون (آنتی ژنها) با علامتِ ضربدر قرمز مشخص شده اند. در جدول زیر به عملکردِ فازهای یادگیری و آزمون متد خط دوّم دفاعی در تخصیص احتمالِ برچسب کلاسِ نمونه آزمون می پردازیم.



شکل ۱۵ - آنتی بادیهای تولید شده در فضای کاهش یافته ابعاد مسئله (بازنمایشِ ابعاد ۵ و ۶ از دادگان نفوذ)

بسته به وضعیتهای مختلفی که بُردارِ موقعیّتِ نمونه آزمون ممکن است در فضای ابعاد مسئله داشته باشد استراتژی خاص و متفاوتی به عنوان راه حل در جهت پیش بینی احتمالِ برچسب کلاس و با هدف تعدیلِ مجموعه تش.د. ها، کاهش خطا و در نهایت تضمین ویژگی "خود ایمنیِ سیستم" می بایست به کار گرفته شود. در جدول پس از مشخص شدن وضعیّت ، ابتدا راه حل به اجرا در آمده و سپس وضعیت نمونه بررسی می گردد.

جدول ۱۳ – بررسی حالات ممکن در پیش بینی و تخصیص احتمال برچسب کلاس نمونه آزمون در فاز شناسایی الگوریتم انتخاب منفی بهبود یافته حقیقی مبنا

وضعیت	کلاس واقعی	کلاس پیش بینی شده	فضای تحت پوشش	نوع خطا	وضعیت نمونه	راه حل
۱	نرمال	نرمال	تش.د. خودی- نرمال	-	نمونه مشکوک تحت پوشش فضای خودی	-
۲	آنومالی			FN		کاهش (تعدیل [۱]) شعاع خودی- نرمال و سپس تعدیل فضا بوسیله فاز یادگیری ثانویه
۳	آنومالی	آنومالی	تش.د. خودی- آنومالی	-		-
۴	نرمال			FP		کاهش(تعدیل) شعاع خودی- آنومالی و سپس تعدیل فضا بوسیله فاز یادگیری ثانویه
۵	نرمال	نرمال	حفره	-	فاصله نمونه مشکوک نزدیکتر به نزدیکترین تش.د. خودی- نرمال	ارسال سیگنال بروز رسانی- تعدیل به فاز یادگیری اوّلیه جهت تولید تشخیص دهنده های غیر خودی بالغ و پوشش حفره
۶	آنومالی			FN		
۷	آنومالی	آنومالی		-	فاصله نمونه مشکوک نزدیکتر به نزدیکترین تش.د. خودی- آنومالی	
۸	نرمال			FP		
۹	نرمال	آنومالی	تش.د.بغ.م	FP	چالش عدم بروز بودن پروفایل خودی نرمال	لزوم بروز رسانی پروفایل تشخیص دهنده های نرمال و سپس ارسال سیگنال بروز رسانی- تعدیل به فاز یادگیری اوّلیه
۱۰	آنومالی			-	ثبت به عنوان تش.د.بغ.م.حظ [۲]	انتظار جهت بروز رسانی پایگاه امضاء های حملات از اینترنت

[۱] استراتژی پیشنهادی برای تعیین شعاع خودی نرمال/ آنومالی استفاده از کمینه / میانگین/ میانه فاصله بین نمونه های خودی بوده و بستگی به تراکم توزیع آنها در فضای ابعاد مسئله دارد. این استراتژی در بخش های بعدی ارائه شده است.

[۲] مزیّت بسیار مهم ثبت یک تشخیص دهنده ی بالغ به عنوان آنتی بادی حافظه آنست که امکان صحّت در پیش بینی احتمالِ برچسب کلاس نمونه مشکوک توسط چنین تشخیص دهنده ای در آینده بسیار نزدیک به واقعیّت بوده و صحّت عملکرد دسته بندی آن تایید می شود. بگونه ای که من بعد با هر بار شناسایی نمونه مشکوک، میزان بلوغ آن (از نظر صحّت پیش بینی برچسب کلاس) افزایش می یابد.

۵-۲-۳- قاعده حالات ممکن برچسب زنی[1] و احتمال درستی آنها:

میانگین نرخ های دسته بندی FP و TN برای خط دفاعی اوّل در شش آزمایش به ترتیب در حدود ۵۴,۸۷ و ۴۵,۱۳ و میانگین نرخ های DR و FN به ترتیب در حدود ۸۳,۲۵ و ۱۶,۷۵ می باشند. برای خط دفاعی دوّم نیز این چهار نرخ در واقع میانگین ۱۸ آزمونی است که صرفاً با الگوریتم انتخاب منفی انجام شده اند که به ترتیب برابر با $TNR = ۹۴,۵۳$ ، $FPR = ۵,۴۷$ ، $Dr = ۹۱,۹۱$ و $FNR = ۸,۰۹$ می باشند.

ملاحظه می شود که مهمترین مزیّت خط دفاعی نخست علیرغم خوشه بند بودن آن ، میانگین بالای نرخ تشخیص و نرخ نسبتاً پایین خطای منفی کاذب می باشد. امّا در خصوص نرخ های صحیح منفی و بخصوص مثبت کاذب ، مقادیر نسبتاً بالایی ندارند. این نکته نشان می دهد که استراتژی دفاعی پیشنهادی به کار برده شده برای نخستین خط دفاعی در سیستم پیشنهادی که بر مبنای تئوری خطر می باشد توانایی نسبتاً بالایی در شناسایی درست ترافیک نفوذ دارد (چیزی حدود بالای ۸۳ درصد) بطوریکه میزان خطای آن در شناسایی اشتباه ترافیک نفوذ پایین بوده و در حدود ۱۶ درصد ارزیابی می گردد.

ولی خط دفاعی اوّل نقطه ضعف بزرگی در شناسایی صحیح ترافیک نرمال دارد بطوریکه با شناسایی اشتباه ترافیک نرمال به عنوان نفوذ ، بی جهت خطا را بالا می برد. بطوریکه در بالا مشاهده می شود میانگین این نوع خطا چیزی حدود ۴۵ درصد می باشد که برای یک سیستم نیمه نظارتی مانند DCA پیشنهادی، بد نبوده و در مقایسه با خوشه بندهای متمرکز و غیر توزیع شده مانند $K-means. PSO. GA. ACO. BHA$ مناسب به نظر میرسد. زیرا هدف اصلی هر سیستم تشخیص ، شناسایی عوامل نفوذ می باشد که با ارتقای نرخ تشخیص و کاهش خطای منفی کاذب می توان به این هدف تا حدود زیادی دست یافت.

از طرفی، وجود خطای مثبت کاذب تنها هزینه انجام محاسبات را بالا می برد. پس طبق مواردی که در انتهای فصل دوّم – ادبیات نظری (۲-۱-۲-۸) و آزمایشات بخش ۳-۲-۳ (نمودار ۹) در خصوص سیستم های تشخیص نفوذ مبتنی بر ایمنی مصنوعی مطرح گردید ، مشاهده گردید که مهمترین چالش این سیستمها همواره پایین آوردن نرخ مثبت کاذب در آزمایشات میباشد. علاوه بر این در ارزیابی خط دوّمِ دفاعی و میانگین های بدست آمده از هجده آزمایش انجام شده به این نتیجه نیز میرسیم که الگوریتم انتخاب منفی پیشنهادی دارای ثُبات کافی در شناسایی ترافیک نفوذ و ترافیک نرمال، هر دو به یک میزان نسبتاً بالا (چیزی در حدود بالای ۹۱ یا ۹۴ درصد) را دارا می باشد.

لازم به توضیح است که وجود حتّی چند دهم درصد برای خطاهای مثبت و به خصوص منفی کاذب درصد بالایی است زیرا نفوذ یک ترافیک آنومال به شبکه داخلی سازمان ، در موارد بسیاری به حملات متعاقب دیگری

به عبارت بهتر، در صورت حافظه بودنِ یک آنتی بادیِ غیر خودیِ بالغ ، در آزمونهای بعد در صورت شناسایی نمونه ای مشکوک ، به احتمال زیاد صحّت نتیجه برچسب زنی به واقعیت نزدیک خواهد بود.

[1] TP(Recall), TN, FP, FN Rates

منجر می گردد که چه بسا جلوگیری - بلاک نمودن ترافیکی که قصد شناسایی شبکه را داشته - اگر از همان ابتدا انجام گردد - قدم مهمی در پیشگیری از حملات بعدی به شمار خواهد رفت.

ما در سیستم پیشنهادی خود ضمن آنکه تا حدود زیادی به هدف اصلی مان که شبیه سازی و ارزیابی حیات آنتی ژنها (منظور ترافیک مشکوک شبکه) به شکل مصنوعی بود رسیدیم ، به هدف مهم تر که همان مقابله با چالش فوق الذکر سیستم های ایمنی مصنوعی (پایین آوردن خطای منفی و بخصوص مثبت کاذب) است نیز دست یافتیم که در آزمایشات بخش بعدی ثُبات این ادّعا را با طی چند سیکل از این حیات مصنوعی و مشاهده و ارزیابی نتایج کافی در بخش بعد به اثبات رسانده ایم.

در بخش [پ-الف-۱۴]، قاعده احتمالی فوق را به سیستم تشخیص نفوذ پیشنهادی بسط داده و ضمن در نظر گرفتن تمامی حالتهای ممکن در برچسب زنی ، اثر بخش بودن این ادّعا را در کاهش نرخ های خطا و افزایش عملکرد دسته بندی نهایی به دقّت و بصورت نظری تحلیل نموده ایم. با نتیجه گیری از تحلیل نظری در [پ-الف-۱۴] می توان این مسئله را با مثالی ساده به صورت زیر بیان نمود. در بخش های بعدی به آنالیز تجربیِ روش ترکیبی پیشنهادی در محیط شبیه سازِ متلب پرداخته ایم.

یک مثال

فرض کنید ما دادگانی به صورت زیر در اختیار داریم و قصد داریم آنرا با استفاده از روش ترکیبی پیشنهادی (دو الگوریتم DC و NS پیشنهادی به عنوان خطوط دفاعی) دسته بندی نماییم.

جدول ۱۴ - دسته بندی فرضی بوسیله مکانیسمِ خط نخست دفاعی (DCA پیشنهادی)

Antigen ID	Target	Output Label	Conf.
T1	0	0.12	TN
T2	1	0.49	FN
T3	1	0.81	TP
T4	0	0.98	FP
T5	1	0.45	FN
T6	0	0.27	TN
T7	1	0.99	TP
T8	1	0.74	TP

جدول ۱۵ - دسته بندی فرضی بوسیله مکانیسمِ خط دوّم دفاعی (NSA)

Output Label	Conf.
0.23	TN
0.66	TP
0.34	FN
0.76	FP
0.91	TP
0.37	TN
0.45	FN
1	TP

جدول ۱۶ - نتیجه دسته بندی سیستم پیشنهادی و تصمیم گیری در مورد برچسب احتمالی نمونه های عدم قطعیّت با استفاده از رابطه ضرایب وزنی پیشنهادی

Final Label	Conf.	Final Label by equation. [3-12] $\alpha = 0.3 \, . \beta = 0.7$	Conf.	Final Label by equation. [3-12] $\alpha = 0.5 \, . \beta = 0.5$	Conf.
0.28	TN	-			
0.78	Not-Sure	0.609	FN	0.575	TP
0.91	Not-Sure	0.481	TP	0.575	FN
0.84	FP	-			
0.48	Not-Sure	0.772	FN	0.68	TP
0.09	TN	-			
0.47	Not-Sure	0.612	TP	0.72	FN
0.52	TP	-			

مطابق اطلاعات جداول فوق ملاحظه می گردد که در این آزمایشِ فرضی ، چهار نمونه ، وضعیّت عدم قطعیّت دارند. حال اگر حدآستانه سیکل اجرا برای سیستم پیشنهادی اجازه اجرای مجدّد این ۴ آنتی ژن را بدهد. در این صورت این نمونه ها وارد سیکل اجرای بعدی می شوند و بهمراه نمونه های جدید مجدّداً مورد آنالیز و دسته بندی خطوط دفاعی قرار میگیرند تا در نهایت تصمیم گیری شود که برچسب احتمالی آنها چیست؟ ولی اگر از حدآستانه سیکلهای قابل اجرای سیستم گذشته باشد در این صورت با استفاده از رابطه وزنی پیشنهادی در خصوص برچسب نهایی این نمونه ها مطابق زیر تصمیم گیری می شود. توضیح اینکه اگر فرض کنیم که این حدآستانه گذشته باشد سیستم پیشنهادی با استفاده از رابطه وزنی به صورت زیر در مورد برچسب نهایی نمونه های عدم قطعیّت تصمیم گیری می نماید.

$$\mathrm{Conf_{Final}}(t_i) = \alpha \times \mathrm{Conf_{DCA}}(t_i) + \beta \times \mathrm{Conf_{NSA}}(t_i) \qquad (۳-۱۲)$$

مقادیر آلفا و بتا یکبار به ترتیب برابر با 0.3 و 0.7 و بار دوّم هر دو برابر 0.5 تعیین گردید. طبق اطلاعات جدول ملاحظه می گردد که سیستم پیشنهادی در مجموع نتایج رضایت بخشی را در دسته بندی ارائه نموده است. بخصوص زمانی که مقدار دو پارامتر ضریب وزنی آلفا و بتا برابر با 0.5 در نظر گرفته شود در این حالت

نرخ تشخیص صحیح نفوذ بسیار بالاتر خواهد رفت. البته تعیین مقادیر این دو ضریب وزنی پیشنهادی بستگی زیادی به تجربه کاربر متخصّص در کار با سیستم دارد که به مرور زمان این تجربه قابل کسب است.

در ادامه آزمایشات مشاهده می شود که اگر دادگان آزمون بوسیله هر یک از خطوط دفاعی به طور مجزّا مورد آزمون قرار گیرند نرخ های عملکرد دسته بندی آنها نسبت به عملکرد سیستم تشخیص نفوذ پیشنهادی بهتر نخواهد بود.

۶-۲-۳- پارامتر $AVote$

اگر تش.د.بغ نزدیک به فضای خودی – نرمال قرار بگیرد به شرط وقوع حالت قطعیّت در خطوط، دو وضعیّت ممکن در آزمون نمونه ها رخ خواهد داد : $Semi - D - TN$ و یا $Mat - MD - TN$.

در صورتی که آنتی بادی بالغ نزدیک فضای خودی –نرمال باشد و برچسب اوّلیه نمونه دن.نبغ. (نرمال) شناسایی گردد ، معیاری لازم است تا پتانسیل بالقوّه و استعداد تش.د.بغ. مربوطه را در شناسایی عوامل نرمال ارزیابی کند. بدین منظور پارامتر AVote را پیشنهاد نمودیم. ایده پیشنهادی بدین صورت است که هر کدام از تش.د. ها در فضای غیر خودی ، دارای یک پارامتر AVote هستند که مقدار اوّلیه آن صفر است. با هر تشخیص عوامل نرمال به شرط وقوع حالت قطعیّت در خطوط ، مقدار این پارامتر در تش.د. ای که آن نمونه آزمون را شناسایی نموده ، یک واحد کمتر می شود و در واقع منفی می شود.

هر چه مقدار آن منفی تر و از صفر دورتر باشد ، نشان دهنده ی پتانسیل بالقوّه ی آن تش.د. در شناسایی عوامل نرمال است. اگر مقدار $AVote \geq 1$ یک گردد ، تش.د. تبدیل به آنتی بادی حافظه می گردد و من بعد به ازاء شناسایی هر نمونه ی آزمون ، مقدار این پارامتر در این حالت بالاتر می رود. در بخش بعد یک سناریوی فرضی را به منظور درک بهتر مطلب ارائه نموده ایم.

۷-۲-۳- درجه تشخیص – استراتژیِ پیشنهادی برای تخصیصِ برچسبهای نمونه ها در فاز آزمون خط دوّم دفاعی

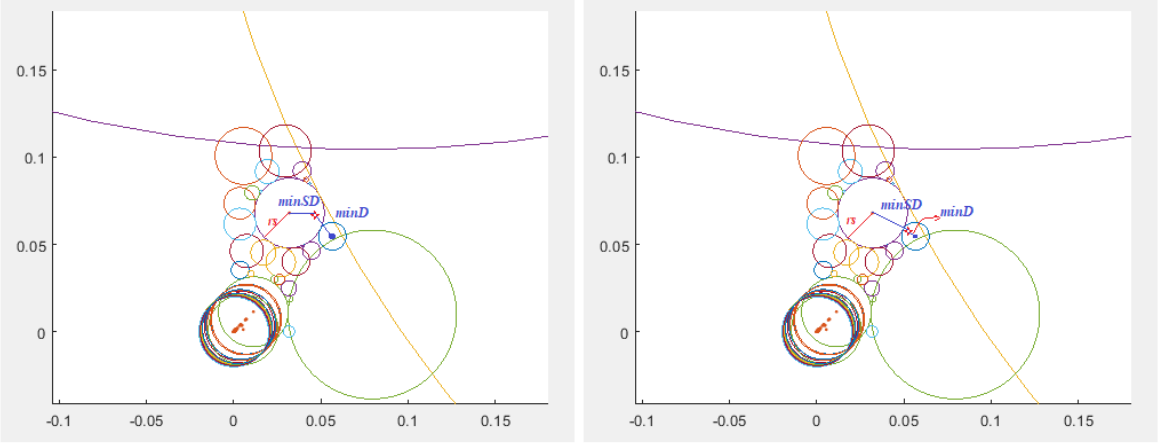
در فاز آزمون برای یک نمونه آزمون t بسته به بردار موقعیّت آن در فضای ابعاد مسئله چهار وضعیت ممکن وجود دارد. موقعیّت t :

الف) در فضای تحت پوشش خودی - نرمال قرار گیرد.
ب) در فضای تحت پوشش خودی - آنومالی قرار گیرد.

ج) در فضای تحت پوشش غیر خودی قرار گیرد بطوریکه نزدیک به فضای خودی- نرمال باشد.
د) در فضای تحت پوشش غیر خودی قرار گیرد بطوریکه نزدیک به فضای خودی - آنومالی باشد.

رخدادِ حُفره در تمام این چهار وضعیت ممکن است. در صورت رخداد حفره ، فاز یادگیری فراخوانی می شود تا با تعدیل و بروز رسانی تش.د. ها حفره ها تا حد ممکن پوشش داده شوند. این فاز همواره به شکل موازی و مستقل از فاز آزمون و دیگر خطوط دفاعی ، پیوسته در حال پوشش حفره ها و آموزش شامل تولید و تکثیر آنتی بادیهاست. اگر مقادیر برچسب نمونه ها به صورت گسسته تعیین گردند (صفر برای نرمال و یک برای آنومالی)، در اینصورت این شیوه برچسب زنی دقّت چندانی نداشته و نمی تواند معیار مناسبی برای محاسبات احتمالی در ماژول تصمیم گیری سیستم تشخیص نفوذ پیشنهادی باشد.

بخصوص به دلیل آنکه ما دو خط دفاعی داریم بنابراین اعمال یک رابطه وزنی برنتیجه برچسب خطوط عملی نخواهد بود.



شکل ۱۶ – سمت چپ) وضعیّتهای الف و ب را نشان می دهند که نمونه آزمون تحت پوشش فضای خودی نرمال/غیر نرمال قرار دارد. سمت راست) وضعیتهای ج و د را نشان می دهد که نمونه آزمون تحت پوشش تش.د. غیر خودی قرار داشته و بسته به فاصله آن از هر یک از مجموعه های خودی نرمال و خودی غیر نرمال در مورد برچسب احتمالی نهایی آن تصمیم گیری می شود.

بنابراین نیاز داریم که بدانیم نمونه ای که برچسب آن 0 شناسایی شده واقعاً تا چه حد نرمال بوده و این احتمال بین 0 تا 0.499 متغیّر خواهد بود، بطوریکه عدد صفر نشان دهنده ی آن خواهد بود که نمونه صد در صد نرمال است. بدین منظور ما ایده "پیوسته سازی برچسب نهایی t " را به عنوان یک استراتژی مناسب که

می تواند برچسب t را به صورت احتمالی با دقّت بالا در بازه $[0.1]$ برچسب بزند به صورت روابط زیر پیشنهاد نمودیم.

جدول ۱۷ – روابط پیشنهادی برای محاسبه برچسب احتمالی نمونه آزمون در وضعیت الف

$\mathbf{min_{NSD}}$	**0**	$\mathbf{r_{nsd}/2}$	$\mathbf{3r_{nsd}/4}$	$\mathbf{r_{nsd}}$	$\mathbf{r_{nsd} + (r_D/4)}$	$\mathbf{r_{nsd} + (r_D/2)}$	$\mathbf{r_{nsd} + r_D}$
$\min_D$	$r_{nsd} + r_D$	$)r_D + (r_{nsd}/2$	$)r_D + (r_{nsd}/4$	r_D	$3r_D/4$	$r_D/2$	0
P_{norm}	1	0.75	0.625	0.5001	0.375	0.25	0
P_{anom}	0	0.25	0.375	0.4999	0.625	0.75	1

جدول ۱۸ – روابط پیشنهادی برای محاسبه برچسب احتمالی نمونه آزمون در وضعیت ب

$\mathbf{min_{ASD}}$	**0**	$\mathbf{r_{asd}/2}$	$\mathbf{3r_{asd}/4}$	$\mathbf{r_{asd}}$	$\mathbf{r_{asd} + (r_D/4)}$	$\mathbf{r_{asd} + (r_D/2)}$	$\mathbf{r_{asd} + r_D}$
$\min_D$	$r_{asd} + r_D$	$)r_D + (r_{asd}/2$	$)r_D + (r_{asd}/4$	r_D	$3r_D/4$	$r_D/2$	0
$P_{knownAttack}$	1	0.75	0.625	0.5001	0.375	0.25	0
$P_{UnknownAttack}$	0	0.25	0.375	0.4999	0.625	0.75	1

روابط مربوط به وضعیت الف)

$$\text{if } \mathrm{Dist(t.nearestNSD)} \le r_{nsd}$$

$$\rightarrow \quad \begin{cases} P_{norm} = 1 - \dfrac{2r_{nsd} - \min_{NSD}}{2r_{nsd}} \\ P_{anom} = 1 - P_{norm} \end{cases}$$

$$\text{elseif } r_{nsd} < \mathrm{Dist(t.nearestNSD)} < \mathrm{Dist(t.nearestNSD)} + \mathrm{Dist(t.nearestD)}$$

$$\rightarrow \quad \begin{cases} P_{anom} = \dfrac{2r_D - \min_D}{2r_D} \\ P_{norm} = 1 - P_{anom} \end{cases}$$

(۳-۱۳)

روابط مربوط به وضعیت ب)

دقّت گردد که در این وضعیّت اگر نمونه تحت پوشش فضای خودی غیر نرمال ، ASD قرار گیرد احتمال آنکه حمله شناخته شده باشد بیشتر بوده و در غیر اینصورت احتمال ناشناخته بودن حمله بالا میرود. در این وضعیّت فرض نرمال بودنِ نمونه مردود می باشد.

$$\text{if } \mathrm{Dist(t.nearestASD)} \le r_{asd}$$

$$\rightarrow \quad \begin{cases} P_{knownAttack} = \dfrac{2r_{asd} - \min_{ASD}}{2r_{asd}} \\ P_{UnknownAttack} = 1 - P_{knownAttack} \end{cases}$$

$$\text{elseif } r_{asd} < \text{Dist}(t.\text{nearestASD}) < \text{Dist}(t.\text{nearestASD}) + \text{Dist}(t.\text{nearestD}) \rightarrow \begin{cases} P_{\text{UnknownAttack}} = \dfrac{2r_D - \min_D}{2r_D} \\ P_{\text{knownAttack}} = 1 - P_{\text{UnknownAttack}} \end{cases} \quad (۳-۱۴)$$

روابط مربوط به وضعیت ج و د)

در این وضعیّت سه حالت ممکن است برای نمونه t اتفاق بیافتد که به صورت روابط زیر می توان بیان نمود:

$$\text{if Dist}(t.\min_{NSD}) < \text{Dist}(t.\min_{ASD})$$

$$\text{if at least one Covered } Ab_i \text{ be a memorized}_{Ab_i}$$

$$P = 1 - (\text{dist of The nearest detected Ag to the Current Ag/ Sum of distances from Current Ag to all detected Antigens into coverer mature} - \text{Antibody})$$

$$L = P \times \text{Nearest detected Ag class label} \quad \textit{; If label be lower than 0.49 Then L= [label – (P×label)] +label}$$

$$\text{Final Label} = \text{mean}(L.P_{anom}) \quad (۳-۱۵-۱)$$

else

$$P_{norm} = \left(\frac{\text{Dist}(t.\min_{NSd})}{\text{Dist}(t.\min_{NSD}) + \text{Dist}(t.\min_{ASD})}\right)$$

$$P_{anom} = 1 - P_{norm} \quad (۳-۱۵-۲)$$

else

$$P_{anom} = 1 - \left(\frac{\text{Dist}(t.\min_{ASd})}{\text{Dist}(t.\min_{NSD}) + \text{Dist}(t.\min_{ASD})}\right)$$

$$P_{norm} = 1 - P_{anom} \quad (۳-۱۵-۳)$$

end

تحلیل این روابط بهمراه اینفوگرافیک مربوط به استراتژی پیشنهادی برای فاز آزمون خط دوّم در [پ-الف-۱۵] بتفصیل بیان شده اند.

۱-۷-۲-۳- یک سناریوی فرضی

فرض کنید تشخیص دهنده ی بالغ غیر حافظه ی Ab_i نزدیک به فضای خودی – آنومالی یا ASD ، قرار گیرد. مقدار پارامتر AVote این تشخیص دهنده برابر صفر است. حال اگر نمونه ی آزمون فرضی Ag_1 تحت پوشش آنتی بادی مذکور قرار گرفته و توسط آن شناسایی گردد ، به دلیل نزدیک بودن فاصله نمونه به فضای ASD

و بشرط وقوع حالت قطعیّت در خطوط (یعنی اگر خط دفاعی اوّل نیز برچسب آنومالی به نمونه زده باشد) ، حالت $Mat - D - TA$ رخ داده و مقدار پارامتر $AVote++$ تشخیص دهنده مذکور یک واحد اضافه می گردد. می دانیم مقدار این پارامتر به محض یک شدن منجر به این می شود که آنتی بادی بالغ تبدیل به حافظه گردد. بدین ترتیب Ab_i که با این تشخیص تبدیل به حافظه شده، پتانسیل بالقوّه ی آن در شناسایی عوامل ناشناخته بالاتر رفته و در سیکلهای اجرای بعدی در صورت برقراری شرط قطعیّت و نزدیک بودن آنتی بادی به فضای خودی – غیر نرمال ، وضعیّت $Mat - MD - TA$ رخ می دهد.

حال اگر نمونه جدید آزمون Ag_2 در وضعیت $Mat - MD - TA$ قرار گیرد ، به دلیل آنکه آنتی بادی حافظه مربوطه قبلاً یک شناسایی موفّق داشته و پارامتر آن برابر یک می باشد ، در این حالت نیز با شناسایی عامل ناشناخته، مقدار پارامتر آن افزوده و ۲ شده و بدین ترتیب پتانسیل حافظه ای آن بالاتر می رود. به عبارت بهتر ، با افزایش پارامتر AVote و هر چه که از صفر دورتر می شویم ، میزان تثبیت / تکثیر آنتی بادی حافظه مربوطه بیشتر می شود.

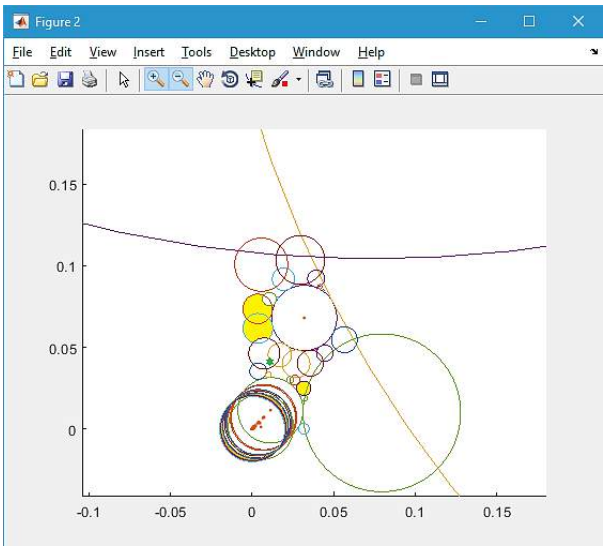
در این پژوهش منظور از تکثیر ، همان تثبیت تشخیص دهنده در فضای ابعاد مسئله بوده و با اجرای فازهای یادگیری و تعدیل فضای غیر خودی ، تشخیص دهنده حافظه دچار تعدیل نمیگردد.

حال اگر نمونه آزمون دیگری را مانند Ag_3 در نظر بگیریم که واقعاً نرمال بوده و تحت پوشش فضای Ab_i قرار گیرد وضعیت $Semi - MD - TA$ ممکن الوقوع خواهد بود. این وضعیت غیر قطعی بوده و بدین ترتیب نمونه Ag_3 به منظور بررسی و آنالیز بیشتر در FeedBack بافر شده و در سیکلهای اجرای بعدی سیستم مورد شناسایی قرار خواهد گرفت. البته به شرطی که حدآستانه ی مربوط به حداکثر سیکلهای قابل اجرای سیستم فرا نرسیده باشد و بتوان نمونه های موجود در فیدبک را بازهم مورد آنالیز قرار داد.

به همین ترتیب آن دسته از نمونه های موجود در بافر فیدبک که تاکنون حدآستانه ی سیکلهای اجرا در مورد آنها گذشته است به منظور تعیین برچسب نهایی وارد فاز تصمیم گیری نهایی می شوند تا رابطه وزنی برچسب نهایی (۱۲-۳) را به صورت احتمالی در بازه ۰ تا ۱ محاسبه و تصمیم گیری نماید. حدآستانه سیکلهای قابل اجرا نیز توسط کاربر متخصّص بر حسب تجربه ی کار با سیستم تعیین می گردد که متناسب با حداکثر زمان تاخیر قابل قبول در آنالیز ترافیک ورودی شبکه می باشد که معمولاً زمان کمی بوده ولی انتخاب این حدآستانه خود یک چالش بزرگ محسوب میگردد.

تا به اینجای کار ، Ab_i دو بار اقدام به شناسایی موفّق نفوذ نموده و مقدار پارامتر آن ۲ می باشد. در این وضعیت اگر هر دو پروفایلها و امضاءها (ASD. NSD) طی یک سناریوی فرضی بروز شده و این بار به Ab_i فضای خودی – نرمال نزدیکتر گردد به دلیل مجاور بودن با این فضا، به محض شناسایی عوامل نفوذ جدید وضعیت قطعیتِ $Mat - MD - TN$ رخ داده و یک واحد از مقدار پارامتر AVote کم می شود و به همین ترتیب با شناسایی عوامل نفوذ دیگر نیز مقدار این پارامتر بازهم کمتر و کمتر شده و زمانی میرسد که برابر با صفر گردد.

در صورت صفر شدن ، آنتی بادی ، پتانسیل و استعداد لازم برای شناسایی عوامل نرمال را پیدا نموده (پلاتِ پایین) و من بعد در صورت رخداد وضعیّت قطعیّتِ $Semi-D-TN$ قادر خواهد بود که هر نمونه نرمال را با احتمال بالایی شناسایی نماید.



شکل ۱۷ – شناسایی آنتی ژن نرمال (سبز رنگ) توسط آنتی بادی بالغ غیر حافظه - وضعیّت $Semi-D-TN$

۸-۲-۳- مرگِ برنامه ریزی شده آنتی بادیهای نابالغ و تاثیر آن در افزایشِ زمانِ محاسبات

در سیستم ایمنی بیولوژیک بدن انسان ، لنفوسیتهای T پس از تولید در مغز استخوان بلافاصله وارد غدّه تیموس می شوند تا طی فرایند انتخاب منفی آموزش لازم را ببینند. در این آموزش آنها می آموزند که با آنتی ژنهای خودی واکنش نشان ندهند. همچنین لنفوسیتهای B نیز که سطح آنها حاوی پرزهایی است که اصطلاحاً به آن آنتی بادی گفته می شود نیز آموزش خاص خود را دارند که در مغز استخوان صورت میپذیرد .

هر لنفوسیت B بر روی سطح خود آنتی بادیهایی با الگوهای مشخصی دارد که فقط به آنتی ژنهای نفوذی با الگوی خاص می چسبند. معمولاً تولید این نوع لنفوسیتها و نوع الگو و فرمت خاص آنتی بادیهای روی سطح آنها به صورت تصادفی در مغز استخوان اتفاق می افتد و تابع مورد خاصی نیست. آنتی بادیها پس از طی فرایند انتخاب منفی و آموزش لازم بالغ شده و در بدن جریان می یابند تا با الگوهای غیر خودی و نفوذی تطبیق بخورند و آنها را شناسایی نمایند. انتخاب منفی به طور مفصّل در فصل دوّم ادبیات نظری تشریح شد.

حال جای سوال است که اگر لنفوسیتی با الگوی آنتی بادی خاص پس از تولید در مغز استخوان بالغ نشد چه اتفاقی رخ می دهد ؟ اگر تعداد آنتی ژنهای نابالغ زیاد باشد سیستم ایمنی بدن چه عکس العملی خواهد داشت؟

نابالغ بودن لنفوسیت B بدان معناست که در فرایند آموزش انتخاب منفی ، الگوی آنتی بادی تصادفی تولید شده سطح آن لنفوسیت طوری بوده که با حداقل یک مورد از آنتی ژنهای خودی بدن تطبیق موفق انجام داده و آنرا شناسایی نموده و در پروسه سخت انتخاب منفی حذف شده و بالغ نشده است. به این فرایند غربالگری مرگ برنامه ریزی شده آنتی بادیهای نابالغ گویند. حال این پروسه بیولوژیک چگونه به مسئله تشخیص نفوذ شبکه در پژوهش جاری نگاشت می شود؟

در مسئله تشخیص نفوذ نیز در هر لحظه الگوریتم انتخاب منفی تعداد زیادی تشخیص دهنده کاندید با بردارهای موقعیت کاملاً تصادفی یکنواخت را در فضای ابعاد مسئله [0,1] تولید می کند. حال نابالغ بودن تشخیص دهنده ها به این معناست که بردار موقعیت آنتی بادی مذکور توسط فضای خودی- نرمال پوشش داده می شود بنابراین نمی تواند بالغ گردد ، بنابراین حذف شده و بردار موقعیت تصادفی دیگری انتخاب و چک می شود که تحت پوشش فضای خودی نباشد. اگر تعداد آنتی بادیهای حذف شده زیاد باشد زمان محاسباتی الگوریتم بالا می رود چرا که پیوسته باید در فضای ابعاد مسئله به تولید تصادفی آنتی بادی ها بپردازد. همچنین تراکم بالای ناحیه خودی نیز اثر مستقیمی بر کاهش زمان محاسباتی تولید آنتی بادیها و بلوغ موفق آنها می گذارد. بلوغ یک تشخیص دهنده کاندید در مسئله تشخیص نفوذ به معنای عدم پوشش آن توسط هیچ یک از نمونه های خودی می باشد.

ایده ای که در راستای بهینه سازی و بلوغ آنتی بادیهای تصادفی تولید شده به ذهن می رسد استفاده از استراتژی مختصات قطبی[1] می باشد که در [۸] محقق از استراتژی این ایده در پیشنهاد " تابع جهش" آنتی بادیهایی که هیچ تشخیصی انجام نداده اند استفاده نموده است. زیرا همانگونه که می دانیم در حل مسائل با فضای n بعدی از مختصات قطبی استفاده می شود. به نظر میرسد استفاده از این ایده راه کار مناسبی در جهت کاهش زمان محاسباتی تولید تصادفی تشخیص دهنده ها و افزایش بلوغ آنها ارائه دهد.

چالش مذکور در تولید تشخیص دهنده ها در فاز یادگیری ثانویه و در نتیجه تعدیل فضای خودی نیز وجود دارد که زمان محاسباتی یادگیری را بالا می برد. زیرا در اکثر موارد بردار های موقعیت تشخیص دهنده های تصادفی تولید شده در داخل فضای خودی قرار نمیگیرند و حذف می شوند. در نتیجه استفاده از ایده مختصات قطبی به طریق زیر ، می تواند تشخیص دهنده های کاندید را دقیقاً در داخل محدوده تحت پوشش خودی [2] با بردارهای موقعیت کاملاً تصادفی تولید نموده و ضمن جلوگیری از حذف تشخیص دهنده های خودی از زمان محاسباتی فاز یادگیری ثانویه نیز بسیار بکاهد.

[1] Polar Coordination
[2] Super − ring

۱-۸-۲-۳- یک ایده مُوثّر - استفاده از استراتژی مختصات قطبی برای تعدیل فضای خودی و کاهش زمان محاسباتی

با استفاده از مختصات قطبی می توان تولید تصادفی و یکنواخت آنتی بادیهای خودی را محدود به ناحیه خاصی[۱] نمود. این ناحیه خاص ، درون تشخیص دهنده های خودی با شعاع ثابت r_s می باشد. بطوریکه به محض تولید نمونه های کاندید در درون هر تشخیص دهنده ی خودی ، این نمونه ها هر یک می توانند با شعاع متغیّر حفره های موجود در تمامی فضای خودی را پوشانده و از لزوم تعداد زیاد تشخیص دهنده های خودی با شعاع ثابت بکاهند. زیرا این حجم زیاد زمان محاسباتی را در فاز یادگیری ثانویه افزایش می دهد. روابط زیر مختصات قطبی مربوط به فضای n بعدی می باشد. [۸]

$$
\begin{aligned}
d'y_1 &= dy_1 + \rho.cos(\theta_1) \\
d'y_2 &= dy_2 + \rho.sin(\theta_1)\,cos(\theta_2) \\
&\vdots \\
d'y_{n-1} &= dy_{n-1} + \rho.sin(\theta_1)\,sin(\theta_2)\,\dots\,cos(\theta_{n-1}) \\
d'y_n &= dy_n + \rho.sin(\theta_1)\,sin(\theta_2)\,\dots\,sin(\theta_{n-1})
\end{aligned}
$$

(۳-۱۶)

$\mathrm{d'}$ تشخیص دهنده کاندید جدید با n بعد $\mathrm{d'y_1 d'y_2 \dots d'y_{n-1} d'y_n}$ می باشد. همچنین d نیز تشخیص دهنده خودی به مرکزیّت $\mathrm{dy_1 dy_2 \dots dy_{n-1} dy_n}$ و شعاع قطبی [۲] ρ است. ρ در بازه $[\frac{\sqrt{\mathrm{n}}}{\exp(\mathrm{t})}.\frac{\sqrt{\mathrm{n}}}{\exp(\mathrm{t-1})}]$ متغیر است. t میزان تکامل در تولید تش.د. هاست و مقدار آن هر چه کمتر و به سمت صفر تعیین گردد به میزانی مشخص توزیع تصادفی تش.د. های کاندید تولید شده از مرکز دور تر شده و میزان تولید در حاشیه فَراشکل و حتّی بیرون از آن بیشتر خواهد بود و بالعکس هر چه مقدار این پارامتر بیشتر تعیین گردد گریز از مرکز تشخیص دهنده های تولید شده کمتر بوده و توزیع آنها متمرکز و نزدیک به مرکز فراشکل خواهد بود. در واقع t پارامتری برای کنترل و میزان سازی توزیع یکنواخت داده های تصادفی درون فراشکل

[۱] منظور اشکال n بعدی در مسائل با ابعاد بالا می باشد. بطور مثال معادل یک دایره در فضای دو بعدی ، کره ای با شعاع مشابه در فضای سه بعدی می باشد. همینطور برای ابعاد بالاتر نیز به همین صورت بوده و همین کره تبدیل به یک $\mathrm{Super-Ring}$ می شود که یک فراشکل میباشد. البته درک انسان از تصوّر فرا اشکال $\mathrm{Shape-Space}$ با ابعاد بالا ناتوان است.

[2] Polar Diameter

می باشد به نحوی که داده ها نه خارج از فراشکل و نه زیاد به مرکز تولید شوند. همچنین $\theta_1\theta_2.\ldots.\theta_{n-1}$ در بازه $[0.360]$ انتخاب می شود.

در [۶] نسخه ای از RNSA تحت عنوان FtNSA [1] با دو فاز یادگیری اولیه و ثانویه ارائه شده که دلیل پیشنهاد فاز ثانویه بیشتر به منظور تعدیل فضاهای خودی NSD و ASD (خودی نرمال و خودی آنومالی) و در نتیجه آن کاهش زمان محاسباتی در فاز آزمون عنوان شده است. ما ضمن استفاده از این استراتژی در فاز یادگیری خط دوّم روش پیشنهادی ، با استفاده از ایده پیشنهادی مختصات قطبی سعی نمودیم تا پوشش حفره های موجود در فضاهای کوچک در لا به لای تشخیص دهنده های بالغ از حیث زمان تولید و پوشش حفره بهبود ببخشیم.

میدانیم که چالش فاز یادگیری نخست ، زمان بالای تولید تشخیص دهنده ها در فضای غیر خودی بوده بطوریکه طبق آزمایشات انجام شده در [پ-الف-۶] زمان لازم برای تولید تش.بغ بخصوص در اوّلین بار اجرای الگوریتم به صورت نمایی افزایش می یابد: $\sum_{i=1}^{d}\left(\left(\sum_{j=1}^{i}C_j\right)N_i+i-1\right)\leq O(n_2)$. که در آن C_j ناحیه پوشش برای تشخیص دهنده j و $\mathrm{N_i}\leq\mathrm{N_t}=\mathrm{N_{sn}}+\mathrm{N_{sa}}+\mathrm{N_d}$ تعداد j نرمال خودی و آنومالی خودی و تشخیص دهنده های غیر خودی هستند که بوسیله تشخیص دهنده i پوشش داده شده اند. ایده پیشنهادی استفاده از مختصات قطبی می تواند این زمان را کاهش داده و از حجم فضاهای حفره پدید آمده در فضاهای خودی و غیر خودی نیز به شدّت بکاهد. $\sum_{i=1}^{d}C_iN_im_i$ که در آن m_i تعداد تشخیص دهنده های کاندید قابل قبولی است که در داخل تشخیص دهنده ی i قرار گرفته اند در واقع هر m_i کاندید ، فاصله خود را از N_i خودی یا غیر خودی که همگی داخل این کاندید به شعاع R واقعند بررسی می کند. ملاحظه می شود که پیچیدگی محاسباتی ایده پیشنهادی از دیدگاه تئوری بهتر از زمان لازم برای تولید تشخیص دهنده ها در حالت پیشفرض است. وجود چالش زمان بالای مورد نیاز جهت تولید تشخیص دهنده های بالغ تا حدی است که به مرور زمان که درصد بالایی از فضای غیر خودی تحت پوشش آنتی بادیهای بالغ قرار میگیرد زمان مورد نیاز برای تولید تش.د. بالغی که موقعیّت آنها تصادفاً بتواند حفره های باقی مانده در فضا را پوشش دهد بسیار بالا میرود.

این چالش زمانی را می تواند با استفاده از ایده مختصات قطبی حل نمود. استفاده از مختصات قطبی راه حل مناسبی است تا نمونه های تصادفی صد درصد و در طی زمان مورد نیاز کمتری در درون فضای خودی تولید شوند. آزمایشات به انجام رسیده در مقاله [۸] آنرا اثبات می کند. مثال زیر استراتژی این ایده را بهتر توضیح می دهد.

[1] Real Valued Negative Selection Algorithm with future training phase

فرض کنید که در فضای دو بعدی [0,1] ، دایره ای به مرکز $(X = 0.3. Y = 0.4)$ به شعاع 0.1 موجود باشد. مطلوبست تولید ۳۰ نقطه تصادفی بطوریکه تمام موقعیتهای این نقاط در فضای درون دایره به شکل تصادفی و یکنواخت تولید و محاط شوند. برای حل این مسئله از مختصات قطبی رابطه ۳-۱۶ استفاده نموده و $n = 2 . \theta_1 = 2\pi$ تعیین می شوند. ما یک شبیه سازی متلب مربوط به الگوریتمِ ایده پیشنهادی را برای n بُعد [1] انجام دادیم. شکل ۲۸ در [پ-ب-۴] خروجی این سورس کد را برای دو بُعد نشان می دهد.

از این شکل مشاهده می شود که تولید تصادفی نمونه های داده در فضای دو بعدی بازه پیوسته [0.1,0.1] بگونه ای که موقعیت تمامی این نقاط در محدوده خاص (دایره مُحاط) قرار داشته باشند مسئله سختی است که نیازمند استراتژی خاصی می باشد (روابط ۶-۸ در [۸]). این استراتژی استفاده از مختصات قطبی در دو بعد (مطابق شکل بالا) می باشد که قابل تعمیم برای مسائل با ابعاد بالا نیز هست.

۲-۸-۲-۳- پدیده تکثیر [2] و جهش تشخیص دهنده ها

مطابق مطالب مطرح شده در بخش " پدیده بلوغ وابستگی " فصل ادبیّات نظری (پیوست [پ-الف-۱۳]) بسته به میزان وابستگی تطبیق الگوی نمونه آزمون و تشخیص دهنده غیر خودی ، دو پدیده ممکن است رخ دهد که با مثالی توضیح می دهیم :

فرض کنید فضای ابعاد مسئله n بعدی بوده و Ag_i در داخل محدود تحت پوشش $(Hyper - Shape)$ تشخیص دهنده Ab_j واقع شده باشد. در این وضعیت ، هر چه فاصله ی Ag_i از مرکز محدوده ی تحت پوشش Ab_j دورتر بوده و اندازه این فاصله نزدیک به اندازه شعاع Ab_j باشد به همان میزان درجه وابستگی در تطبیق دو الگوی Ag_i و Ab_j کمتر است و در نتیجه نیاز به جهش بیشتر می باشد. [۸]

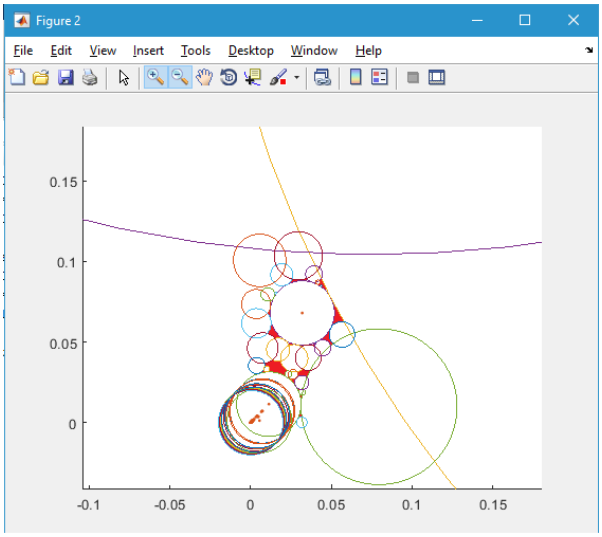
در این پژوهش ، مقصود از جهش یعنی تعدیل/ تغییر [3]بردار موقعیّت Ab_j با تولید تصادفی تشخیص دهنده ها در اطراف Ab_j . به عبارت بهتر ، هر چه Ag_i به مرکز محدوده تحت پوشش Ab_j نزدیکتر باشد به معنای وابستگی بالا در تطبیق الگوهای Ag_i و Ab_j است که در اینصورت تشخیص دهنده ی بالغ مربوطه (Ab_j) ، به حافظه ی سیستم ایمنی منتقل شده و تکثیر می شود. در این پژوهش مقصود از تکثیر Ab_j

[1] https://www.researchgate.net/publication/325252515_Generating_Uniformly_Random_Numbers_within_a_Hyper_Shape_Super-Ring_in_N_dimentional_Space_A_MATLAB_Simulation
[2] Cloning
[3] modification

بالغ موثّر حافظه [۱]، تثبیت بردار موقعیّت Ab_j و عدم تعدیل و تغییر آن علیرغم تکرار فازهای یادگیری اوّلیه در طول اجرای الگوریتم می باشد.

تثبیت بردار موقعیت Ab_j موثّر حافظه در اینجا به مفهوم تکثیر بکار رفته است. همچنین مفهوم موثّر بودن به معنای آنست که Ab_j بالغ ، حداقل یک Ag_i را در داخل $Hyper - Shape$ خود تحت پوشش قرار داده و شناسایی نموده است. به محض احراز موثّر بودن Ab_j ، این آنتی بادی موثّر و بالغ بالافاصله "حافظه" می شود. در FtNSA در مقاله [۶] با اجرای فازهای یادگیری ، رخدادهای زیر در فضای ابعاد مسئله محتمل خواهد بود :

۱- **پوشش حفره ها** – با هر بار شناسایی عوامل نفوذ ، فاز یادگیری اوّلیه فراخوانی و اجرا می شود. نتیجه اجرای فاز یادگیری معمولاً پوشش برخی حفره ها را در فضای مسئله در پی دارد امّا بطور کامل نمی توان همه این حفره ها را پوشش داد. به دلیل آنکه در لا به لای محدوده های تحت پوشش $Hyper - Shape$ تشخیص دهنده ها همواره حفره هایی وجود خواهند داشت که قابل پوشش نخواهند بود. این چالش بوسیله ی ایده پیشنهادی در بخش قبل و استفاده از استراتژیِ مختصات قطبی در تولید غیر تصادفی تشخیص دهنده های بالغ را می توان رفع نمود و به این ترتیب در زمان کمتری بیشترین فضای حفره ها را پوشش داد.



[۱] دلیل آنکه به تشخیص دهنده Ab_j بالغ گفته می شود که مشخص است چون در مرحله تولید در فضای مسئله ، ابتدا با الگوهای خودی چک شده و با آنها تطبیق نخورده بدین معنی که در محدوده فضای خودی ها جنریت نشده است. همچنین دلیل آنکه موثر گفته می شود آنست که حداقل یک تشخیص موفّق Ag_i انجام داده است که میزان وابستگی تطبیق دو الگو مشخص است. در نهایت علّت آنکه حافظه گفته می شود آنست که الگوی Ab_j به دلیل تطبیق موثر الگویی ارزشمند است که می تواند در فازهای تشخیص بعدی الگوریتم ، الگوهایی با میزان وابستگی مشابه را شناسایی و بایند کند.

شکل ۱۸ - عدم امکان پوشش کامل حفره های ظاهر شده در فضای لا به لای آنتی بادیها و همچنین در فضای مرز خودی- غیر خودی

به عنوان مثال در فضای دو بعدی که شکل محدوده های تحت پوشش آنتی بادیها دایره می باشد علیرغم تولید و تکثیر آنتی بادیها در فضای غیر خودی ، بازهم حفره هایی در ناحیه مرز میان فضای خودی – غیر خودی ظاهر خواهند شد که به دلیل دایره بودن این نواحی نمی توان به طور کامل این حفره های موجود در لا به لای دایره ها را پوشش داد ، در نتیجه حفره های لا به لای آنتی بادیها در نهایت غیر قابل پوشش کامل بوده و بنابراین خطای مثبت و منفی اشتباه همواره وجود خواهد داشت. (شکل ۱۸)

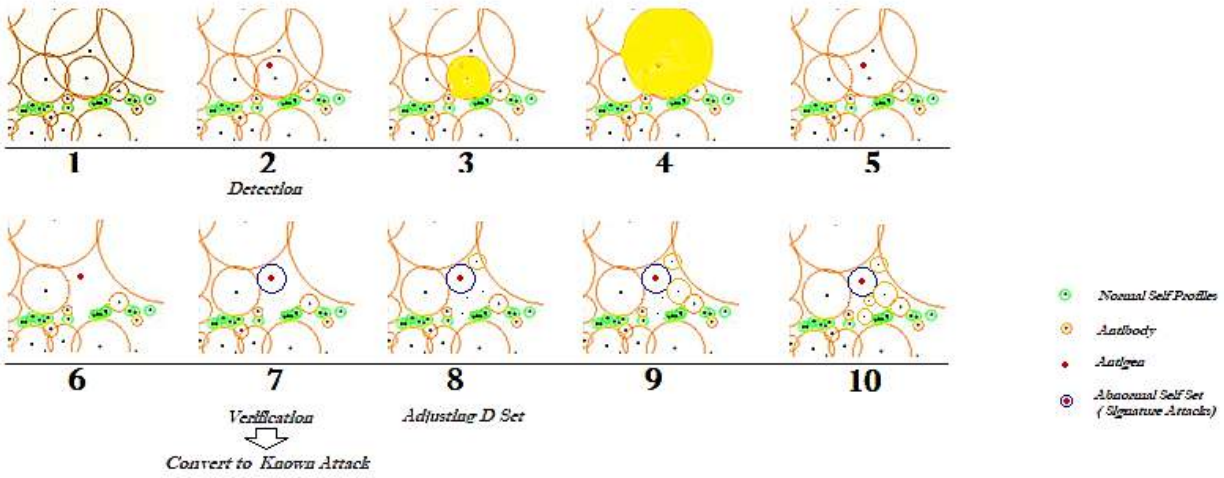
بنابراین استفاده از ایده ی مختصات قطبی در این موارد می تواند مثمِر الّثمر باشد. از طرفی وجود این دو خطای اشتباه و اینکه به هیچ عنوان نمی توان نرخ تشخیص را به صد در صد افزایش داد، دلایل و مبنای فلسفی و تجربی دارد که در بخش بعد با این دو دیدگاه بیشتر آشنا می شویم.

۲- تولید تش.د.بغ.م دقیقاً در موقعیّت بردار Ag_i ای که قبلاً Ab_j را تبدیل به تشخیص دهنده ی حافظه کرده و توسط آن شناسایی شده است. به عبارت دیگر در جریان تولید تش.د. (های) بالغ جدید ، Ab_j مربوطه به دلیل حافظه بودن آن کماکان در فضای ابعاد مسئله تثبیت می شود. همانطور که بیان گردید تثبیت Ab_j در موقعیت مربوطه و عدم امکان تولید آنتی بادیهای جدید در محدوده تحت پوشش آن ، به معنای تکثیر Ab_j است. پیش زمینه بیولوژیکی تکثیر به معنای توزیع آنتی بادیهایی با بلوغ وابستگی بالا در الگو در بافت های سلولی بدن می باشد که این خاصیت توزیع شدگی بهمراه تکثیر آنتی بادیها در واقع به معنای " برقراری ایمنی و مقاومت غیر متمرکز سیستم" در برابر آنتی ژن نفوذی دارای الگوی شناخته شده نیز هست.

بنابراین تکثیر و پخش شدن این آنتی بادیهای حافظه در بدن یک مفهوم واکنش سریع و پایدار سیستم در برابر نفوذ شناخته شده در جای جای بافت های بدن می باشد. پس نگاشت مفهوم تکثیر به مسئله تشخیص نفوذ شبکه در پژوهش جاری، تثبیت موقعیّت آنتی بادی بالغ می باشد.

۳- انتظار تا زمان تایید صحّت (Verification) برچسب پیش بینی شده ی Ag_i (امضای حمله [1]). شکل زیر مراحل این انتظار را نشان می دهد. پس از تایید صحّت – امضای Ag_i ، این عامل شناخته شده به مجموعه خودی – آنومالی (پایگاه امضاهای حملات) اضافه می شود. بدین ترتیب تشخیص دهنده (های) بالغ موثّری که آن Ag_i را پوشش و شناسایی نموده اند نیز حذف شده و فضای خالی موجود نیز با اجرای فاز یادگیری اوّلیه ، سریعاً تعدیل و بروز می گردد. با تعدیل این فضا تمام حفره های بوجود آمده/ یا از قبل موجود نیز پوشش داده می شوند.

[1] Sign

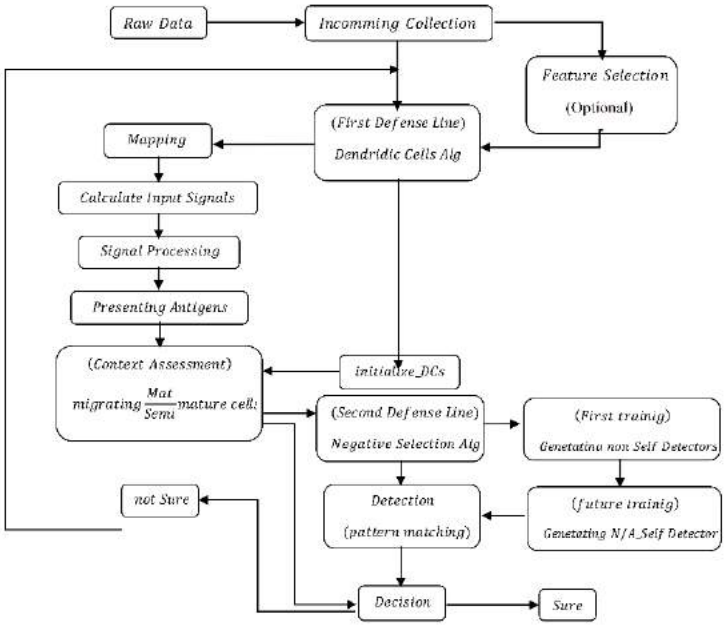


شکل ۱۹ - گامهای شناسایی/ پیش بینی الگوی نمونه آزمون و انتظار جهت تایید صحّت برچسب نمونه پیش بینی شده

۹-۲-۳- سیستم تشخیص نفوذِ هیبریدیِ پیشنهادی

روش پیشنهادی ما ترکیبِ هیبریدیِ دو مکانیسمِ ایمنی مصنوعی (تئوری خطر و انتخابِ منفی) در قالب دو خط دفاعی با الهام از دو مص.ذت و مص.تط. با شبیه سازیِ کاملِ این دو می باشد. این دو مصونیّت در برابر آنتی ژن مکمّل یکدیگر عمل می کنند. برای استراتژیِ فاز تصمیم گیری نهایی نیز قوانینی[1]را وضع نمودیم.

۱-۹-۲-۳- ساختار مدل سیستمِ تشخیصِ نفوذِ هیبریدیِ پیشنهادی



شکل ۲۰ – مدلِ مربوط به روشِ هیبریدیِ تشخیص نفوذِ پیشنهادی

[1] Rules

۱-۱-۹-۲-۳- توصیف مدل

سیستمِ هیبریدیِ پیشنهادی از سه لایه تشکیل شده و در لایه اوّل یک ماژول اصلی و یک ماژول فرعی (اختیاری) به کار رفته است. لایه های بعدی در واقع همان خطوط دفاعیِ روش پیشنهادی هستند. الگوریتم ۷ در شکل ۲۲ ، شبه کد مربوط به الگوریتمِ هیبریدیِ پیشنهادی را نشان می دهد.

۱- لایه نخست

۱-۱- **ماژول اصلی (پیش پردازش)** : ترافیک شبکه ورودی به سیستم را کَپچِر نموده و پس از آنالیز و تخصیص ویژگیها لاگ ها را بافر می کند. این بافر همان دادگانِ آزمون می باشد که برای آنالیز در فازهای بعدی مورد نیاز است.

۱-۲- **ماژول فرعی- اختیاری (متد جستجو و انتخاب ویژگی راپِر)** : در این پژوهش الگوریتم انتخاب ویژگی دَه - پا با متد شبکه عصبی تک لایه به عنوان دسته بندِ یادگیری به کار رفت که البته تاثیر منفی آن از آزمایشات بخش های قبل به خوبی به اثبات رسید. [1] البته اعمال کاهش ابعاد مسئله اختیاری بوده ولی تاثیرِ مثبتِ سایر متدهای انتخاب ویژگی در بهبود نرخ های دسته بندیِ متدهای ایمنی مصنوعی هنوز چالش هست که نیاز به بررسی بیشتر دارد. این ماژول نمونه های ترافیک شبکه را از بافر خوانده و ابعاد مسئله را با استراتژی جستجو و انتخاب ویژگی دهَ - پا کاهش می دهد. پس از کاهش ابعاد ، نمونه های موجود در دادگان آزمون به خط نخست (لایه دوّم - مکانیسمِ DCA پیشنهادی) ارسال می شوند.

۲- لایه دوّم

استراتژی دفاعی خطّ نخست دفاعی (الگوریتم سلولهای دندریت- شبیه سازی تئوری خطر)

این لایه از چهار ماژول تشکیل شده است. به نمونه های ترافیک شبکه که به این لایه می رسند مانند آنتی ژن برخوردِ مشابه زیستی می شود. بدین معنا که دقیقاً همان کاری را که سلّولهای دندریت در بافتهای سلّولی بدن انجام می دهند [2]همان وظیفه نیز توسط ماژولهای این لایه دفاعی صورت می پذیرد.

۲-۱- **ماژول نخست (زایش سلّولهای دِندریت در بافتهای بدن و نمونه برداری از آنتی ژنها [3])** : این ماژول نمونه های ترافیک شبکه در دادگان آزمون را تکثیر می کند. شیوه تکثیر پیشنهادی بدین صورت است که کل دادگان آزمون بر مبنای دو ویژگی Service – Protocol به زیر مجموعه هایی تقسیم می شوند. بدین ترتیب نمونه های با پروتکل-سرویس یکسان در یک زیر مجموعه (اصطلاحاً یک بافت سلّولی) قرار میگیرند. این عمل سرعت نمونه برداری را افزایش می دهد. سپس نمونه های ترافیکی هر بافت در فضای بافت سلّولی

[1] CFA – ANN

[2] یعنی دریافت و پردازش سیگنالها از بافت های سلّولی آسیب دیده /سالم و نمونه برداری از آنتی ژنها و تحویل آنها بهمراه سیگنالهای محاسبه شده به لنفوسیتهای دارای مص.تط.

[3] GetAntigens()

Service – Protocol خود ، به صورت تصادفی یکنواخت تکثیر[1]می شوند و یک بُردار طولانی از نمونه های تکثیر شده ساخته می شود. (در بخش جزئیات تئوری خطر بیان شد)

عمل تکثیر به معنای کپی شدن نمونه های ترافیک شبکه با الگوی مشخص و تکرار شدن آنهاست و به مفهوم Clone می باشد. در نهایت به ازاء هر نمونه از ترافیک شبکه یک سلّول دندریتِ متناظر ایجاد میشود که به شکل تصادفی اقدام به نمونه برداری از آنتی ژنهای اطراف می نماید. ما ایده نمونه برداری پیشنهادی را به شکل الگوریتم زیر بیان و آنرا شبیه سازی نمودیم :

Algorithm . 6 – Psudo code of proposed strategy for Multiplying of Antigens to form an antigen$_{\text{Vector}}$

/* initialization phase */
* At first. test_set are devided to cell tissues */

$\mathbf{tissues} = \mathbf{tissue}(\mathbf{test_{set}}\,.\,\mathbf{service} - \mathbf{protocol\ features})$ */ create tissues

/* Antigens Uniformely random multiply to form an antigen Vector */

foreach tissues

$$\mathbf{AgVector(i)} = [\mathbf{ceil}(\mathbf{minimumCloneRange}.\mathbf{maximumCloneRange}) \times \mathbf{size_{of}tissues(i)}] + \mathbf{size_{of}tissues(i)}$$

$$\mathbf{Vector(i)} = \mathbf{Random\_Dist}(\mathbf{size_{of}tissues(i)}.\mathbf{AgVector(i)})$$

/* randomely distribute AgVector(i) in size$_{of}$tissues(i)

End for

شکل ۲۱ – شبه کد مربوط به الگوریتمِ پیشنهادی برای تکثیر آنتی ژنها جهت تشکیل بردار آنتی ژن

۲-۲- **ماژولِ تخصیص سیگنالها** : ضمن اینکه مکانیسمِ سلولهای دندریت اقدام به نمونه برداری از ترافیک مشکوک می نماید ، سیگنالهای ورودی نیز محاسبه شده و در محاسبه سیگنالهای خروجی مورد استفاده قرار میگیرند.

۳-۲- **ماژول محاسبه ی وضعیت مهاجرتِ آنتی ژنها** : در این مرحله ، پس از آنکه ترافیک های مشکوک نمونه برداری و سیگنالهای خروجیِ سلّولها نیز تخصیص و محاسبه شدند نوبت به ارزیابی وضعیّت برچسب نهایی هر ترافیک مشکوک فرا میرسد. خروجی این مرحله در نهایت ترافیک های مشکوک (آنتی ژنها) پس از ارزیابی نهایی به یکی از دو زیر مجموعه ی بالغ کننده یا نیمه بالغ کننده مهاجرت می یابند.

۳- لایه سوّم

[1] Multiply

استراتژی خط دوّم دفاعی (مکانیسمِ انتخاب منفی حقیقی مبنا بهبود یافته)

۱-۳- **ماژول نخست** : فاز یادگیریِ اوّلیه و تولید و تکثیر تشخیص دهنده های غیر خودی جهت پوشش فضای غیر خودی و حُفره ها.

۲-۳- **ماژول دوّم** : فاز یادگیری ثانویه ، همروندِ با فاز نخست می تواند اجرا شده و نمونه های فضاهای خودی را تعدیل کند. در نتیجه از تعداد نمونه های خودیِ اضافی در پروفایلهای سیستم کاسته می شود.

۳-۳- **ماژولِ آزمون انتخاب منفی** : در این مرحله نمونه ی ترافیک مشکوک بسته به موقعیت قرارگیری در فضای ابعاد مسئله (خودی ، غیر خودی و یا حفره) و اینکه تحت پوشش تشخیص دهنده ی بالغ حافظه قرار گرفته باشد یا بالغِ غیر حافظه ، در نهایت برچسب نهایی آن با استراتژیِ پیشنهادی که در بخش قبل ارائه گردید مشخص می شود. البته شایان ذکر است که در تعیین برچسب احتمالی نهایی ، برچسب احتمالیِ اولیه که در خط نخست دفاعی به نمونه مربوطه تخصیص یافته نیز در ارزیابی احتمال برچسب نهایی تاثیر گذار است. بطوریکه ممکن است نتیجه ی ارزیابی دو خط دفاعی ناسازگار با هم باشد که در اینصورت نمونه های با وضعیت عدم قطعیت یا ناسازگار مجدداً در سیکل بعدی اجرای این دو خط طی یک حیات مصنوعی وارد پروسه ی ارزیابی می شوند. البته حدآستانه قابل تعریف برای حداکثر تعداد سیکلهای مجاز باید بر اساس حدآستانه ی قابل تحمل برای تاخیر زمانی پردازش ترافیکِ ورودی شبکه در نظر گرفته شود که بیشتر به تجربه کاربر یا مدیر شبکه وابسته بوده و مقداری نسبی می باشد.

Algorithm . 7 – Psudo code of proposed Immune Inspired NIDS

Inputs: $\text{Antigen}_{\text{set}}$,$\text{Train}_{\text{set}}$. $\text{Weight}_{\text{Matrix}}$. $\text{minSize}_{\text{of}}\text{CloneAg}$. $\text{maxSize}_{\text{of}}\text{CloneAg}$, Tdmax, Tmax, m

Outputs : Probabilities of assigned Class Labels

Loop cycles

/* preprocessing phase include feature selection

$\text{Features}_{\text{Subset}}$ is determined by a feasible feature selection method. for e. q CFA. (Optional phase)

Segregate Train Set to two subsets: Normal Self and Abnormal Self, are known by abbreviations of NSD and ASD respectively

/* Initial Defend Line

MCAV Values are computed by the Proposed DCA mechanism. (Alg.2)

/* Second Defend Line

/* Two Loops must be run concurrently, Loop1 has two phases of training and also it is run test phase in Loop2

Loop 1:

The detector set is null at first cycle

/* AVote parameter is determined for all of new generated mature detectors via initial training stage

[Detectors, Radiuses, AVotes] = RNSAInitialTrainingPhase (NSD, ASD, detectors, radiuses, normal self-radius, abnormal self-radius, Tdmax) **/* Using Eq.9 strategies at Table.2, and proposed hierarchy of the polar-coordinates provided based on Eq.8 for regulating non-self-detectors */**

/* Future training stage is run so as to regulate two training subsets */

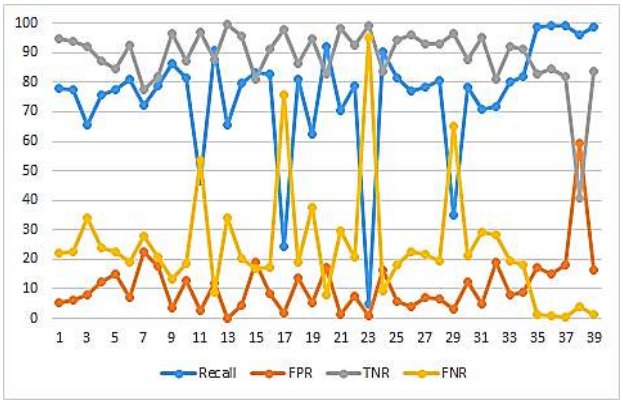
[NSD, ASD] = RNSAFuthureTrainingPhase (NSD, ASD, detectors, radiuses, normal self-radius, abnormal self-radius, Tmax, m**) /* Using proposed hierarchy of the polar-coordinates provided based on Eq.8 for regulating self-region */**

End Loop 1
Loop 2:
If Feedback is not null
PoppedAgs = Pop antigens of Feedback
End if
Add PoppedAgs to AG Set
Add some of new antigens from $antigen_{set}$ to AG_{set} so as to create a fixed size of AG_{set}
[Probabilities of assigned class labels, AVotes, detectors, radiuses, Feedback] = RNSATest (NSD, ASD, AG set, detectors, radiuses, normal self-radius, abnormal self-radius, AVotes, MCAV Values) **/* Using Eq.9**
Buffering not Sure Antigens to Feedback Base
}
End Loop 2
End cycles

شکل ۲۲ - شبه کد مربوط به روش تشخیص نفوذِ هیبریدیِ پیشنهادی با دو خط دفاعی

۱۰-۲-۳- حیات مصنوعی [1]

پس از تعیین شعاع خودی مناسب که با آزمایشات بخش ۳-۲-۳ بدست آمد ($rA = 0.1 \text{ ، } rS = 0.01$) ، در این بخش سیستم تشخیص نفوذ پیشنهادی را آزمایش نمودیم. دادگان آزمون ابتدا به زیر مجموعه های آزمون در اندازه های مشخص تقسیم شده (۲۱۱۰ رکورد) و سپس توسط سیستم پیشنهادی سیکلهای مختلف اجرا را تجربه می کنند. روند اجرای متوالی سیکلها را اصطلاحاً " حیات مصنوعی" می نامیم. روندی که در آن دادگان آزمون با اندازه ثابت و یکسان به ترتیب وارد سیستم می شوند تا سیکلهای حیات مصنوعی را تجربه کنند. دلیل این اصطلاح آنست که آنتی ژنها که در اینجا در اصل همان ترافیک شبکه ما هستند پس از ورود به سیستم تشخیص ، نوعی حیات مصنوعی را در خطوط دفاعی تجربه نموده و در پروسه ی تعیین برچسب احتمالی ، هر یک سرنوشت متفاوتی خواهند داشت. جزئیّات دادگان آزمون استفاده شده و اندازه های آنها به تفکیک در [پ-ب-۲] آمده است.



نمودار ۱۰ - مقایسه عملکردِ دسته بندهای جدولِ ۲۳ در [پ-ب-۴]

[1] Artificial Life

در ارزیابی مقایسه ای فوق حاصل از جدول ۲۳ در [پ-ب-۴]، تمام نتایج بدست آمده از دسته بندی متدها میانگین ده بار اجرای متوالی با بهینه ترین پیکربندی پارامترهای ورودی می باشند که با استفاده از تولباکس Classification Learner به دقّت آزمایش شده اند. با بررسی این نمودار و اطلاعات مربوط به خروجی های بدست آمده این نتیجه حاصل می شود که دو مکانیسمِ ایمنی مصنوعی به کار رفته در خطوط دفاعی (یعنی DCA و RNSA) که به ترتیب متدهای ۳۵ تا ۳۹ جدول را تشکیل می دهند از نرخ خطای منفی کاذب بسیار پایین تری برخوردار بوده و در عوض نرخ خطای مثبت آنها نسبت به سایر متدهای مرسوم کمی بالاتر است. از نمودار تداخلِ منحنی های TNR و FPR مربوط به متد ۳۸ ام یعنی DCA پیشنهادی را می توان مشاهده کرد.

از نمودار قابل استنباط است که متدِ ۳۶ که مربوط به سیستم هیبریدیِ پیشنهادی می باشد پس از طی پنج سیکل حیات مصنوعی و در وضعیتی که متد تصمیم گیریِ آن با پارامترهای یکسانِ $\alpha = \beta = 0.5$ تنظیم شده اند توانسته است بهترین عملکرد را در مقایسه با سایر دسته بندها داشته باشد. خطای مثبت کاذب بالا ناشی از دو عامل است که یکی وجود حفره هاست و دیگری چالش مرز بندی صحیح بین فضای خودی-نرمال با غیر خودی و خودی-آنومالی با غیر خودی می باشد. البته نرخ تشخیص الگوریتم های ایمنی مصنوعی و خصوصاً روش هیبریدی پیشنهادی و همچنین دقّت آنها بهتر از سایر متدهاست.

از طرفی نرخ کامل بودن (Completeness) یا حساسیّت متدهای ایمنی مصنوعی و سیستم پیشنهادی در مقایسه با الگوریتم های دیگر بسیار بهتر و بالای ۹۹ درصد می باشد درحالیکه برای سایر متدها این نرخ در حدود ۷۵ تا ۹۰ درصد می باشد. در مقام مقایسه ی سیستم پیشنهادی با هر دو متد ایمنی مصنوعی ملاحظه می شود که نرخ خطای آن کمتر بوده و در حدود ۲,۶۵ میباشد درحالیکه میزان خطا برای الگوریتم سلولهای دندریت پیشنهادی بالا بوده و در حدود ۲۱,۳ درصد است.

مهمترین مزیّت روش هیبریدی پیشنهادی افزایش صحّت ، کاهش خطا و همچنین افزایش نرخ دسته بندی و کاهش خطای منفی اشتباه می باشد. در نتیجه می توان این گونه نتیجه گیری کرد که در مجموع عملکرد روش هیبریدی پیشنهادی بخصوص با گذشت سیکلهای متوالی (درصورت رعایت حدآستانه سیکلهای اجرا که متناسب با حداکثر تاخیر قابل قبول در پردازش ترافیک ورودی شبکه تعیین می گردد) افزایش و اندکی بهبود می یابد. میزان این بهبود بیشتر بستگی به فواصل تعدیل فضاهای خودی-غیر خودی و بروز رسانی پایگاه امضاءها دارد. شکل ۳۲ در [پ-ب-۴] معماری سیستم تشخیص نفوذ هیبریدی پیشنهادی را نشان می دهد.

این به این دلیل است که در طول اجرای سیکلها ، ضمن تعدیل فضای غیر خودی از حجم تشخیص دهنده ها نسبتاً کاسته شده و فازهای یادگیری نیز شروع به تعدیل فضای های خودی می نمایند. بخصوص اگر در حین اجرای سیکلها سیستم تشخیص دو پایگاه پروفایلها و امضاء های حملات را بروز رسانی نماید در نهایت به مرور منجر به تعدیل فضای غیر خودی و بهبود عملکرد دسته بندی سیستم می گردد که این فرایند را اصطلاحاً " رُشد" سیستم می نامیم.

یک دلیل دیگر بهبود کیفیت دسته بندی با پیشرفت اجرای سیکلهای متوالی آنست که این روند منجر به تغییر پارامتر AVote تشخیص دهنده های بالغ شده و پتانسیل بالقوّه ی آنها در شناسایی حملات / ترافیک نرمال نیز دچار تغییرات محسوسی می شود.

البته این پروسه کاملاً متاثّر از میزان تشخیص صحیح / اشتباه سیستم در سیکلهای اجرای قبل می باشد. بطوریکه نرخ تشخیص بالا و خطای پایین در سیکل اجرای n ام ، در عملکرد دسته بندی ترافیک های جدید در سیکلهای بعدی تاثیر مستقیم می گذارد و دلیل آن همانطور که بیان شد تغییر صحیح یا اشتباه پارامتر AVote مربوط به تشخیص دهنده هایی است که حافظه شده اند مستعد شناسایی نفوذ هستند. هرچند وجود خط دفاعی نخست که اغلب دارای خطای مثبت اشتباه و نرخ تشخیص بالاتری است منجر به این می گردد که عدم قطعیّت سیستم هیبریدی پیشنهادی بالا رود. با افزایش میزان عدم قطعیّت دادگان ترافیک شبکه نیاز به اجرای سیکلهای بیشتری خواهند داشت. در نتیجه برای آنکه سیستم بیشتر بتواند از حجم داده های با وضعیت عدم قطعیّت بکاهد ، میبایست مقدار حافظه ی پردازشی را ارتقا داده تا حدآستانه سیکل های اجرا متناسب با حداکثر تاخیر قابل قبول افزایش یابد.

تجربه بدست آمده از انجام آزمایشات مختلف با الگوریتم های پیشنهادی خطوط دفاعی در بخش های قبل نشان میدهند میزان صحّت و درستی برچسبهای اختصاص داده شده با توجه به حالات ممکن برچسب زنی نمونه ها از قاعده بدست آمده از نتیجه تحلیل و ارزیابی نظری در بخش ۱-۳-۵ و جدول ۱۱ در [پ-الف-۱۴] پیروی می کنند.

۱۱-۲-۳- انجام آزمایش نهایی و ارزیابی نتایج بدست آمده

تا به اینجا دادگان نخستِ آزمون با الگوریتم های دسته بندی مختلفی از جمله روش هیبریدی پیشنهادی و دو متد ایمنی مصنوعی استفاده شده در خطوط دفاعی مورد بررسی و آنالیز قرار گرفت. این بخش به آنالیز کامل سه دادگان آزمون استفاده شده (جزئیات آنها در [پ-ب-۲]) در سیستم پیشنهادی اختصاص دارد. اینفوگرافیکِ شکل ۲۹ در [پ-ب-۴] نمایانگر روند اجرای حیات مصنوعی را نشان می دهد. مشاهده می شود که دادگان آزمون که در رنگ های مختلفی مشخص شده اند همه آنها اندازه ثابتی داشته و به ترتیب وارد سیستم می شوند. در اینجا نکته ای که مطرح است وجود درصدی از نمونه های آزمون با وضعیت عدم قطعیّت در تمامی سیکلهای اجرای سیستم می باشد بطوریکه در هر سیکل اجرا ، ما تعداد مشخصی از این نمونه ها را به تفکیک از تمامی دادگان آزمون خواهیم داشت.

واضح است که پس از اتمام پردازش هر سیکل اجرا ، همواره درصدی از حجم نمونه های با وضعیت عدم قطعیّتِ موجود کاسته شده و سیستم ظرفیّت آنرا خواهد داشت که در سیکل اجرای بعدی تعداد بیشتری نمونه

جدیدِ باقی مانده از دادگان آزمون های قبل را بافر نموده و بهمراه نمونه های عدم قطعیّت باقیمانده / جدید مورد آنالیز قرار دهد. علاوه بر این در مورد نمونه های عدم قطعیّتی که حداکثر سیکلها در موردشان اجرا شده و سیستم دیگر فرصت زمانی کافی جهت پردازش مجدّد وضعیت آنها را در سیکلهای بعدی ندارد ، ماژول تصمیم گیری نهایی پیشنهادی با استفاده از رابطه ی وزنی ۳-۱۲ در مورد آنها به اجرا در آمده و برچسب احتمالی نهایی آنها را نهایتاً ماژول تصمیم گیری نهایی محاسبه خواهد نمود. جزئیات بیشتر نحوه بافرینگ نمونه ها و سازوکار پردازش وضعیّت نمونه های با وضعیت عدم قطعیّت را در اینفوگرافیکِ شکل ۳۰ در [پ-ب-۴] ترسیم نموده ایم.

۱-۱۱-۲-۳- انتخاب استراتژی مناسب برای تخصیص سیگنالهای ورودی DCA

علاوه بر این در آزمایشاتی که در جدول زیر انجام شد، کاربرد دو متد بهره اطلاعات و دانش متخصصان در نگاشت زیر مجموعه ی ویژگیهای انتخاب شده و تخصیص آنها به سیگنالهای ورودی مورد بررسی و ارزیابی مقایسه ای قرار گرفت. اندازه استخر تشخیص دهنده های بالغ حاصل از اجرای سیکل دوّم ۴۶۹۹ می باشد که در هفت آزمایش مستقل زیر همین میزان آنتی بادی برای دسته بندی به کار رفت. مطابق اطلاعات جدول از نتایج آزمایش چنان بر می آید که متد بهره اطلاعات عملکرد دسته بندی به مراتب بهتری دارد.

جدول ۱۹ - بررسی تاثیر استفاده از تجربه و دانش متخصّصان امنیّت در Proposed DCA در سیستم هیبریدی پیشنهادی

Classification Performances. On Sure Subset						$\alpha.\beta$	DCA 's Signal Generation Method
Recall	FPR	TNR	FNR	Acc	Err		
95.2	9.4	90.6	4.8	93.7	6.3	0.5, 0.5	IG
98.2	15	85	1.8	94.1	5.9	0.2,0.8	
92.7	13	87	7.3	90.9	9.1	0.8,0.2	
98	14.5	85.5	2	94.1	5.9	0.25,0.75	
97.7	14.1	85.9	2.3	94	6	0.3,0.7	
97.2	12.4	87.6	2.8	94.2	5.8	0.35,0.65	
96.8	11.4	88.6	3.2	94.2	5.8	0.4,0.6	
96.54285714	12.83	87.17	3.457	93.6	6.4	AVG	
78.8	7.1	92.9	21.2	83.2	16.8	0.5, 0.5	Expert Knowledge
95.9	14.8	85.2	4.1	92.6	7.4	0.2,0.8	
71.4	9.7	90.3	28.6	77.3	22.7	0.8,0.2	
94.2	14.2	85.8	5.8	91.6	8.4	0.25,0.75	
92	12.3	87.7	8	90.7	9.3	0.3,0.7	
89.4	10.6	89.4	10.6	89.4	10.6	0.35,0.65	
86.5	8.3	91.7	13.5	88.1	11.9	0.4,0.6	
86.88571429	11	89	13.11	87.6	12.4	AVG	

از طرفی استفاده از روش دانش متخصصان برای تخصیص سیگنالها در الگوریتم خط نخست دفاعی میزان تولید تشخیص دهنده های بالغ را کمتر افزایش میدهد و این میزان در این آزمایش ۵۵۸۷ آنتی بادی بوده است ولی در صورتی که متد بهره اطلاعات استفاده شود میزان تولید آنتی بادیها پس از انجام آزمون ها ۶۰۳۳ است.

همچنین در مورد تعداد نمونه های با وضعیت عدم قطعیّت نیز وضع به همین صورت است. با کاربرد متد بهره اطلاعات، میزان نمونه های با وضعیّت قطعیت $TP_1 - TP_2$ بالا بوده و این میزان در آزمون اوّل برابر با ۱۶۷۵ است در حالیکه در صورت استفاده از روش تجربه متخصصان این میزان به ۱۲۱۶ نمونه مطابق آزمون هشتم در جدول فوق کاهش محسوسی می یابد.

مشاهده می شود که استفاده از این روش تجربی (دانش متخصّصان) در تولید سیگنالها ، علیرغم مزیّتهای آن که بالا بودن نرخ تشخیص مثبت و منفی صحیح می باشد چالش مهم افزایش عدم قطعیت نمونه ها را نیز با خود بهمراه دارد که سیکلهای اجرا را در مقایسه با متد دیگر طولانی مینماید. ضمن اینکه بنا به تجربِه ی حاصل از ارزیابی انجام شده در اینفوگرافیکِ شکل ۲۹ از [پ-ب-۴] که برای آزمون به انجام تحلیل تئوری در زمینه ی تاثیر کاربرد این دو متد جداگانه در تخصیص سیگنالهای الگوریتم سلولهای دندریت پرداختیم. نتایج این تحلیل در جداول شکل ۳۱ در [پ-ب-۴] آمده است.

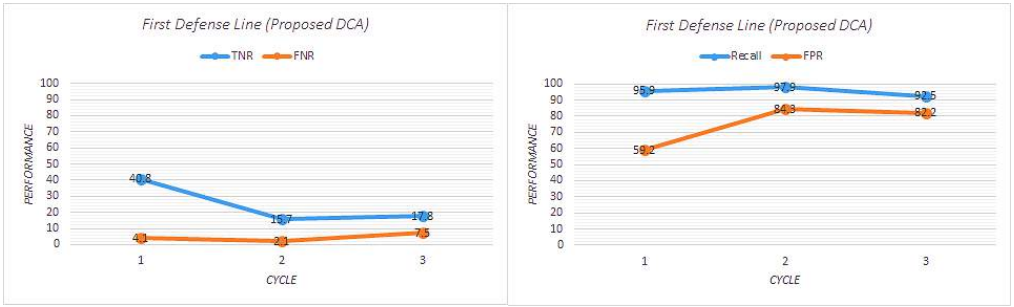
مطابق این تحلیل ، اگر از متد بهره اطلاعات استفاده گردد به دلیل بالا رفتن نرخ تشخیص و نرخ خطای مثبت کاذب در خط نخست، احتمال وقوع عدم قطعیّت با وضعیّت $Mat - MD - TN$ بیشتر از سایر وضعیّتهای عدم قطعیّت و درصورتیکه سیستم دسته بند به تازگی اقدام به تولید و تکثیر آنتی بادیها نمونه باشد تجربِه ی لازم را نداشته و میزان تشخیص دهنه های حافظه ی آن پایین است. بنابراین به جای وضعیّت فوق ، $Mat - D - TN$ رخ می دهد و این سیستم را به عدم قطعیّت می برد. امّا در صورت استفاده از دانش متخصّصان احتمال وقوع هر دو وضعیّت قطعیت $TN_1 - TN_2$ و $TP_1 - TP_2$ بالا رفته امّا در عوض میزان وقوع دو وضعیّت قطعیّت $FN_1 - FN_2$ و $FP_1 - FP_2$ بسیار پایین خواهد بود.

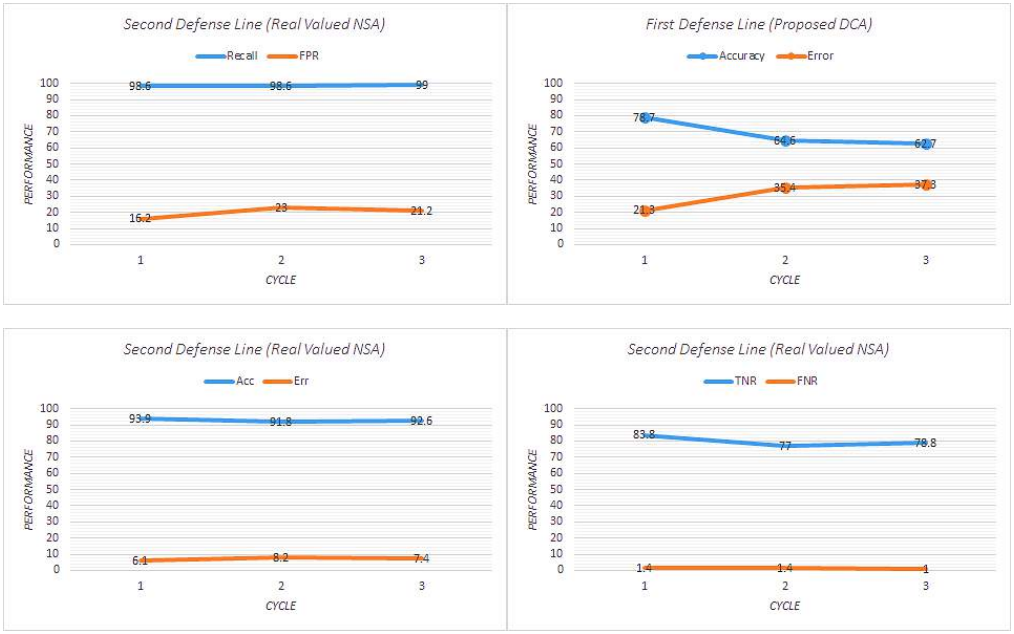
همچنین در بهره اطلاعات ، احتمال وقوع $TN_1 - TN_2$ به دلیل پایین بودن نرخ منفی صحیح در خط نخست در مقایسه با حالت دانش متخصصان کم می باشد. تحلیل تجربِی این نتایج را در محیط شبیه ساز در آزمون نهایی نیز بررسی نموده ایم. به عبارت دیگر ، در هر سیکل در حالتی که از متد دانش متخصّصان در تخصیص سیگنالها در خط نخست استفاده گردد درصد زیادی از نمونه های قطعیّت واقعاً صحیح هستند و این نکته مثبتی است (TN و TP). نتایج این ارزیابی نشان می دهند که مهمترین چالش استفاده از متد بهره اطلاعات بالا رفتن نرخ خطای مثبت کاذب بوده که این وضعیت منجر به وقوع بالای ردیف های سوّم و چهارّم (عدم قطعیّت و قطعیّت) در جدول شکل ۳۱ دارد. در مجموع انتظار می رود در صورت استفاده از دانش متخصصان در خط نخست دفاعی ، سیستم سیکلهای طولانی اجرا را تجربِه نماید که البته چالش بزرگی است. در نقطه مقابل در صورت استفاده از متد بهره اطلاعات میبایست صرفاً بر روی مسئله ی چگونگی کاهش میزان نرخ خطای مثبت خط نخست بیشتر کار شود. استفاده از دانش متخصصان آنگونه که در بخش ۳-۲-۶ مطرح شد همواره با چالش کشف زیر مجموعه ویژگی (های) مناسب جهت تخصیص به سیگنالهای ورودی روبه رو است. چالشی که تاکنون توسط محقّقان راه حل مناسبی برای آن ارائه نشده است. البته به نظر میرسد استفاده از

استراتژی جستجوی بهینه ترین زیر مجموعه ویژگیها با کاربرد متدهای کاهش ابعاد می تواند راه حل مناسبی برای رفع این چالش باشد. ولی نکته مهم آنست که به دلیل آنکه هک و نفوذ خصوصاً الگوی ترافیک شبکه پیوسته در حال تغییر و تحوّل بوده و حملات جدید با سوء استفاده از آسیب پذیری های روز صفر پیوسته تولید میشوند لذا دانش متخصصان این حوزه نیز مستلزم بروز شدن و شناخت دقیق الگوی ترافیک مشکوک است. در نتیجه بحث شناسایی الگو با استفاده از متدهای یادگیری ماشین در این رابطه مطرح می شود که می تواند متخصصان را در آنالیز ترافیک شبکه و کشف متناسب ترین زیر مجموعه ویژگیها و نگاشت آنها به سیگنالهای ورودی یاری دهد. در آزمایش نهایی ، سه دادگان آزمون در اندازه ثابت به ترتیبی که در اینفوگرافیکِ اشکالِ ۲۹ ارائه شده وارد خط نخست دفاعی سیستم پیشنهادی می شوند. نتایج این آزمایش به صورت زیر می باشد. حدآستانه سیکلهای اجرا را متناسب با حداکثر زمان تاخیر قابل قبول در پردازش نمونه های آزمون ، مقدار ۳ سیکل تعیین نمودیم.

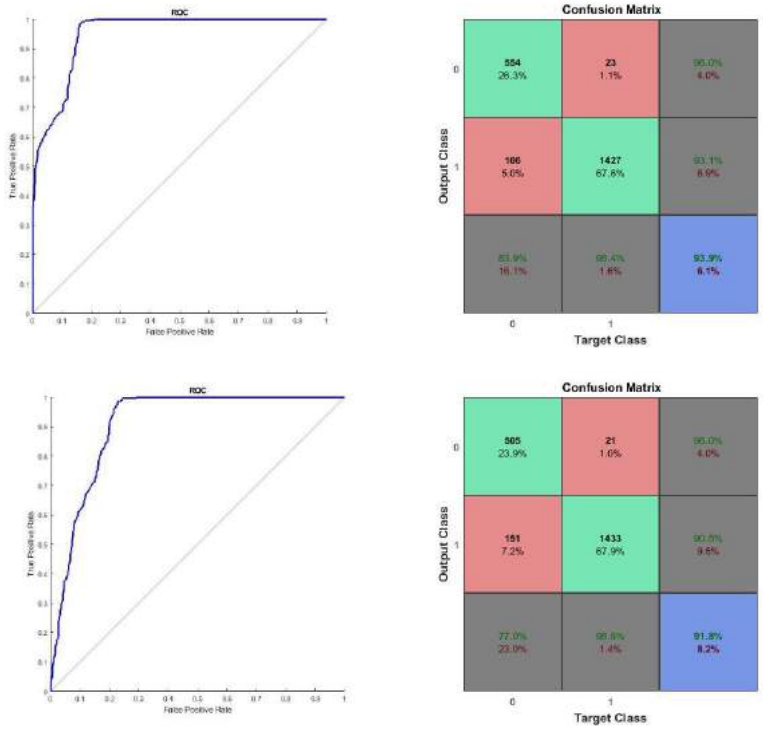
بدین معنی که روش هیبریدی پیشنهادی نمونه های با وضعیت عدم قطعیّت را حداکثر تا سه سیکل مورد پردازش مجدّد قرار خواهد داد و در صورتی که پس از سیکل سوّم ، سیستم در مورد وضعیت برچسب این نمونه ها بازهم وضعیت قطعی و اتّفاق نظر ارزیابی خطوط دفاعی را ارائه ندهد ماژول تصمیم گیری نهایی وضعیت احتمالی این نمونه ها را با استفاده از رابطه ی وزنی ۳-۱۲ مشخص می کند. در این آزمایش مقادیر دو پارامتر وزنی $\alpha . \beta$ در این ماژول را ۰,۵ تعیین نمودیم. تعیین مقادیر پارامترهای وزنی کار مشکلی است زیرا این وزنها ارتباط نزدیکی به متوسط کیفیت دسته بندی خطوط دفاعی دارند.

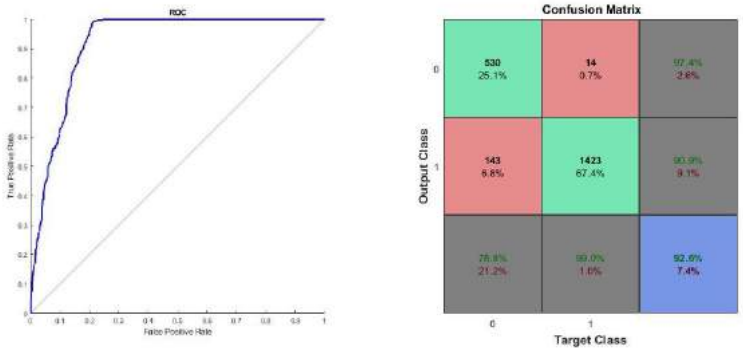
بطور مثال اگر کیفیت دسته بندی خط نخست (فیتنس) در مقایسه با عملکرد الگوریتم خط دوّم دفاعی خیلی کمتر باشد در نتیجه فاصله ای مقادیر این وزنها از هم بیشتر شده و $\alpha \ll \beta$ می گردد ولی مجموع آنها همواره برابر یک میباشد. ($\alpha + \beta = 1$) نتیجه این آزمایش در مقایسه با حالتی که این سه دادگان آزمون با استفاده از هر یک از خطوط دفاعی بطور مستقل مورد آنالیز و دسته بندی قرار گیرند ارزیابی شده است.





نمودار ۱۱ - نتایج آزمایش مستقلّ سه دادگان آزمون در خط نخست (Proposed DCA) و خط دوّم (FtRNSA)



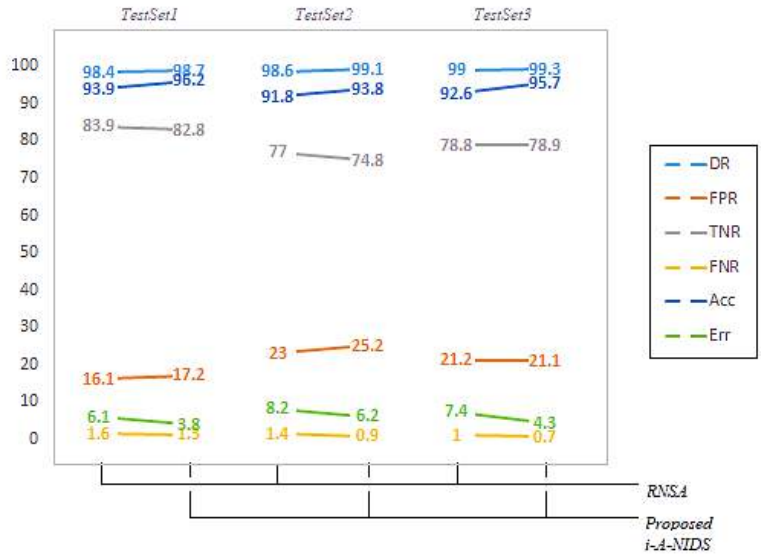


نمودار ۱۲ – ماتریس درهم ریختگی و منحنیِ ROC مربوط به سه آزمایش مستقل سه دادگان ترافیک شبکه در FtRNSA های خودی به ترتیب برابر $r_A = 0.1$ و $r_S = 0.01$ (به تفکیک از بالا دادگان آزمون اوّل تا سوّم)

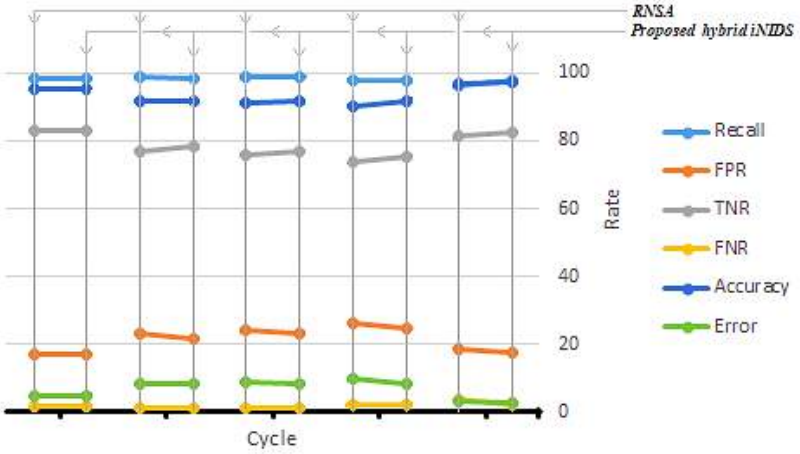
جدول ۲۰ - نتیجه اجرای پنج سیکل های متوالی حیات مصنوعی در مورد سه دادگان آزمون [۱]
(برای تخصیص سیگنالها در خط دفاعی نخست از متد بهره اطلاعات استفاده شد)

Cycle	Antibody Pool Size (Before)	Antibody Pool Size (After)	Runtime	Sure Size		Buffer Size (not-Sure Records)		Recall	FPR	TNR	FNR	Accuracy	Error	Size of Samples
1	29	2397	23.498482	Sure1	1728	not-Sure1	382	98.5	17.2	82.8	1.5	95.4	4.6	1728
	FtRNSA							98.5	17.2	82.8	1.5	95.4	4.6	
2	2397	4699	37.062048	Sure2	1640	not-Sure2	88	98.6	21.7	78.3	1.4	91.7	8.3	1865
				Sure1	225	not-Sure1	157							
	FtRNSA							98.8	22.9	77.1	1.2	91.8	8.2	
3	4699	6827	81.374148	Sure3	1370	not-Sure3	113	99.1	23	77	0.9	91.8	8.2	1761
				Sure2	313	not-Sure2	157							
				Sure1	78	not-Sure1	79							
	FtRNSA							99.1	24.1	75.9	0.9	91.3	8.7	
Total DataSet1				Total Sure Size	2031	Total not-Sure Size	79	98.7	17.2	82.8	1.3	96.2	3.8	2110
Total DataSet2					1953		157	99.1	25.2	74.8	0.9	93.8	6.2	
Total DataSet3					1370		740	99.3	21.1	78.9	0.7	95.7	4.3	
4	6827	8129	100.5629	Sure4	989	not-Sure4	224	97.8	24.6	75.4	2.2	91.9	8.1	1722
				Sure3	614	not-Sure3	126							
				Sure2	119	not-Sure2	38							
	FtRNSA							97.6	26.1	73.9	2.4	90	10	
5	8129	7653	97.3456	Sure5	625	not-Sure5	159	97.7	17.6	82.4	2.3	97.1	2.9	1737
				Sure4	1005	not-Sure4	195							
				Sure3	107	not-Sure3	19							
	FtRNSA							96.4	18.5	81.5	3.6	96.6	3.4	

[۱] با توجه به اینکه در این آزمایش حدآستانه تعداد سیکلهای اجرای قابل قبول برابر ۳ تعیین شده در نتیجه پس از سه سیکل اجرای متوالی ، سیستم در مورد نمونه های عدم قطعیّت که به ترتیب ۷۹ ، ۱۵۷ و ۷۴۰ نمونه ترافیک شبکه باقی مانده از سه دادگان آزمون ما را تشکیل می دهند تصمیم میگیرد.



نمودار ۱۳ – مقایسه ی عملکرد تشخیصِ روش پیشنهادی در پایانِ سیکل سوّم حیات در مقایسه با FtRNSA



نمودار ۱۴- عملکرد دسته بندی زیر مجموعه های قطعیّت در سیکلهای اجرای متوالی حیات مصنوعی[۱]
روش پیشنهادی در مقایسه با FtRNSA

در نمودار بالا ، نرخ های دسته بندی نمونه های قطعیّت در هر سیکل در مقایسه با نتایج دسته بندی همین نمونه های در FtRNSA ارزیابی شده اند. همانگونه که از نمودار فوق قابل استنباط است در سیکل نخست به دلیل آنکه سیستم هنوز به آن حدّ بلوغ خود در تولید تشخیص دهنده های موثّر حافظه نرسیده ، میزان نرخ های دسته بندی و خطا در مقایسه با دسته بندیِ FtRNSA بوده و بهبودی مشاهده نمی گردد. امّا به مرور

[۱] دقّت شود که اندازه ی دادگان مورد آزمایش در سیکلها متفاوت می باشند.

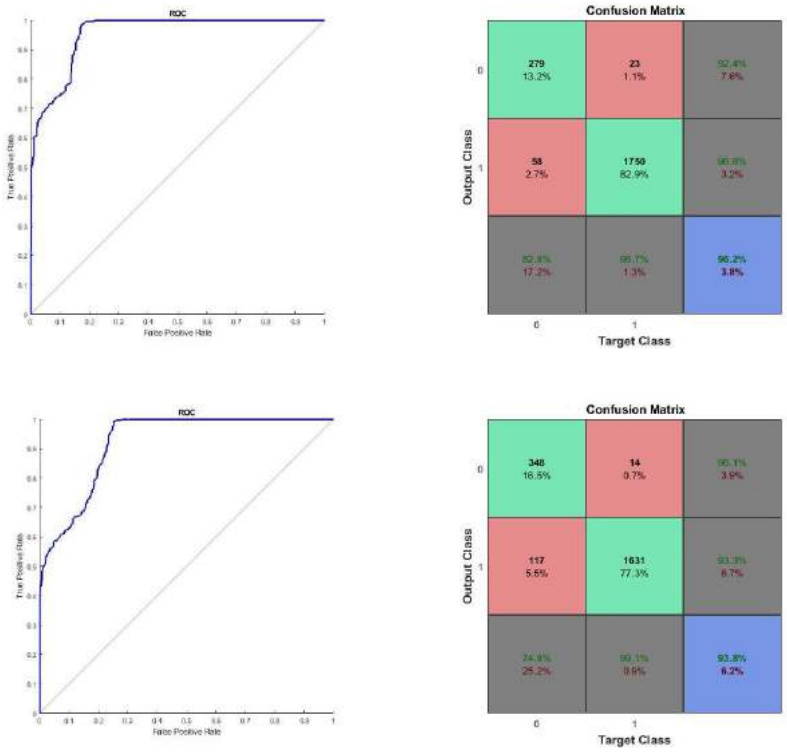
زمان و اجرای سیکلهای متوالی این تفاوت محسوس را در همه نرخ های دسته بندی مشاهده می کنیم. این نشان می دهد که کاربرد الگوریتم سلولهای دندریت به عنوان یک استراتژی پیشگیرانه از نفوذ در خط نخست دفاعی روش پیشنهادی موثّر بوده است. ترکیب الگوریتم سلولهای دندریت در روش هیبریدی پیشنهادی تاثیر خود را در افزایش دقّت در شناسایی و کاهش خطا و بویژه در افزایش نرخ شناسایی ترافیک نرمال و کاهش خطای مثبت کاذب نشان می دهد و تاثیر چندانی در افزایش نرخ شناسایی حملات نمی گذارد.

این نکته ای است قابل توجه چراکه محققان این حوزه همواره به دنبال این بوده اند که نرخ مثبت اشتباه کاذب الگوریتم های انتخاب منفی را در متدهای ایمنی مصنوعی پیشنهادی خود ارتقا دهند. زیرا همانگونه که در فصول گذشته اشاره گردید و به عنوان یک پرسش و فرضیه در این پایان نامه مطرح گردیده است ، تاثیر کاربرد تئوری خطر در پیشگیری از نفوذ و کاهش خطا بود. تجربه ی حاصل از انجام چنین آزمایشی با این وسعت در شبیه ساز متلب تجربه بسیار خوبی بود و امکان بلوغ حاصل از این تجِربه می بایست قاعدتاً هنگام پیاده سازی روش پیشنهادی در عمل بررسی شده و به تحقیقات آتی موکول میگردد.

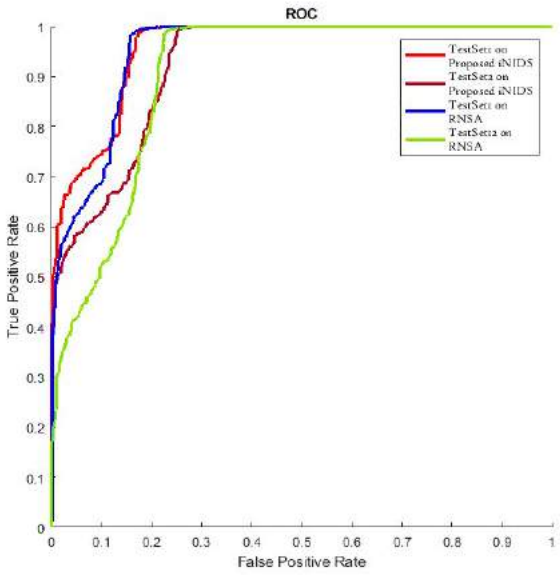
از طرفی در انجام آزمایشات بالا مشاهده گردید که کُلّ تعداد حملات ناشناخته برای دادگان آزمون اوّل و دوّم به ترتیب برابر با ۶۷ و ۱۸ رکورد ترافیک شبکه می باشند. در نتیجه هم سیستم پیشنهادی و هم دو متد ایمنی مصنوعی (DCA پیشنهادی و FtRNSA) که ارزیابی های مقایسه ای روشِ هیبریدیِ پیشنهادی با آنها انجام شده اند بیشترِ این حملات ناشناخته را به عنوان آنومالی شناسایی می نمایند ولی به دلیل آنکه دسته بندی ما از نوع باینری می باشد لذا نمی توان اطلاع دقیقی از کیفیّت برچسب زنی حملات و تفکیک صحیح آنها در دست داشت. بدین منظور میبایست از همان ابتدا دسته بندی غیر باینری (به تعداد نوع حملات) را انجام دهد. جدول زیر نرخِ شناسایی حملات ناشناخته تعبیه شده در هر سه دادگان آزمون را در پایانِ سیکلِ سوّمِ حیاتِ مصنوعی نشان می دهد.

جدول ۲۱ - نرخ شناسایی حملات ناشناخته در سه دادگانِ آزمون با روش پیشنهادی در مقایسه با FtRNSA

Method	Test Set	Number of All Unknown Attacks	Number of Detected Unknown Attacks	Detection Rate of Unknown Attacks	Total DR	Total FP
Proposed i-A-NIDS	First	86	61	70.93	98.7	17.2
	Second	101	72	71.28	99.1	25.2
	Third	38	33	86.84	99.3	21.1
RNSA	First	86	55	63.95	98.4	16.1
	Second	101	63	62.38	98.6	23.0
	Third	38	27	71.05	99.0	21.2



نمودار ۱۵ - ماتریس درهم ریختگی و منحنی های ROC مربوط به نتیجه دسته بندی نهایی دو دادگان آزمون اوّل و دوّم پس از طی سه سیکل اجرای متوالی در روش هیبریدی پیشنهادی

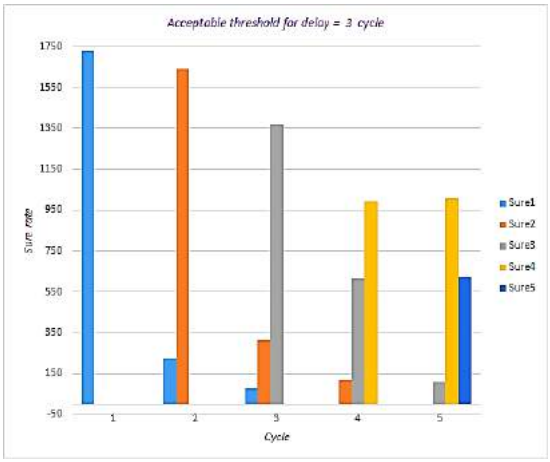


نمودار ۱۶ - منحنیِ تجمّعی ROC -

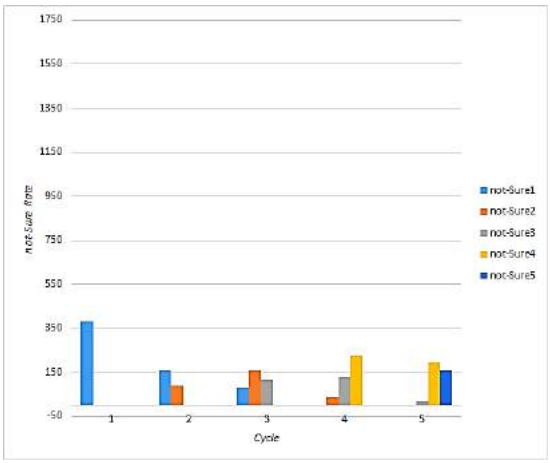
ارزیابی مقایسه ای دو سیکل نخست حیات مصنوعی سیستم هیبریدی پیشنهادی در مقایسه با FtRNSA

در ارزیابی سطح زیر منحنی (AUC) تجمّعی مشاهده می شود که سطح زیر نمودار سیکل نخست (قرمز رنگ) نسبت به سطح زیر نمودار آبی رنگ (FtRNSA) بیشتر بوده و عملکرد بهتری دارد. همین نسبت برای AUC نمودار قهوه ای نسبت به سبز در دادگان آزمون دوّم نیز صادق است. از طرفی به دلیل آنکه دو نمودار همدیگر را قطع می کنند بنا به موارد مطرح شده در بخش ۲-۲-۸ (تحلیل روابط) نمی توان مقایسه ای صحیح و دقیقی در خصوص بهتر بودن عملکرد دسته بندی داشت. بدین منظور نیازمند آن هستیم که متدهای مقایسه شده را با معیار استعداد تشخیص (CID) چک کنیم.

با بررسی خروجیها ، استعداد تشخیص متد آبی رنگ برابر $1.7052 - 0.0000\,i$ و استعداد تشخیص متد قرمز رنگ (سیستم پیشنهادی در آزمون دادگان نخست) برابر $1.9246 - 0.0000\,i$ بدست آمد. همچنین استعداد تشخیص متد سبز و قهوه ای نیز به ترتیب برابر 1.6418 و $1.6855 + 0.0001\,i$ بدست آمدند. ضمناً خروجی معیار استعداد تشخیص به دلیل آنکه این معیار از لُگاریتم در مبنای ۲ در محاسبات خود استفاده می کند، خروجی به صورت یک عدد مختلط می باشد که البتّه یک چالش است. در نتیجه اگر صرفاً قسمت های حقیقی این اعداد را در نظر بگیریم ملاحظه می شود که روش هیبریدی پیشنهادی در آزمون هر دو دادگان اوّل و دوّم استعداد بالاتری در تشخیص نسبت به RNSA دارد.



نمودار ۱۷ - تعداد نمونه های با برچسب قطعی در پنج سیکل اجرای حیات مصنوعی



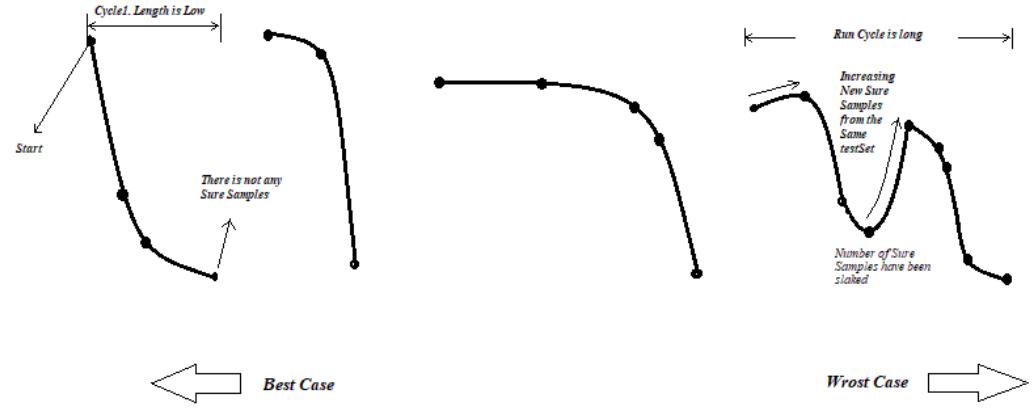
نمودار ۱۸ - تعداد نمونه های عدم قطعیّت باقی مانده در پنج سیکل اجرای حیات مصنوعی

از طرفی اگر روش هیبریدی پیشنهادی را بر اساس معیار همبستگی نیز مورد ارزیابی قرار دهیم همبستگی نمودار آبی رنگ در مقایسه با نمودار قرمز ۰,۸۵۶۶ به ۰,۸۵۲۵ بوده و خروجی همین معیار برای دو متد سبز و قهوه ای (آزمون دادگان دوّم) به ترتیب برابر با ۰,۸۰۸۲ و ۰,۸۱۳۵ و برای آزمایش این دو متد در دادگان سوّم نیز ۰,۷۹۴۴ به ۰,۷۹۸۸ برای سیستم پیشنهادی بدست می آیند.

این نتایج که حاکی از آن اند که روش هیبریدی پیشنهادی با استعداد تر است و در دراز مدّت در پیش بینی برچسب حملات در مجموع بهتر عمل می نماید.

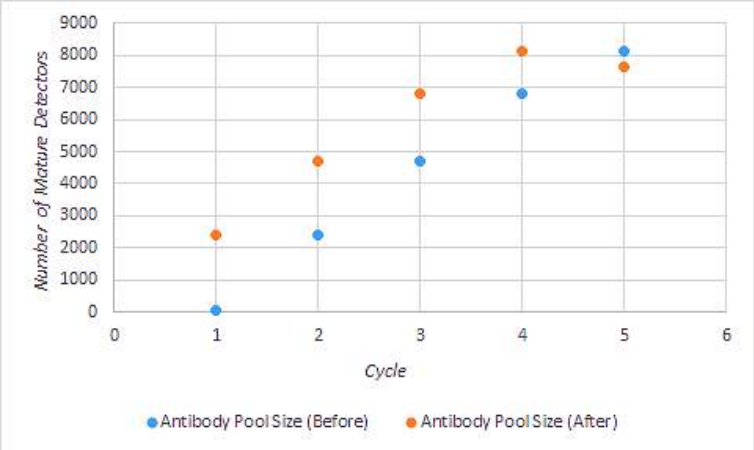
در خصوص دو نمودار فوق لازم است توضیحی ارائه گردد. نمودار اوّل وضعیت قطعیّت نمونه های آزمون را در روند اجرای سیکلهای متوالی حیات مصنوعی نشان داده و از خروجی نمودار دوّم نیز می توان میزان عدم قطعیّت را در سیکل ها مشاهده نمود. ابتدا در سیکل اوّل ، دادگان آزمون نخست با ۲۱۱۰ رکورد ترافیک شبکه وارد سیستم تشخیص شده و تحت آنالیز و بررسی خطوط دفاعی قرار میگیرند.

نتیجه اجرای خطوط در سیکل اوّل همیشه تعدادی نمونه را می تواند به عنوان نمونه های عدم قطعیّت اعلام نماید. بدین ترتیب نمونه هایی که قطعیّت برچسب آنها در سیکل اوّل اثبات نشده بهمراه تعدادی از نمونه های جدیدتر در سیکل دوّم مورد بررسی خطوط دفاعی قرار میگیرند و به همین ترتیب. سیکل های اجرای حیات مصنوعی را در حالت کُلّی می توان به صورت اینفوگرافیک زیر مورد تجزیه و تحلیل تئوری قرار داد. مطابق شکل ۲۳ هر چه مشابهت نمودار سیکلهای اجرا به اشکال سمت چپ نزدیکتر باشد طول اجرا کمتر بوده و سیستم سریعتر به بیشترین میزان قطعیّت می رسد.



شکل ۲۳ - اینفوگرافیکِ مربوط به تحلیل کلّی وضعیتهای قطعیّت بر اساس طول اجرای سیکلهای حیات مصنوعی

البته ذکر این نکته ضروری است که شکل سمت راست بدترین حالت نیست و بدتر بودن آن بیشتر به افزایش طول اجرا و تعداد سیکلهای متوالی نسبت به اشکال سمت چپ بر میگردد. طول اجرا بستگی به منابع پردازشی و حافظه ای سیستم دارد.



نمودار ۱۹ - اندازه استخر آنتی بادی های بالغ قبل و بعد از هر سیکل اجرا

اگر این منابع به اندازه کافی موجود باشند و سیستم بتواند از منابع حافظه ای و پردازشی رزرو شده در مواقع لزوم استفاده نماید در اینصورت سیستم خواهد توانست در حین یادگیری تشخیص دهد و بالعکس. در نتیجه می توان گفت که سیکلها سریعتر اجرا شده و مشابهت شکل نمودار به سمت اشکال سمت چپ بیشتر متمایلتر خواهد بود. در سیکل اجرای پنجم اقدام به بروز رسانی و تزریق تعدادی حمله ی جدید به پایگاه امضاء ها نمودیم و این منجر به اجرای فاز یادگیری اوّلیه و سپس فاز یادگیری ثانویه به منظور تعدیل دو فضای غیر خودی و خودی گردید. در نتیجه سیستم به طرز هوشمندانه ای شروع به تعدیل خود می نماید. بطوریکه از شکل فوق قابل مشاهده است نتیجه این تعدیل در سیکل پنجم تعداد آنتی بادیهای بالغ را از ۸۱۲۹ به ۷۶۵۳ آنتی بادی کاهش داده است.

از سیکل پنجم به بَعد، استراتژی پیشنهادی مبتنی بر مختصّات قطبی ارائه شده در ۳-۲-۸-۱ برای پوشش حفره های از قبل موجود و خصوصاً حفره های جدیدِ بوجود آمده در اطراف تشخیص دهنده های تعدیل شده نیز می بایست بطور همزمان با حیات مصنوعی اجرا گردد که به دلیل وجود محدودیتهای نرم افزار متلب در پردازش موازی و جریانی اطلاعات و عدم وجود منابع حافظه کافی امکان این آزمایش و ادامه ی روندِ اجرای سیکلهای حیات مصنوعی مقدور نبود.

توضیحی لازم است در مورد خروجی های فوق داده شود. دادگان آزمون نخست سه سیکل اجرا را طی نموده و در نهایت ۷۹ نمونه با وضعیت عدم قطعیّت وارد ماژول تصمیم گیری نهایی می گردند تا وضعیت برچسب نهایی آنها با اعمال رابطه وزنی مربوطه مشخص گردد. همچنین به دلیل آنکه در سیکل اوّل ۱۷۲۸ نمونه ی قطعی از دادگان آزمون اوّل از سیستم با موفقیّت خارج می شوند ، به جای همین تعداد نمونه ی خارج شده ، از دادگان آزمون دوّم در سیکل دوّم به سیستم بافر شده و مورد پردازش قرار میگیرند.

به همین ترتیب ۳۸۲ نمونه ی جدید در دادگان آزمون دوّم به عنوان نمونه های باقی مانده با وضعیت عدم قطعیّت همچنان منتظر می مانند تا در سیکل (های) بعدی (سوّم به بعد) در کنار نمونه های جدیدتر مورد آنالیز و بررسی مجدّد قرار بگیرند. همانطور که از نتایج نمودارهای ۱۳ و ۱۵ مشاهده می گردد درصد بهبود کیفیّت و دقّت دسته بندی در سیستم تشخیص نفوذ پیشنهادی بدون اعمال کاهش ویژگی چشمگیر بوده و این روند به مرور زمان و با رشد سیستم (تولید و تکثیر – تثبیت تشخیص دهنده های حافظه ی سیستم در اثر کسب تجربه ی آزمون ها ، بروز رسانی و تعدیل فضاهای خودی و غیر خودی و همچنین پوشش حفره ها) افزایش خواهد یافت.

تکنیکهای کاهش ابعاد (الگوریتم انتخاب ویژگی ده-پا در این پایان نامه) صرفاً سرعت اجرای آزمون ها را افزایش می دهند و به دلایلی که در بخش بعد در خصوص تاثیر منفی اعمال فاز انتخاب ویژگی در محاسبات فواصل در متدهای دسته بندی مبتنی بر ایمنی مصنوعی بیان شده الگوریتم انتخاب ویژگی به هیچ وجه نمی تواند دقّت و عملکرد دسته بندی را در مورد سیستم های ایمنی مصنوعی بهبود ببخشد. از طرفی از نتایج

آزمایشِ دو دادگان آزمون نخست در (نمودار ۱۵ در مقایسه با نمودار ۱۲) مشاهده گردید که کاربرد تئوری خطر در قالبِ یک سیستم دسته بند نیمه نظارت شده در خط نخست دفاعی توانسته این دقّت را در دسته بندی بهبود ببخشد بنابراین مطابق آزمایشات انجام شده در بخش ۳-۲-۲-۱ و نتایج آزمایشات آن ، به نظر نمیرسد که کاربرد متد کاهش ابعاد ده-پا بجز افزایش سرعتِ آزمون فایده ی دیگری نداشته باشد.

این نتیجه گیری صرفاً در مورد متدهای ایمنی مصنوعی انتخاب منفی و سلّولهای دندریت و بخصوص روش هیبریدی پیشنهادی صادق است و در مورد متدهای دسته بندی مبتنی بر یادگیری ماشین تاثیر کاربرد یک متد کاهش ابعاد در ارتقای کیفیت دسته بندی و افزایش دقّت بسیار موثّر است. در بخش بعدی دلیل این ادّعا و چرایی عدم تاثیر متد کاهش ابعاد بر متد های ایمنی مصنوعی را تحلیل نموده ایم.

۲-۱۱-۲-۳- شناسایی زیر حملات [1]

خوشبختانه سیستم دسته بندیِ روشِ هیبریدیِ پیشنهادی تمامی حملات شناخته شده و بیشترِ حملات ناشناخته ی موجود در سه دادگان آزمون را شناسایی نموده و در دسته ی آنومالی قرار می دهد. (جدول ۲۱) امّا نمی توان صرفاً با دسته بندی باینری نوع حمله را شناسایی نمود مگر آنکه تعداد خوشه ها بیشتر از دو باشد که در آن صورت مسئله به یک مسئله دسته بندی غیر باینری تبدیل می شود. در این پژوهش مسئله صرفاً با دسته بندیِ باینری حل شده است. به منظور تفکیک حملات به چهار دسته ی معمول DoS. Probe. U2R . R2L که در دادگانِ NSL – KDD به این صورت است ما نیاز داریم که دسته بند از همان با زیر حملات آموزش ببیند و فاز آزمون خط دوّم ، نیز ترافیک شبکه را به چهار دسته حمله فوق دسته بندی نماید.

در بخش های قبل تجربه ی حاصل از آزمایشِ الگوریتم های فراابتکاری مبتنی بر جمعیّت نمونه اوّلیه و ضعف این الگوریتم ها را در دسته بندی باینری ترافیک شبکه مشاهده نمودیم. امّا نکته ای که مطرح می شود آنست که این الگوریتم ها علیرغم ضعف در دسته بندی باینری پتانسیل این را دارا هستند که دادگان ترافیکِ شبکه را به بیش از دو دسته به خوبی دسته بندی نمایند. در نتیجه ، چالشِ روشهای فراابتکاری همانطور که گفته شد " متمرکز گرا بودن" آنهاست بدین معنی که سعی دارند صرفاً بهترین مرکزیّتها را به تعداد خوشه های مطلوب مشخص نمایند و با تابع مشخصه [2] این موقعیتها را در طی فازهای تکرار ارزیابی میکنند تا به بهترین موقعیتهایی برسند که بالاترین فیتنس را در خروجی تابع داشته باشند. توضیحاتِ تکمیلی در خصوصِ دسته بندی متمرکزگرا و چالشها و پتانسیلهای آن در [پ-الف-۷] بررسی شده است.

[1] Sub Attacks
[2] Objective Funstion

۳-۱۱-۲-۳- تحلیلی کوتاه بر تاثیرِ منفیِ کاربردِ الگوریتم انتخابِ ویژگی بر متدهای ایمنیِ مصنوعی

متدهای دسته بندی متمرکز گِرا در مورد ترافیک شبکه نمی توانند به دسته بندی مطلوبی را به انجام برسانند به دلیل آنکه اوّلا استراتژی آنها وابستگی زیادی به نحوه ی توزیع داده ها در فضای ابعاد مسئله دارد بطوریکه اگر توزیع دادگان مورد استفاده برای دسته بندی مشابه دادگان نفوذ ، داده های آن در هم تنیدگی و چگالی بالایی در توزیع داشته باشند نمی توان با متدهای دسته بندی متمرکز گِرا مرکزیّت ها را مشخص نمود زیرا چگالی داده ها آنقدر زیاد بوده و نزدیک به هم اند که امکان دسته بندی باینری در این گونه موارد بسیار مشکل و حتّی ممکن نیست. بنابراین تاثیرالگوریتم انتخاب ویژگی در دسته بند متمرکز گِرا به شرطی که آزمون در دادگانی چَگال و موقعیّتهای داده های آن به شدّت در هم تنیده نباشد مفید بوده و منجر به افزایش دقّت سرعت اجرا و در نهایت ارتقای عملکرد و کیفیت دسته بندی میگردد.

در خصوص متدهای ایمنی مصنوعی وضع به گونه ی دیگری است . اصولاً همانگونه که در فصول و بخش های گذشته بحث شد این متدها مسئله ی دسته بندی را به روش نامتمرکز (کاملاً توزیع شده و مشارکتی) حل می کنند بطوریکه در مورد مسائلی مانند تشخیص نفوذ که دادگان آن از در هم تَنیدگی و چگالی بالایی در توزیع برخوردارند بسیار مفید و راهگشاست. از طرفی استراتژی حل مسئله در این متدها بر مبنای محاسبه ی فاصله است که در اغلب موارد از فاصله ی اقلیدسی استفاده می شود. در نتیجه استراتژی تعیین فاصله تفاوت بسیاری با استراتژی مبتنی بر تعیین مرکزیّت ها دارد.

زیرا در تعیین فاصله اگر ویژگی خاصی (وابسته یا غیر وابسته) حذف گردد دقّت در تعیین فاصله ی نمونه ها از یکدیگر کاهش می یابد و ممکن است چندین نمونه روی هم بیافتند. پس کاهش ابعاد مسئله تاثیری به جز افزایش سرعت اجرای آزمون ها در سیستم های ایمنی مصنوعی ندارد. در نتیجه در متدهای ایمنی مصنوعی بهتر است ابعاد بالا باشد تا بتوان به طور دقیق دقّت فواصل را تعیین نمود و بر همین اساس دسته بندی صحیحی را انجام داد.

همانگونه که در بالا بیان گردید ، متدهای ایمنی مصنوعی و روش هیبریدی پیشنهادی تشخیص ترافیک آنومالی شبکه را به صورت غیر متمرکز و کاملاً توزیع شده دسته بندی نموده و نیز مانند سایر الگوریتم های دسته بندی و یادگیری ماشین می توانند دسته بندی غیر باینری را نیز انجام دهند به شرطی که از همان ابتدا پایگاه امضاء های حملات آنها بر اساس نوع حمله از قبل تفکیک شده و سیستم با آنها آموزش دیده باشد.

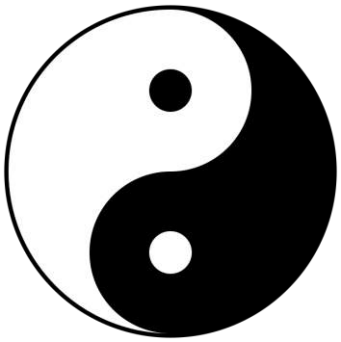
به عنوان نمونه در دادگان UNSW – NB15 ما نُه زیر حمله داریم. برای اینکه سیستم ایمنی مصنوعیِ شبیه سازی شده از همان ابتدا بتواند نوع حمله را شناسایی و دسته بندی غیر باینری انجام دهد باید ۹ زیر مجموعه ی خودی – آنومالی باضافه ی یک زیر مجموعه ی پروفایل خودی-نرمال داشته باشیم و الگوریتم انتخاب منفی را با این دَه زیر مجموعه آموزش دهیم و یا این نُه زیر حمله را با چهار حمله ی استاندارد به

طریقی معادلسازی نماییم. بدین منظور از تجربیّات متخصصان داده کاوی در پیام رسان ریسرچ گیت کمک گرفتیم و از آنها درباره نحوه ی این معادلسازی پرسشی را مطرح نمودیم. (جدول ۱۸ در [پ-ب-۱])

در اکثر موارد در آزمایشات شبیه سازی ها و مقالات دسته ی چهارگانه ی فوق را معمولاً در آزمون ها مدّ نظر قرار می دهند بدین منظور به عنوان یک پیشنهاد میتوان این ۹ زیر حمله را به چهار زیر حمله به صورتی که در جدول ۱۸ در [پ-ب-۱] بیان شده نگاشت نمود.

۱۲-۲-۳- تحلیلِ فلسفه وجودیِ امنیّت اطلاعات – فلسفه ی یین و یانگ

یین و یانگ ، مفهومی است در نگرِش چینیانِ باستان به نظام هستی. یین و یانگ شکل ساده شده ای از مفهومِ یگانگیِ متضادهاست. از دیدگاهِ چینیانِ باستان و تائوباوران ، در همه پدیده ها و اشیاء غیر ایستا در جهانِ هستی ، دو اصل متضاد ولی مکمّل وجود دارد. یین و یانگ نشان دهنده ی قطب های مخالف و تضادهای جهان هستند. البته این بدان معنی نیست که یانگ خوب است و یین بد است. بلکه یین و یانگ مانند شب و روز و یا زمستان و تابستان بخشی از چرخه ی هستی هستند. وقتی تعادل و احساس خوبی به وجود می آید که تعادلِ بینِ یین و یانگ برقرار باشد. [۴۰]



شکل ۲۴ – تایجیتو [1] نماد سنّتی نشان گر نیروهای یین و یانگ ، برگرفته از [۴۰]

نقطه های متضادی که در شکلِ بالا دیده می شود به این مفهوم است که یین وقتی به حداکثرِ خود برسد و می خواهد تمام شود در درونش یانگ را دارد و وقتی هم که یانگ می خواهد به حداکثر خود برسد در درونش یین را دارد. یعنی وقتی یکی تمام می شود دیگری در درونش رُشد می کند و این چرخه ادامه پیدا می کند.

با الهام از این نظریه و بسط دادن موضوع آن در دنیای امنیت فناوری اطلاعات می توانیم اینگونه این فلسفه را بیان کنیم که به دلیل آنکه مبنای دو نیروی خیر و شرّ ، هوش انسانی می باشد ، پس هر فعالیتی که منشاء

[1] TaiJitu

آن هوش انسانی باشد را نیز شامل میشود. پس در بسط دادن این قاعده ، فعالیتهای هوش مبنا مانند هک و نفوذ و شناسایی آن از این قاعده مستثنی نبوده و به نظر میرسد که فلسفه ی وجودی آنها در نهایت به بین و یانگ میرسد.

با این نتیجه گیری ، از دیدگاه فلسفی قابل اثبات است که چون خیر و شرّ همواره وجود داشته و خواهند داشت پس در بحث شناسایی نفوذ (بخش سپید رنگِ شکل فوق) هیچ گاه نمی توان به طور صد در صد به نرخ های خطای منفیِ کاذبِ صفر و نرخ های تشخیصِ صد در صد دست یافت. همینطور برای بخشِ سیاه رنگ که در واقع بیانگر اقدامات هک و نفوذ می باشد نمی توان به طور صد در صد ادّعا کرد اثر تمامی نفوذها به سیستم های فناوری اطلاعات همواره پایدار خواهد بود و هیچ عاملی نمی تواند اقدام به شناسایی آن بنماید. زیرا بالاخره با گذرِ زمان دیر یا زود اقدامات کشف و شناسایی آن از سوی مختصصان با استفاده از ابزارها ، اتّخاذِ سیاستها انجام خواهد گردید ولی همانطور که در بالا نیز اشاره شد آن فعالیت نیز پایدار نیست و نخواهد بود. بنابراین خطاهای مثبت و منفی کاذب همواره وجود خواهند داشت و هیچ موقع به صفر درصد بطوریکه این عدم وجود خطا پایدار باشد نخواهند رسید. آسیب پذیریهای روز صفر ، کاتالیزوری در تسریع این نبرد میان این دو نیرو (نفوذ و شناسایی) انجام می دهند.

به دلیل وجود محدودیّت در منابع پردازشی و عدم وجود حافظه کافی جهت اجرای سریع شبیه سازی ها و محدودیتهای نرم افزار متلب همچون زمان اجرای بالا در پردازش داده های حجیم ، فُرصت پیاده سازی و صنعتی سازیِ این طرح (بررسی و تحقیقات بیشتر در خصوص امکان تبدیل روش هیبریدیِ پیشنهادی به محصولِ اوّلیه) به تحقیقاتِ آینده در دوره تخصصیِ دکتری موکول خواهد گردید.

۱۳-۲-۳- چرا روش هیبریدیِ پیشنهادی یک سیستم خِبره است ؟

بنا به تعاریف ارائه شده از یک سیستمِ خِبره که نوعی سیستم هوشمندِ مدیریت دانش نیز محسوب می گردد و نظر به عقیده محقّق در این پژوهش ، یک سیستم تشخیص نفوذِ خبره سیستمی است که توانایی کسب دانش حاصل از تجربه یادگیری (های) قبلی را داشته و ضمن مدیریتِ پایدارِ دانشِ کسب شده بتواند از آن در جهت پیش بینی صحیحِ برچسب کلاسِ نمونه ها حداکثر بهره را ببرد.

بنابراین روش هیبریدیِ پیشنهادی که در واقع یک سیستم تشخیص نفوذ هوشمند شبکه نیز هست از جَهاتی خِبره محسوب می گردد. زیرا اوّلاً ویژگیهای یک سیستم هوشمند را دارا می باشد همچون قدرت یادگیری از آموخته های قبلی و ویژگی خاص خود یادگیری و ثانیاً دانش کسب شده حاصل از نتیجه یادگیری خود را می تواند در جهت ایجاد مدل و پیش بینی برچسب استفاده نموده و این دانش را با تَعدیل تش.د.بغ. م.حظ مدیریت نماید.

فصل چهارم

Conclusion and Future Directions

نتیجه گیری
و
پیشنهادها

۱-۴- مقدمه

به عنوان فصل پایانی این پژوهش ، به جمع بندی کلّی مطالب ، مباحث مطرح شده و دستاوردهای حاصل از این تحقیق و ارزیابی برتری آن نسبت به کارهای مشابه سایر محققان پرداخته شده و ضمن ارائه نتیجه گیری لازم ، جهت گیری ها و چشم انداز تحقیقاتی آینده را به خوبی ترسیم نموده ایم. در بخش جمع بندی بر روند انجام کار مروری خلاصه شده است. شامل : کل فصول ، ادبیاتِ پژوهش تا ارائه روش پیشنهادی ، روش تحقیق و آنالیز یافته ها از هر سه دیدگاه تئوری ، تجربی (شبیه سازی) و فلسفی. در بخش نتیجه گیری مهم ترین نتایج و دستاوردهای حاصل از اجرای پژوهش ضمن پاسخ به پرسشهای آگاهانه و بررسی صحّتِ فرضیات مطرح شده در جهت رسیدن به اهداف پژوهش ، بیان و امکانِ برتریِ روشِ هیبریدیِ پیشنهادی نسبت به سایر کارهای تحقیقاتی مشابه بحث شده است.

بخش انتهایی نیز ضمن ترسیم چشم انداز پژوهش به ارائه پیشنهادهایی جامع و مبسوط به منظور پیاده سازی عملی و بهبود روشِ فعلی و ادامه زنجیره های تحقیقاتی مرتبط با پروژه تحقیقاتی جاری در آینده، پرداخته است.

۲-۴- جمع بندی

محتوای گزارش کارِ پروژه تحقیقاتیِ پایان نامه شامل چهار بخش در قالب دو فصل می باشد. فصل دوّم با مُرور ادبیات نظری به عنوان پیش زمینه مطالعات زیستی پژوهش آغاز شده و در بخش دوّم این فصل ، آندسته از کارهای پژوهشی که از سیستم ایمنی مصنوعی در کار خود استفاده کرده اند طبقه بندی و بررسی شده اند. در فصل سوّم بر دو مقوله مهم تئوری خطر و انتخاب منفی و امکان ترکیب این دو جهت ایجاد یک سیستم تشخیص نفوذ با دو خط دفاعی مبتنی بر مفاهیم دفاع در عمق تمرکز شده است. ارائه روش ترکیبی پیشنهادی با ایجاد نوعی حیات مصنوعی و آنالیز و تجزیه و تحلیل یافته ها از سه دیدگاه تئوری، تجربی (شبیه سازی) و فَلسفی موضوع بخش دوّمِ فصل سّوم این پژوهش است.

محتوای بخش نخستِ فصل دوّم به ارائه پیش زمینه مطالعات زیستی پژوهش در قالب ادبیات نظری اختصاص دارد. مفاهیمِ ایمنی زیستی همچون چَرخه دفاع و اهمیت آن در پیش گیری از نفوذ ، اصولِ ایده پردازی و مدلسازی تا مهندسی و بطور کلّی شیوه ی الهام گرفتن از طبیعت در حل مسئله ، سیستم ایمنی زیستی و مصونیّتهای ذاتی و اکتسابی و سیستم ایمنی مصنوعی و چهار متد آن بطور مبسوط معرفی و بررسی شده اند. در نهایت در قسمت انتهایی این بخش به بررسی نحوه نگاشت (انطباق) مفاهیم ایمنی مصنوعی به تشخیص نفوذ و همچنین پتانسیلها و چالشهای سیستم های تشخیص نفوذ مبتنی بر ایمنی مصنوعی طبقه بندی و بررسی شده اند. با مطالعه ادبیات نظری و تجربی در فصل دوّم ، پیش زمینه مطالعاتی کاملی از دو جنبه ایمنی-زیستی و تجربی بدست می آید. اینکه سایر محققان چگونه از سیستم ایمنی مصنوعی در ارائه سیستم تشخیص نفوذ بهره برده و اصولاً سیستم ایمنی مصنوعی چگونه مسئله را حل می نماید؟

بحث و بررسی در خصوص تفاوت شیوه حل مسئله تشخیص نفوذ بوسیله متدهای ایمنی مصنوعی موضوع بخش دوّم از فصل سوّم می باشد. متدهای ایمنی مصنوعی مسئله تشخیص نفوذ را به شیوه غیر متمرکز و کاملاً توزیع شده و مشارکتی حل می کنند. تشخیص دهنده های بالغ که رویکرد مشارکتی و توزیعی دارند در حکم سنسورهای تشخیص نفوذ هستند که با تعدیلِ خود ضمن پوشش بهتر فضای غیر خودی ، منجر به رُشد ایمنی سیستم می شوند.

در فصل سّوم ، ارائه روش پیشنهادی ، به مرور ایده هایی نو جهت بهبود عملکرد دو مکانیسمِ سلّولهای دندریت و انتخاب منفی پرداخته شده است. در بخش نخست ، مفهوم تئوری خطر با مدلِ انتزاعیِ آن یعنی DCA استاندارد معرفی و تشریح شد و سعی گردید تا با ارائه ایده هایی در طول روند تحقیق ، مکانیسمِ آن بهبود داده شده و چالشهای آن مرتفع گردد. چالشهایی مانند نمونه برداری غیر تصادفی آنتی ژنها و نحوه تشکیل بُردار آنتی ژن، لزوم از پیش مرتّب بودنِ دادگان آزمون بر اساس ویژگی کلاس ، تخصیص احتمالات سیگنالهای ورودی.

به عنوان نمونه به منظور تخصیص و محاسبه احتمالات سه سیگنال ورودی از الگوریتم خوشه بند سیاه چاله استفاده شد و در پیوست ایده هایی بهمراه روابط مربوطه ارائه گردید که متاسفانه نتایج منفی ارزیابی این ایده مانع از کاربرد آن در روند بهبود DCA گشت. علت این مسئله این بود که استراتژیِ پردازش در الگوریتم سیاه چاله در دسته بندی باینری به صورت متمرکزگِرا بوده و با این روش قادر به حل مسئله تشخیص نفوذ شبکه نیست.

همچنین بررسی و ارزیابی کاربرد دو متد بهره اطلاعات و دانش متخصصان در تخصیص سیگنالهای ورودی نیز انجام گردید و مشاهده شد که نتیجه کاربرد متدِ بهره اطلاعات، عملکرد نرخ تشخیص را بالا برده ولی نرخ خطای مثبت کاذب را نیز بالا می برد. در صورتیکه استفاده از دانش متخصصان نرخ های تشخیص صحیح مثبت و منفی ثابت و یکسانی دارد که البته نیاز به بهبود دارند. بطوریکه به نظر میرسد در صورتیکه تحلیل و آنالیز دقیق ترافیک شبکه جهت کشف و نگاشت به سیگنالها صورت گیرد استفاده از این روش امیدوار کننده تر است. در کاربرد دانش متخصصان ، دادگان مربوطه کاملاً آنالیز گردید و با اعمال روابطی بر روی شش ویژگی مد نظر جهت نگاشت به سیگنالهای ورودی استفاده گردید. کشف این زیر مجموعه ویژگیها و اعمال رابطه ای دقیق بر روی ترافیک شبکه جهت نگاشت به سیگنالهای ورودی چالش بزرگی است.

از طرفی ایده هایی مانند تقسیم دادگان آزمون به بافت های سلّولی بر اساس ویژگیهای پروتکل- سرویس جهت تسریع روند شناسایی خطر و پردازش سیگنالها ، پویا نمودن حدآستانه مهاجرت سلّولهای دندریت و پیشنهاد استراتژی شعاع نمونه برداری آنتی ژنها که با سه معیار بیشینه ، میانگین و میانه بررسی و ارزیابی گردیدند، مشاهده شد که استفاده از این ایده ها تاثیر زیادی در بهبود عملکرد دسته بندی DCA دارند. علاوه بر این بررسی اِعمال تغییراتی بر روی ماتریس وزنها جهت بهبود در پردازش و تولید سیگنالهای خروجی نیز نتیجه ای در بر نداشت و همان مقادیر وزنی ماتریس پیشفرض مناسب دیده شد.

در بخش ادبیات نظری بحث گردید که اصل دوّم دفاع در عمق می گوید که هر لایه دفاعی به طور مستقل کار می کند و این لایه ها در شناسایی نفوذ مکمّل یکدیگر هستند. در بخش دوّم از فصل سوّم نیز مکانیسمِ انتخاب منفی در دادگان آزمون نفوذ بررسی و ارزیابی گردید و پس از کسب تجربه از آزمون های مکرّر RNSA، مقادیر شعاع های خودی متناسب در هر دادگانِ آزمون بدست آمد. مقادیر این دو شعاع برای دادگان آزمون نخست به ترتیب برابر 0.01 برای خودی – نرمال و 0.1 برای خودی- آنومالی بودند. این مقادیر جهت انجام آزمون های بعدی با روش ترکیبی پیشنهادی لازم بودند. در نتیجه امکان ترکیب دو مکانیسمِ DCA پیشنهادی (حاصل از بخش نخست فصل سوّم) و RNSA با مقادیر شعاع بدست آمده ، در قالب دو خط دفاعی بر مبنای اصل دوّم امنیّت ، یعنی دفاع در عمق بررسی و ارزیابی گردید. این بررسی از سه دیدگاهِ نظری ، شبیه سازی و فلسفی انجام شد و نتیجه مطلوبی را به دنبال داشت. در رابطه با امکان بهبود RNSA جهت پوشش بهتر حفره های فضای غیر خودی و تسریع روندِ تولید تشخیص دهنده ها در موقعیت های موثر در این فضا، قسمتی

از روند تحقیقات منجر به ارائه ایده استراتژی مختصات قطبی گشت. از این استراتژی جدید در فاز یادگیری اوّلیه جهت تولید تشخیص دهنده های بالغ و پوشش بهتر حفره های بوجود آمده در فضای غیر خودی استفاده گردید که نتایج مطلوبی را از نظر تسریع روند تولید تشخیص دهنده ها و بلوغ آنها و همچنین کاهش نرخ خطای مثبت کاذب به دنبال داشت. همچنین محاسبات مربوط به توجیه تئوریِ هیبریداسیونِ خطوط دفاعی بر مبنای احتمالات ممکن در رخداد های صحیح و اشتباه در برچسب زنی نمونه های آزمون در هر یک از خطوط انجام گردید. توجیه تجربی نیز بر مبنای مکمّل سازیِ خطوط دفاعی با ایجاد نوعی حیاتِ مصنوعی و ارزیابی نتایج آزمون و شبیه سازی با اجرای سیکلهای متوالی در این حیات انجام شد.

آزمون های نهایی روش ترکیبی پیشنهادی دو بار انجام شدند یکبار DCA پیشنهادی با کاربرد متد بهره اطلاعات و بار دوّم همین مکانیسم با کاربرد متدِ دانش متخصّصان در تخصیص سیگنالهای ورودی. از طرفی استفاده از بهره اطلاعات علیرغم نرخ تشخیص بالای آن در خط نخست، بروز خطای مثبت کاذب بالای آن منجر به این می شود که عملکرد خط دوّم سیکلهای حیات طولانی تری را طی نماید تا سیستم دیرتر به قطعیت بالایی برسد و این زمان بَر خواهد بود بطوریکه چالشِ بلادرنگ بودنِ تشخیص را در پی خواهد داشت.

نتایج آزمایشهای نهایی روشِ هیبریدیِ پیشنهادی در مقایسه با RNSA میزانِ بهبودِ چشمگیری را در طی سه سیکل نخست حیات مصنوعی نشان می دهند. جدول و نمودار زیر این میزان را به تفکیک سه دادگانِ آزمون آزمایش شده بر اساس نتایجِ جدول ۲۱ و همچنین نمودار ۱۶ از فصل سوّم به خوبی نمایش داده است.

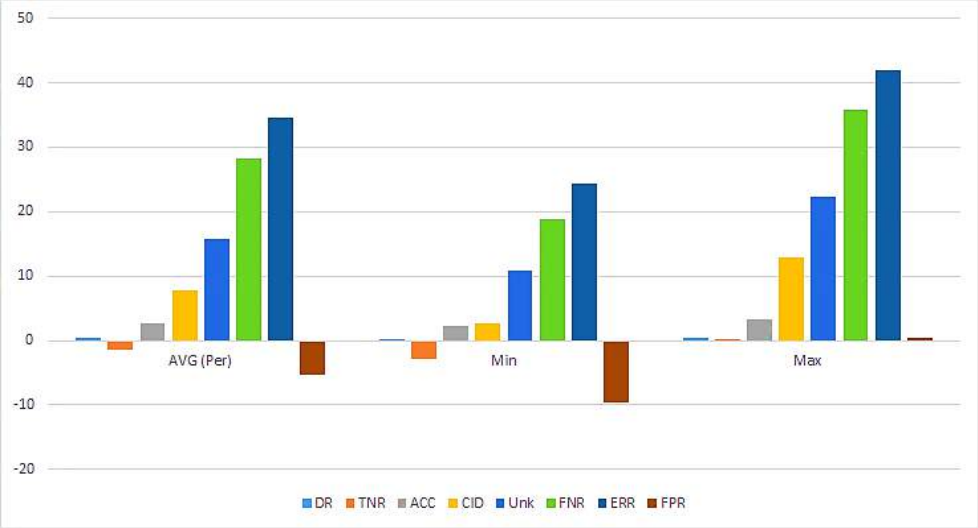
جدول ۲۲ – مقایسه میزانِ بهبودِ عملکردِ دسته بندیِ روشِ هیبریدیِ پیشنهادی در پایان سیکلِ سوّم ، نسبت به عملکردِ مکانیسمِ RNSA به تفکیکِ سه دادگانِ آزمون

	Improvement of the Proposed hybrid method Compared with RNSA's performance (%)			
Measure	**TestSet1**	**TestSet2**	**TestSet3**	**Average (%)**
DR	0.304878049	0.507099391	0.303030303	0.371669
TNR	-1.31108462	-2.85714286	0.126903553	-1.34711
Accuracy	2.449414271	2.178649237	3.347732181	2.658599
CID	12.86652592	2.661712754		7.764119
Unknown DR	10.91477717	14.2673934	22.22378607	15.80199
FNR	18.75	35.71428571	30	28.15476
Err	37.70491803	24.3902439	41.89189189	34.66235
FPR	-6.83229814	-9.56521739	0.471698113	-5.30861

مطابق اطلاعات جدول در پایان اجرای سیکلِ سوّمِ حیاتِ مصنوعی ، میانگینِ میزانِ بهبود در شناسایی حملات ناشناخته بوسیله متدِ پیشنهادی نسبت به RNSA در حدود ۷ درصد می باشد. همینطور در مورد کاهش نرخ های خطا نیز بهبود چشمگیری مشاهده می شود. علاوه بر این در خصوصِ میزانِ کاهشِ خطای منفیِ

کاذب نیز عملکردِ متد پیشنهادی در مقایسه با RNSA به طور میانگین در حدود ۳۴ درصد بهبود داشته است. نمودار زیر کمینه ، بیشینه و میانگینِ میزانِ بهبودِ نرخ های دسته بندی را نشان می دهد. به جز خطای مثبت کاذب در مابقی نرخ های دسته بندی بهبودهای چشمگیری مشاهده می گردد. داده های این نمودار حاکی از آنند که متد پیشنهادی از ضعفِ عمده ی شناساییِ صحیحِ نمونه های نرمال، برخوردار است که به مرور زمان و با تولید وتکثیر تش.د.بغ و تعدیلِ فضاهای خودی و غیر خودی می توان این خطا را پایین آورد. تسریع در کاهش این نوع خطا بستگیِ مستقیمی به وجود منابع قوی پردازشی و مهمتر از آن اعمالِ استراتژی مناسب جهت تولید تش.د.بغ و پوشش حفره ها و فضاها در ابعاد مسئله دارد که البته مورد دوّم موضوعِ این پژوهش بود که به خوبی بر روی آن کار شد.

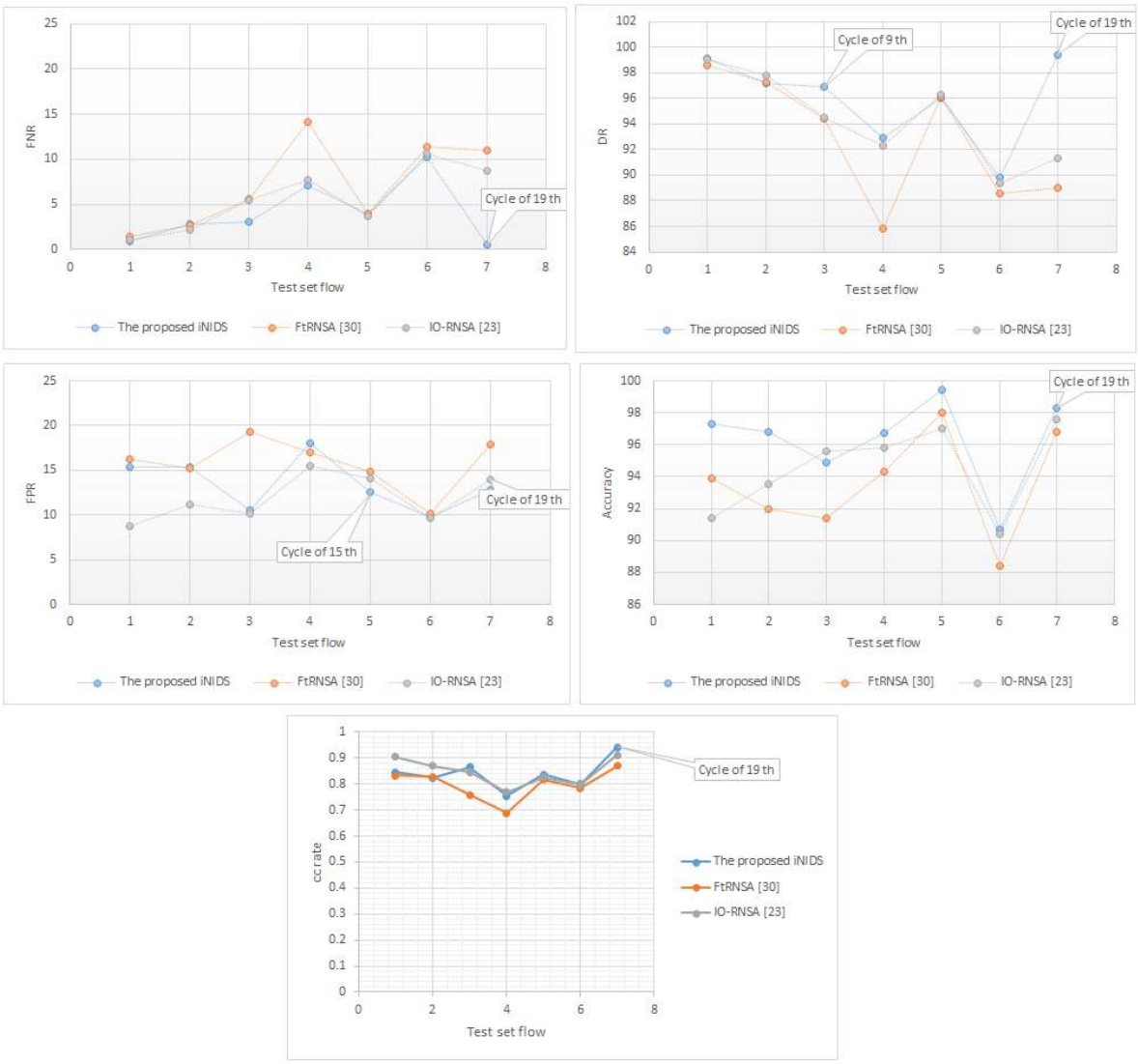
نمودار ۲۰- بازه پیشرفت در بهبودِ نرخ های دسته بندیِ متدِ پیشنهادی پس از اجرای سه سیکلِ متوالی [1]



علیرغم تمام پتانسیلهایی که سیستم پیشنهادی از نظر نرخ تشخیص حملات ناشناخته ، صحّت و خطا دارد ولی به نظر میرسد که نسبت به متد مشابه دیگر از نظر دو معیار FPR و TNR ضعیف عمل کرده است. (میله های نارنجی و قهوه ای در شکل فوق). اطلاعات این نمودار و جدول فوق این نکته را آشکار می کنند که سیستم پیشنهادی از لحاظ تشخیص صحیحِ نمونه های نرمال مشکل دارد. این مشکل می تواند با تولید و تکثیر تشخیص دهنده های بالغ و فرایند احیای حافظه در دراز مدت برطرف گردد. سرعت کاهش FPR نیازمند منابع پردازش و حافظه بالایی است. ما سیستم پیشنهادی را تا سیکل ۱۹ ام حیات آزمودیم. در طول اجرای سیکلها ۷ بار از دادگان UNSW دادگان نفوذ به همان اندازه (۲۱۱۰) بصورت برداشت بدون جایگذاری آماده و به سیستم تزریق گردید. نمودارهای زیر نشان میدهند که سیستم پیشنهادی (منحنیِ آبی) نسبت به دو

[1] Max – Min

متد دیگر مشابه، در شرایط یکسان و تحت دادگان نفوذ یکسان در هر مرحله چه طور رقابت می کنند. اندازه پارامتر $Th_{maxPractCycles} = 3$ تعیین گردید.



شکل ۲۵- تداوم حیات مصنوعی تا سیکل ۱۹ اُم

همچنین به عنوان بخش پایانی تحقیق نیز فلسفه یین و یانگ معرفی گردید که به نظر می رسد توجیه مناسبی در عدم امکانِ اجتناب از بروزِ خطاهای کاذب و مقابله با حملات روز صفر (ناشناخته) در مسئله تشخیص نفوذ باشد. توجیهی که به نظر می رسد اساس و زیر بنای فلسفه وجودی امنیّت فنّاوری اطلاعات بوده باشد.

۳-۴- بحث و نتیجه گیری

این بخش به بحث در خصوص دستاوردها و یافته های حاصل از انجام این تحقیق اختصاص دارد. اینکه نتیجه انجام این پژوهش در راستای دستیابی به اهداف اصلی آن به چه میزان توانسته است به پرسشهای مطرح شده پاسخ داده و صحّتِ فرضیّات مربوطه را تایید نماید. هدف اصلی این پژوهش رسیدن از ایمنی به امنیّت با برقراری نوعی حیات مصنوعی از طریق شبیه سازی مصونیّتهای ایمنی ذاتی و اکتسابی با کارکردی مشابه در سیستم ایمنی زیستی در جهت دستیابی به یک سیستم تشخیص نفوذ پایدار و ایمن شبکه می باشد. با در نظر گرفتن این هدف به نظر میرسد در این راستا ، تشخیص دهنده های بالغ به عنوان سنسورها و ناظرانِ فعّالِ سیستم نقش موثری در برقراری این حیات و امکان رُشد ایمنی سیستم داشته اند.

در واقع تشخیص دهنده ها به عنوان بازوان پُر توانِ سیستمِ تشخیص با مشارکت یکدیگر و به صورت کاملاً توزیع شده اقدام به شناسایی نفوذ می نمایند. ایجاد امکان مشارکتِ تشخیص دهنده ها تنها با تعریف و تخصیص پارامتری به نام A Vote ممکن بود تا بدینوسیله بتوانند تعدیلِ فضای غیر خودی و وضعیت ایمنی درون سیستم را کنترل نمایند. ارزیابی نتایج حاصل از انجام آزمایشهای نهایی در بخش دوّم از فصل سوّم نشان می دهند که روش تشخیص نفوذ هیبریدیِ پیشنهادی قادر است به مرور و با طی روند اجرای سیکلهای متوالی از حیات مصنوعی عملکرد دسته بندی خود را بهبود و ثُبات نسبی ببخشد. در ذیل به بحث در خصوص مهمترین دستاوردهای حاصل از انجام این تحقیق با ذکر جزئیات کامل آنها بر مبنای رویکرد پاسخ به پرسشهای مطرح شده پرداخته ایم.

- در رابطه با موفّقیت در ایجاد حیات مصنوعی و پاسخ به پرسش نخست باید گفت طبق نتایج بدست آمده از آزمایش نهایی و با توجه به نمودار ۱۶ و ۲۰ روش هیبریدی پیشنهادی در پنج سیکل اجرای نخست موفّق عمل نموده است بطوریکه نرخ های همبستگی و استعداد تشخیص مربوط به دو دادگانِ آزمون اوّل این موفقیت را در سه سیکل ابتدایی حیات نشان می دهند. امّا چالشی که سیستم با آن مواجه می باشد طبق نمودارهای ۱۷ و ۱۸ هر چه از میزان نمونه های عدم قطعیّت در طی روند اجرای سیکلها کمتر کاسته شده و تعداد نمونه های قطعیّت دیرتر افزایش یابند سیستم نیازمند به منابع پردازشی اضافی خواهد بود. وجود این منابع اضافی جهت پردازش منجر به کاهش زمان طی نمودن سیکلها متوالی گشته و در نتیجه بلادرنگ بودن تشخیص بهبود می یابد. در مجموع میتوان گفت روش هیبریدی پیشنهادی در طی پنج سیکل اجرای متوالی توانسته است کارکردی مشابه سیستم ایمنی زیستی داشته و دو مصونیّت ذاتی و اکتسابی را به خوبی شبیه سازی نماید. البته در این راستا چالش بلادرنگ بودن نیز مطرح بوده است.
- در خصوص تشخیص در حین یادگیری و بالعکس ، در این پایان نامه امکان شبیه سازی همروندِ دو فاز یادگیری و فاز آزمون حاصل نشد. دلیل این مسئله عدم پشتیبانی نرم افزار متلب از اجرای همروند پردازشها می باشد.

- زمان تولید و تکثیر تشخیص دهنده های بالغ بسیار زمان بر بوده بطوریکه با گذشت زمان روند صعودی نمایی دارد. (نمودار ۵-۷ از پ-الف-۶) استفاده از ایده استراتژی مختصات قطبی و کاربرد آن در فاز یادگیری نخستِ مکانیسمِ RNSA جهت پوشش موثّر حفره ها منجر به کاهش زمان و تسریع روند تولید تشخیص دهنده ها گشت. در نتیجه محصول این کاربرد موثّر را در آزمون و ارزیابی دو دادگان نخست می توان مشاهده نمود. به عبارت بهتر ، با کاربرد این استراتژی ، با کمترین تعداد تشخیص دهنده بالغ به بیشترین تاثیرِ مثبت در خروجی دسته بندی می توان دست یافت. به نحوی که این کاربرد در نرخ های تشخیص و کاهش خطا نیز موثّر بوده است به دلیل آنکه فضای غیر خودی تعدیل شده و حفره های کمتری دراین فضا باقی ماندند بنابراین تشخیص دهنده ها ضمن افزایش دقّت دسته بندی خطای مثبت کاذب را نیز کاهش داده اند.
- استراتژی پردازش در دسته بندیِ بسیاری از الگوریتم های الهام گرفته شده از طبیعت (فراابتکاری) متمرکزگِرا بوده و بر مبنای یافتن بهترین موقعیتهای مرکزیّتهاست. در این الگوریتمها ، تابعِ هدف به صورت متمرکز تصمیم میگیرد که در چه موقعیتهایی از مرکزیّت های یافت شده ، فیتنس بالاتر است. با ارزیابی های انجام شده در این پژوهش مشاهده شد که این شیوه حل مسئله در دسته بندی باینری ترافیکِ نفوذِ شبکه جواب نمی دهد.
- کشف این زیر مجموعه ویژگیها و اعمال رابطه ای دقیق بر روی ترافیک شبکه جهت نگاشت به سیگنالهای ورودی چالش بزرگی است. تاکنون دو روش معمول بدین منظور از سوی محقّقان ارائه شده که یکی استفاده از بهره اطلاعات و دیگری استفاده از دانش متخصّصان در ارزیابی ترافیک شبکه می باشد. ما در این پایان نامه روش سوّمی را با ایده پردازی بر روی یک خوشه بند مناسب ، مشخصاً الگوریتمِ سیاه چاله ارائه نموده و ارزیابی هایی را در این خصوص انجام دادیم که نتایج منفی بدنبال داشت. نتیجه ارزیابی این الگوریتم نشان داد که به دلیل آنکه همانند سایر الگوریتم های فراابتکاری با روشِ متمرکز با مسئله دسته بندی رفتار می کند در نتیجه، پتانسیل کافی جهت کشف و نگاشت (تخصیص) سیگنالهای ورودی ندارد.

نتایج استفاده از دانش متخصصان نیاز به آنالیز دقیقِ ترافیک شبکه دارد و به نظر میرسد به منظور کشف و استخراج بهترین زیر مجموعه ویژگیها جهت نگاشت به سیگنالهای ورودی می بایست به دنبال کشفِ زیر مجموعه ویژگیهایی بود که نماینده بهتری از آنومالی های مبتنی بر زمان (آنومالی ضِمنی و تجمّعی معرفی شده در [۳-۲-۲]) باشند. به این دلیل که طبق جدول ۲ توابع نگاشت در بخش [۲-۲-۳] دو سیگنالِ PAMP و خطر از دید محاسباتی معادلِ تعداد پیام های خطا در هر ثانیه و تعداد بسته های شبکه در هر ثانیه هستند که ارتباط تنگاتنگی با زمان دارند. بدین معنی که اگر با آنالیز و کشفِ زیر مجموعه ویژگیهای مرتبط با زمان درترافیک شبکه ، تابعِ نگاشتِ مناسبی پیشنهاد و ارائه گردد می توان سیگنالهای ورودی را به مراتب با دقّت بیشتری تخصیص داد. در نتیجه به موجب تاثیر اعمال تابعِ نگاشتِ مناسب ، خروجی دسته بندی DCA نیز بهبود خواهد یافت.

در خط نخست دفاعی ، مطابق مباحث مطرح شده در بخش نخست فصل سوّم ، مکانیسمِ DCA سعی دارد از طریق نمونه برداری آنتی ژنها و پردازش سیگنالهای وضعیت دریافت شده از بافت های سلّولی و با عملیاتِ توزیع شده و مشارکتی بوسیله سلولهای دندریت، احتمالِ وجود خطرِ نفوذ را پیش بینی نماید. با دقّت در مفهوم تئوری خطر و سازوکار مصونیّت ذاتی مشاهده شد که مکانیسمِ سلولهای دندریت در نهایت همان کارِ یک دسته بند نظارت نشده / نیمه نظارت شده را انجام می دهد. این سیستم با واکنش سریع خود ضمن پیش بینی خطر همواره در شناسایی عوامل نفوذی خطاهایی نیز دارد بطوریکه ممکن است عوامل خودی را به اشتباه به عنوان عاملِ نفوذی شناسایی نمایند. به عبارت بهتر ، خطای مثبت کاذب آنها بالاست و البته این رفتار کاملاً طبیعی میباشد.

۴-۴- پیشنهادهایی برای کارهای تحقیقاتی آینده

- تعیین صحیح دو پارامترِ وزنی β و α یک چالش است. در آزمایشهای نهایی مشاهده گردید که تغییر در انتخاب وزنهای مناسب برای این دو پارامتر مهم ، در خروجی دسته بندی و تعیین اینکه برچسب نهایی نمونه عدم قطعیّت آیا بزرگتر از ۰,۴۹۹۹ است یا کمتر از آن بسیار تاثیر گذار است. علاوه بر این کسب تجربه لازم از دسته بندیِ خطوط دفاعی سیکلهای قبل، منجر به تعیین بهتر این وزنها خواهد گردید. بنابراین پیشنهاد می گردد که در تحقیقات آتی بر روی روابطی به منظور وابسته نمودن این دو وزن به تجربه خروجی دسته بندی حاصل از سیکلهای اجرای قبلِ سیستم بیشتر کار شود.
- روابط ارائه شده در ایده استفاده از خوشه بند سیاه چاله جهت کشف و تخصیص سیگنالهای ورودی الگوریتم سلّولهای دندریت کارآمدی لازم را ندارند بدین منظور استفاده از ایده استفاده از خوشه بند های غیر متمرکز گِرا جهت کشف و تخصیص سیگنالها پیشنهاد می گردد.
- به منظور پیاده سازی عملی روشِ هیبریدیِ پیشنهادی و اجرای سیکلهای طولانی تر از حیاتِ مصنوعی، منابع پردازشی و حافظه بالایی لازم است تا سیستم بتواند به سرعت با آنتی ژنها زیست نموده و در تخصیص احتمالات برچسب ها به قطعیّت لازم دست یابد.
- فاز یادگیری اوّلیه و ثانویه بهمراه فاز آزمون خط دوّم دفاعی می بایست به صورت همروند و کاملاً موازی با هم پیش بروند تا بتوان به هدف تشخیص در حین یادگیری و بالعکس دست یافت. پیاده سازی عملی این فازها به طور اجرای همروند در نرم افزار متلب ممکن نبود که نیازمند پیاده سازی در محیط ابری ، پیاده سازی سخت افزاری و یا استفاده از یک GPU جهت پردازش موازی میباشد. همچنین استفاده از استراتژی مختصات قطبی در فاز یادگیری اوّلیه در خط دوّم دفاعی نیز نیازمند منابع کافی حافظه می باشد. این ایده جهت کاهش زمان تولید تشخیص دهنده های بالغ پیشنهاد گردید امّا در شبیه سازی متلب نمی توان به شکل بلادرنگ آنرا پیاده کرد. بنابراین به نظر میرسد استفاده از چنین منابع پردازشی قوی بتواند چالش های مذکور را رفع نماید.

- رقابت میان دو فاز یادگیری اولیه و ثانویه که به ترتیب فضاهای غیر خودی و خودی را تعدیل می نمایند نقش موثری در تعیین مرزبندی صحیح میان این دو فضا دارد. در صورتیکه این مرز به درستی تعیین نگردد شاهد رخداد خطاهای مثبت و منفی کاذب خواهیم بود. در این پایان نامه ما این رقابت را اصطاحاً مانُور نامیده ایم زیرا مکانیسمِ این دو فاز یادگیری در خط دوّم در تسخیر فضا کاملاً مشابه با مانُورهای نظامی می باشد که در آن هر دو طرف در واقع خودی ها هستند که در نقش دشمن فرضی منجر به این می شوند که سیستم با آزمون و خطا و تجارب حاصل از این نبرد بتواند بلوغ لازم را در نبردهای واقعی کسب نموده و نتیجه آنرا مجدّداً جهت یادگیری بازخورد نماید. بنابراین ایده ای که به نظر میرسد در صورت پیاده سازی، موفق واقع گردد تاثیر دادنِ نتایج کسب تجربه با پیاده سازی نبرد میان تشخیص دهنده ها و عوامل خودی می باشد.
- تعیین شعاع فضای خودی نرمال/ آنومالی همواره یک چالش اساسی برای الگوریتم انتخاب منفی حقیقی مبنا مطرح بوده بطوریکه نقش مستقیمی در نرخ های خروجی دسته بندی دارد. بدین منظور و به جهت رفع این چالش در کارهای تحقیقاتی آینده ، پیشنهاد میگردد که تعیین این شعاع به صورت خودکار و هوشمندانه توسط خود سیستم تشخیص صورت پذیرد. بدین صورت که معیاری مانند میانه/ میانگین/ مینیممِ فواصل بین نمونه های خودی را می توان به عنوان ایده ای در این راستا مطرح نمود و به عنوان استراتژی پیشنهادی در .تعیین پویای شعاع های خودی از آن بهره برد.
- به منظور کاهش زمان تشخیص و افزایش کیفیّتِ دسته بندی ، پیشنهاد می شود پارامتر A Vote پیوسته سازی شده و بروز رسانی مقدار آن به جای گسسته بودن به طریقی به پردازش سیگنالها در خط نخست وابسته گردد. بطوریکه اگر چنانچه سیستم با خطر نفوذ مواجه باشد ، مقدار سیگنال آلارمِ فرضی که از بافت سلّولی مشخص و بافت های همجوار از خط دفاعی اوّل به تشخیص دهنده های موجود در همان بافت (ها) در خط دوّم ارسال می گردد منجر به تغییر هوشمند پارامتر A Vote تشخیص دهنده (های) بالغی گردد که در موقعیت آن بافت سلولی حضور دارند.

به عبارت بهتر ، در بروز رسانی مقدار این پارامتر و بحث مدیریت حافظه تشخیص دهنده های بالغ ، میبایست همواره تغییر پارامترهای A Vote بافت های همجوار و همچنین سیگنالهای خروجی دریافت شده از سلولهای دندریت موجود در این بافت ها نیز تاثیر گذار باشد و این به معنای پیاده سازی دقیقِ نوعی سیستم رأی گیری مشارکتی می باشد. بگونه ای که آرای تشکیل دهنده در تعیین مقدار این پارامتر ، مقادیر سیگنالهای خروجی منعکس شده از خط دفاعی نخست باضافه مقادیر پارامترهای A Vote مجاور همگی میتوانند تعیین کننده باشند. البته تعیین رابطه ای وزنی به منظور محاسبه این سیستم رأی گیری مشارکتی کار آسانی نخواهد بود و به نظر می رسد در تعیین آن استفاده از فواصل معیار مانند فاصله اقلیدسی، دور از انتظار نباشد.

منابع

مقالات منتشر شده در مجلات و کتب

[1]. Wu, H. (2017). **Artificial Immune Systems Based Intrusion Detection Algorithm for Cloud Environment**. *Boletín Técnico*, *55*(1), 11-17.

[2]. Fernandes, D. A., Freire, M. M., Fazendeiro, P. A., & Inácio, P. R. (2017). **Applications of artificial immune systems to computer security: A survey.** *Journal of Information Security and Applications*, *35*, 138-159.

[3]. Okamoto, T., & Tarao, M. (2016). **Toward an artificial immune server against cyber-attacks**. *Artificial Life and Robotics*, *21*(3), 351-356.

[4]. Hosseinpour, F., Amoli, P. V., Farahnakian, F., Plosila, J., & Hämäläinen, T. (2014). **Artificial immune system based intrusion detection: innate immunity using an unsupervised learning approach**. *International Journal of Digital Content Technology and its Applications*, *8*(5), 1.

[5]. Dal, D., Abraham, S., Abraham, A., Sanyal, S., & Sanglikar, M. (2008, June). **Evolution induced secondary immunity: An artificial immune system based intrusion detection system**. In *Computer Information Systems and Industrial Management Applications, 2008. CISIM'08. 7th* (pp. 65-70). IEEE.

[6]. Gong, M., Zhang, J., Ma, J., & Jiao, L. (2012). **An efficient negative selection algorithm with further training for anomaly detection**. *Knowledge-Based Systems*, *30*, 185-191.

[7]. Ramdane, C., & Chikhi, S. (2014). **A new negative selection algorithm for adaptive network intrusion detection system**. *International Journal of Information Security and Privacy (IJISP)*, *8*(4), 1-25.

[8]. Xiao, X., Li, T., & Zhang, R. (2015). **An immune optimization based real-valued negative selection algorithm**. *Applied Intelligence*, *42*(2), 289-302.

[9]. Seresht, N. A., & Azmi, R. (2014). **MAIS-IDS: A distributed intrusion detection system using multi-agent AIS approach**. *Engineering Applications of Artificial Intelligence*, *35*, 286-298.

[10]. Saurabh, P., & Verma, B. (2016). **An efficient proactive artificial immune system based anomaly detection and prevention system**. *Expert Systems with Applications*, *60*, 311-320.

[11]. Fouladvand, S., Osareh, A., Shadgar, B., Pavone, M., & Sharafi, S. (2017). **DENSA: An effective negative selection algorithm with flexible boundaries for self-space and dynamic number of detectors**. *Engineering Applications of Artificial Intelligence*, *62*, 359-372.

[12]. Ramdane, C., & Chikhi, S. (2017). **Negative selection algorithm: recent improvements and its application in intrusion detection system**. *Int. J. Comput. Acad. Res.(IJCAR)*, *6*(2), 20-30.

[13]. Hatamlou, A. (2013). **Black hole: A new heuristic optimization approach for data clustering**. *Information sciences*, *222*, 175-184.

[14]. Eesa, A. S., Orman, Z., & Brifcani, A. M. A. (2015**). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems**. *Expert Systems with Applications*, *42*(5), 2670-2679.

[15]. Chelly, Z., & Elouedi, Z. (2016). **A survey of the dendritic cell algorithm**. *Knowledge and Information Systems*, *48*(3), 505-535.

[16]. Moustafa, N., & Slay, J. (2015, November). **The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems**. In *Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), 2015 4th International Workshop on* (pp. 25-31). IEEE.

[17]. Mukhopadhyay, M. (2014). **A brief survey on bio inspired optimization algorithms for molecular docking**. *International Journal of Advances in Engineering & Technology*, *7*(3), 868.

[18]. Meisel, M., Pappas, V., & Zhang, L. (2010). **A taxonomy of biologically inspired research in computer networking**. *Computer Networks*, *54*(6), 901-916.

[19]. Dressler, F., & Akan, O. B. (2010). **Bio-inspired networking: from theory to practice**. *IEEE Communications Magazine*, *48*(11).

[20]. Hassanien, A. E., Kim, T. H., Kacprzyk, J., & Awad, A. I. (Eds.). (2014). ***Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations*** (Vol. 70). Springer.

[21]. Souici-Meslati, L., & Zekri, M. (2016). **Immunological Approach for Intrusion Detection**. *REVUE AFRICAINE DE LA RECHERCHE EN INFORMATIQUE ET MATHÉMATIQUES APPLIQUÉES*, *17*.

[22]. Yin, C., Ma, L., & Feng, L. (2017**). Towards accurate intrusion detection based on improved clonal selection algorithm**. *Multimedia Tools and Applications*, *76*(19), 19397-19410.

[23]. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). **A survey of network anomaly detection techniques**. *Journal of Network and Computer Applications*, *60*, 19-31.

[24]. Chandola, V., Banerjee, A., & Kumar, V. (2009). **Anomaly detection: A survey**. *ACM computing surveys (CSUR)*, *41*(3), 15.

[25]. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). **Anomaly-based network intrusion detection: Techniques, systems and challenges**. *computers & security*, *28*(1-2), 18-28.

[26]. Liao, H. J., Lin, C. H. R., Lin, Y. C., & Tung, K. Y. (2013). **Intrusion detection system: A comprehensive review**. *Journal of Network and Computer Applications*, *36*(1), 16-24.

[27]. Gu, G., Fogla, P., Dagon, D., Lee, W., & Skorić, B. (2006, March). **Measuring intrusion detection capability: an information-theoretic approach**. In *Proceedings of the 2006 ACM Symposium on Information, computer and communications security* (pp. 90-101). ACM.

[28]. Kuang, F., Xu, W., & Zhang, S. (2014). **A novel hybrid KPCA and SVM with GA model for intrusion detection**. *Applied Soft Computing*, *18*, 178-184.

[29]. Al-Enezi, J.R., Abbod, M.F. and Alsharhan, S., 2010, May. **Artificial immune systems-models, algorithms and applications,** *International Journal of Research and Reviews in Applied Sciences (IJRRAS)*, 3(2): 118-131. https://bura.brunel.ac.uk/handle/2438/4643

[30]. Viegas, E. K., Santin, A. O., & Oliveira, L. S. (2017). **Toward a reliable anomaly-based intrusion detection in real-world environments**. *Computer Networks*, *127*, 200-216.

[31], Dutt, I., Borah, S., & Maitra, I. (2016). **Intrusion Detection System using Artificial Immune System**. *International Journal of Computer Applications*, *144*(12).

[32]. Burlakov, M. E., & Osipov, M. N. (2016). **Research the behavior of elements in artificial immune system for intrusion detection systems in information networks**. In *CEUR Workshop Proceedings* (Vol. 1638, pp. 895-901).

[33]. Čisar, P., Čisar, S. M., & Markoski, B. (2014). **Implementation of immunological algorithms in solving optimization problems**. *Acta Polytechnica Hungarica*, *11*(4).

[34]. Yang, H., Li, T., Hu, X., Wang, F., & Zou, Y. (2014). **A survey of artificial immune system based intrusion detection**. *The Scientific World Journal*, *2014*.

[35]. Al-Sheshtawi, K. A., Abdul-Kader, H. M., & Ismail, N. A. (2010). **Artificial immune clonal selection classification algorithms for classifying malware and benign processes using API call sequences**. *International Journal of Computer Science and Network Security*, *10*(4), 31-39.

[36]. Olubadeji, B., & Adetunmbi, A. O. (2016). **Auto-Immunity Dendritic Cell Algorithm**. *International Journal of Computer Applications*, *137*(2), 10-17.

[37]. Ahmad, A., Idris, N. B., & Kama, M. N. (2017). **CloudIDS: Cloud Intrusion Detection Model Inspired by Dendritic Cell Mechanism**. *International Journal of Communication Networks and Information Security (IJCNIS)*, *9*(1).

[38]. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). **Network anomaly detection: methods, systems and tools**. *IEEE communications surveys & tutorials*, *16*(1), 303-336.

[39]. Iglesias, F., & Zseby, T. (2015). **Analysis of network traffic features for anomaly detection**. *Machine Learning*, *101*(1-3), 59-84.

[40]. Cua, A.S. ed., 2013. **Encyclopedia of Chinese philosophy**, Routledge.

[41]. Piotrowski, A. P., Napiorkowski, J. J., & Rowinski, P. M. (2014). **How novel is the "novel" black hole optimization approach?** *Information Sciences*, *267*, 191-200.

[42]. Yaghoobi, S., & Mojallali, H. (2016). **Modified Black Hole Algorithm with Genetic Operators**. *International Journal of Computational Intelligence Systems*, *9*(4), 652-665.

[43]. Kumar, S., Datta, D., & Singh, S. K. (2015). **Black hole algorithm and its applications**. In *Computational intelligence applications in modeling and control* (pp. 147-170). Springer, Cham.

[44]. Alsharafi, WM., Omar, MN. (2015). **a detector generating algorithm for intrusion detection inspired by artificial immune system**, ARPN journal of engineering and applied sciences, 10(2), February 2015.

[45]. J.L Santanelli and et al., **Network Intrusion detection Using Danger Theory and Genetic Algorithm**, *Intelligent Systems Design and Applications*, Advances in Intelligent Systems and Computing 557, DOI 10.1007/978-3-319-53480-0 39

پایان نامه های دانشجویانِ دانشگاه های خارجی

[46]. Greensmith, J. (2007). **The dendritic cell algorithm** (*Doctoral dissertation*, University of Nottingham, UK).

[47]. Gu, F. (2011). **Theoretical and empirical extensions of the dendritic cell algorithm** (*Doctoral dissertation*, University of Nottingham).

مقالاتِ مندرج در کنفرانسهای بین المللیِ خارجی

[48]. Wlodarczak, P. (2017, April). **Cyber Immunity**. In *International Conference on Bioinformatics and Biomedical Engineering* (pp. 199-208). Springer, Cham.

[49]. Li, Z., & Pi, D. (2017, September). **Data Clustering Algorithm Based on Artificial Immune Network**. In *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 516-527). Springer, Singapore.

[50]. Kayacik, H. G., Zincir-Heywood, A. N., & Heywood, M. I. (2005, October). **Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets**. In *Proceedings of the third annual conference on privacy, security and trust*.

[51]. Saurabh, P., Verma, B., & Sharma, S. (2012, October). **Biologically inspired computer security system: the way ahead**. In *International Conference on Security in Computer Networks and Distributed Systems* (pp. 474-484). Springer, Berlin, Heidelberg.

[52]. Ma, W., Tran, D., & Sharma, D. (2008, August). **Negative selection with antigen feedback in intrusion detection**. In *International Conference on Artificial Immune Systems* (pp. 200-209). Springer, Berlin, Heidelberg.

[53]. Ge, Y., Liang, H., Chen, L., & Zhang, Q. (2015). **The Designation of Bio-Inspired Intrusion Detection System Model in Cloud Computing Based on Machine Learning**. In *International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE)* (pp. 1932-1937).

[54]. Zhang, X., An, J., Wang, Y., & Liu, W. (2017, May). **The application of immune clone algorithm in network intrusion detection**. In *Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on* (pp. 619-622). IEEE.

[55]. Greensmith, J., & Aickelin, U. (2008, August). **The deterministic dendritic cell algorithm**. In *International Conference on Artificial Immune Systems* (pp. 291-302). Springer, Berlin, Heidelberg.

[56]. Gu, F., Feyereisl, J., Oates, R., Reps, J., Greensmith, J., & Aickelin, U. (2011, July). **Quiet in class: classification, noise and the dendritic cell algorithm**. In *International Conference on Artificial Immune Systems* (pp. 173-186). Springer, Berlin, Heidelberg.

[57]. Gupta, H., Gupta, A., Gupta, S. K., Nayak, P., & Shrivastava, T. (2016, December). **How effective is black hole algorithm?.** In *Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on* (pp. 474-478). IEEE.

پادکستهای علمی

[58]. Andersen.P. (Producer), Director. (March 19, 2012). **The Immune System,** [Video podcast]. US. Retrieved from http://www.bozemanscience.com

[59].Nagpal.R, W. e. (Director). (Jun Monday, 29, 2015), **Bio inspired Robotics: Softer, Smarter, Safer**, Wyss Institute's 6th Annual Symposium. [Video podcast]. US. Retrieved from http://wyss.harvard.edu/viewpage/582

موضوعات مندرج در پایگاه های اینترنتی

[60]. **Immunity Cell-mediated.** (n.d.). Retrieved from wikipedia: http://en.wikipedia.org/wiki/Cell-mediated

[61]. Aitkin.J, Andrews.P, et al. (2013), **Basic Immune Inspired Algorithms .AISWEB. The Online Home of Artificial Immune Systems**, http://www.artificial-immune-systems.org/contact.shtml

گزارشهای فنّی

[62]. Brownlee, J. (2005). **Clonal selection theory & CLONALG-the clonal selection classification algorithm (CSCA)**. *Swinburne University of Technology*.

ارجاع به نظراتِ علمی افراد برجسته

[63]. Jain, R. (2007). **Intrusion detection systems**. *WUSTL class lecture*.

[64]. P.F.Zabrodskii, September 2017, (Full Professor – Saratov State Medical University, Russia), **public corresponding,** https://www.researchgate.net/post/Immunity_versus_Security .(Access data : 2018 May).

[۶۵]. مرتضی امینی ، نیمسال اوّل ۹۱-۹۲، (مرکز امنیّت داده و شبکه شریف ، دانشگاه صنعتی شریف ، ایران) ، **امنیّت داده و شبکه- سیستم تشخیص نفوذ** ، (اسلایدِ درسی)،

http://dnsl.ce.sharif.edu

پیوست ها

پیوست الف – توضیحات بیشترِ مطالب

بخشِ نخست - اصل دوّمِ امنیّت (دفاع در عمق)

دوّمین اصل از اصول سیزده گانه امنیّت بحث دفاع در عمق را مطرح می کند. طبق این اصل ، امنیّت باید در چندین لایه برقرار شود به صورتی که هر لایه لایه دیگر را پوشش دهد. با رعایت این اصل اگر یک لایه دفاعی با شکست مواجه شود لایه های دیگر می توانند جلوی حملات را بگیرند.

ایده دفاع در عمق با انواع استراتژیهای دفاعی ، مدیریت ریسک را انجام می دهد. اگر یک لایه دفاعی نامناسب باشد و از دسترس خارج شود . امیدواریم که لایه دفاعی دیگر از نقص به صورت کامل دفاع کند. از اینرو ، افزونگی و لایه بندی در امنیت فکر خوبی است.

بخش دوّم - ارزیابیِ مقایسه ایِ سه الگوریتمِ فراابتکاری

در این قسمت ، به منظور بررسی کیفیت دسته بندی نهایی ترافیک شبکه ، عملکردِ برخی از الگوریتم ها و متدهای منتخب در نرم افزار متلب در دادگان ترافیک شبکه UNSW-NB15 ارزیابی مقایسه ای نموده ایم. بدین منظور آزمایشِ زیر که به ترتیب بدون نظارت و با نظارت می باشند انجام شده اند. نتیجه این آزمایش تفاوت میان شیوه حلّ مسئله با دسته بندی باینری ترافیک شبکه در الگوریتمهای ارزیابی شده و همچنین الگوریتم سلولهای دندریت استاندارد (دسته بندیِ متمرکزگِرا و توزیع شده) را آشکارتر می نماید.

۱-۲- آزمایش

برای انجام این آزمایش سه الگوریتم فراابتکاری[1] PSO, GA, BHA انتخاب شدند. روش حل مسئله دسته بندی باینری در هر سه متد تقریباً یکسان بوده و به شیوه ی متمرکزگرا مسئله را حل می کنند. تفاوت آنها در استراتژی بهینه سازی موقعیت های مرکزیتهای یافت شده می باشد که تحت عنوان تابع هدف/فیتنس نیز شناخته می شود.

بنابراین در تمامی الگوریتم های فرا ابتکاری ، تابع ارزیابی فیتنس ، نقش موثری در تعیین بهترین موقعیّتِ مرکزیّتها دارد. بطوریکه در هر فازِ تکرارِ الگوریتم ، موقعیتهای هر نمونه/ ستاره/ ذرّه به عنوان مرکزیّتِ پیشنهادی آزمون شده و خوشه بندی انجام می شود در نهایت تابع فیتنس بر حسب معیار ارزیابی خود که معمولاً یک رابطه وزنیِ با نرخهای تشخیص و خطا میباشد تصمیم میگیرد و هزینه یا فیتنسِ موقعیتهای انتخاب شده را در خوشه بندی نهایی محاسبه و ارزیابی می کند. از این نظر انتخاب تابع فیتنس / هزینه و همچنین تعیین معیار مناسب در این تابع نقش بسیار موثری در عملکرد نهایی خواهد داشت. امّا در مورد مسائلی مانند دسته بندی باینری ترافیک شبکه ، که توزیع داده

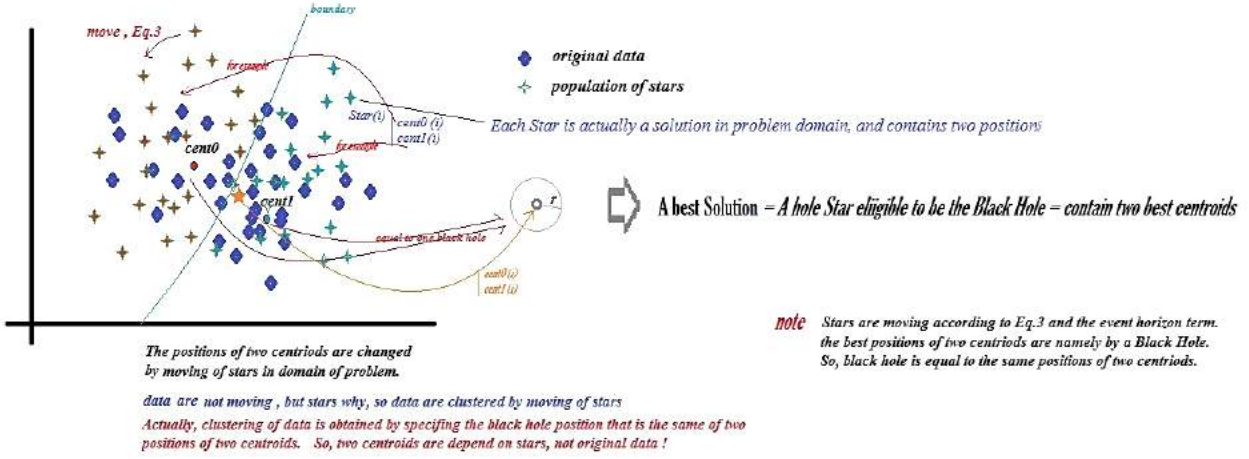
[1] The Black Hole Optimization Algorithm (in Clustering problem)

ها در فضای ابعاد مسئله به گونه ایست که ممکن است به جای دو خوشه ، چندین خوشه آنومالی یا نرمال داشته باشیم پس با ارزیابیِ نتایج آزمایشات اثبات می کنیم که کاربرد الگوریتم های فرا ابتکاری برای حلّ مسئله دسته بندی باینری در تشخیص نفوذ انتخاب مطلوبی نیست.

- ❖ الگوریتم ازدحام ذرّات
- ❖ الگوریتم ژنتیک
- ❖ الگوریتم حفره سیاه

بخش سوّم – الگوریتم سیاه چاله

الگوریتم حفره سیاه ، از پدیده بَلعیدن ستاره ها در فضا توسط سیاه چاله الهام گرفته شده است. این الگوریتم برای اولین بار در سال ۲۰۱۳ در مقاله [۱۳] ارائه گردید. توضیحات مربوط به این الگوریتم در بخش بخش بعدی ارائه شده است. برخی مانند نویسنده مقاله [۴۱] معتقدند که این الگوریتم شکل ساده شده و در واقع حالت سکون " الگوریتم ازدحام ذرّات" است و حتّی در مقالات [۵۶] الهام از بیولوژیک بودن آن مورد انتقاد قرار گرفته شده است. اینفوگرافیک زیر به وضوح سازوکار این الگوریتم جدید را نشان می دهد.

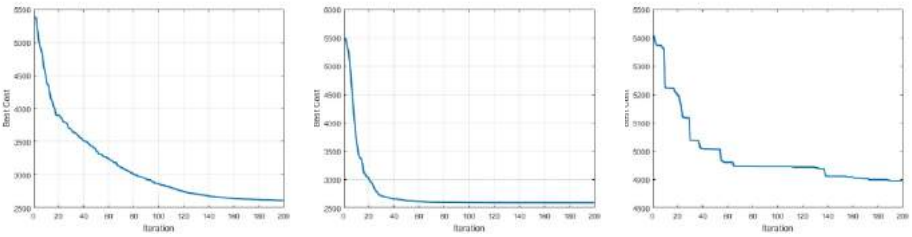


شکل ۱ - اینفوگرافیکِ مکانیسمِ خوشه بندیِ حفره سیاه ، برگرفته از شبیه سازی منتشر شده [۲]

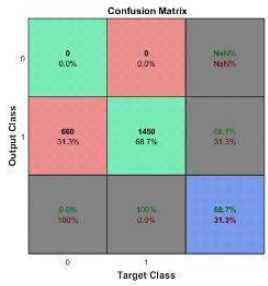
نتایج

نتایج آزمایش نشان می دهد که در تمام آزمون ها ، میزان خطای مثبت اشتباه بسیار بالاست. حتی ما تلاش کردیم و مجدداً پارامترها را تغییر دادیم اما ، میزان خطای مثبت کاهش چشمگیری نیافت. حتی با تغییر تابع فیتنس نیز وضع به همین گونه است. تابع فیتنس برای هر سه الگوریتم ، فاصله درون خوشه ای[1] تعیین شد.

[1] Fitness function: Sum of Within - Cluster Distance (WCD)



نمودار - 1 نتایج الگوریتم به ترتیب از راست با پارامترهای پیشفرض[1]



نمودار ۲ - ماتریس در هم ریختگی برای تمام الگوریتمها با پارامترهای پیشفرض

بخش چهارّم - اهمیّتِ انتخابِ ویژگی (دفاع غیر مستقیم)

در این بخش ضمنِ بیانِ مقدّمه ای در رابطه با انتخابِ ویژگی و اهمیّت آن ، به بررسی استراتژی جستجوی ده پا در بازتاب و انعکاس نور تابیده شده به ماهیچه های پوست این جاندار پرداخته و کاربردِ مکانیسمِ انتخاب ویژگیِ دَه -پا را در انتخاب موثّرترین ویژگیهای دادگان ترافیکِ شبکه ارزیابی نموده ایم.

مهمترین نیازمندیهایی که برای یک سیستم تشخیص نفوذ در شبکه مطرح می باشند عبارتند از افزایش دقت و سرعت فرایند تشخیص در حین یادگیری ، کار با داده حجیم ، دانستن مجموعه ویژگیهای بهینه مرتبط با هر نوع حمله. به منظور افزایش دقت و سرعت طبق تحقیقات انجام شده موثرترین راهکار همان کاهش حجم داده از طریق انتخاب بهترین و بهینه ترین ویژگیها و حتی نمونه هاست. انتخاب و استخراج ویژگی، یک تکنیک، یادگیری ماشینِ ضروری است که در ایجاد سیستم های دسته بندی کارآمد موثر است. وقتی که برای کاهش ویژگی استفاده شود نتیجه آن کاهش هزینه های محاسباتی و دسته بندی بهتر است [۱۰۵]. مسئله ی انتخاب ویژگی تلاش می کند تا زیر مجموعه حداقلی به اندازه $m < n$ را از بین n ویژگی پیدا کند تا بدین طریق عملکردِ دسته بندها را افزایش دهد. در واقع در این وضعیّت ، ابعادِ مسئله کاهش یافته است.

هر نوع حمله ای در شبکه ، از زیر مجموعه ای خاص از ویژگیهای خاص نَشئَت میگیرد. به همین دلیل کاهشِ ابعاد فرایندِ بسیار مهمی می باشد. از طرفی برای انتخابِ بهینه ترین ویژگیهای مرتبط با هر نوع حمله می بایست استراتژی جستجوی مناسبی اتخاذ گردد که در کمترین زمانِ ممکن بتواند ویژگیهای مناسب را انتخاب و حجم مجموعه ی داده را کاهش دهد. به طور کلی رویکردهای مختلفی برای فرایندِ انتخابِ ویژگی در تشخیص نفوذ به کار رفته اند. در این میان، الگوریتمهای مختلفی با رویکردهای فیلتر و راپر و حتی ترکیبی وجود دارند. مسئله اساسی که در زمینه این الگوریتمها وجود دارد انتخاب رویکرد صحیح ، بسته به کاربرد در مسئله دسته بندی در تشخیص نفوذ است.

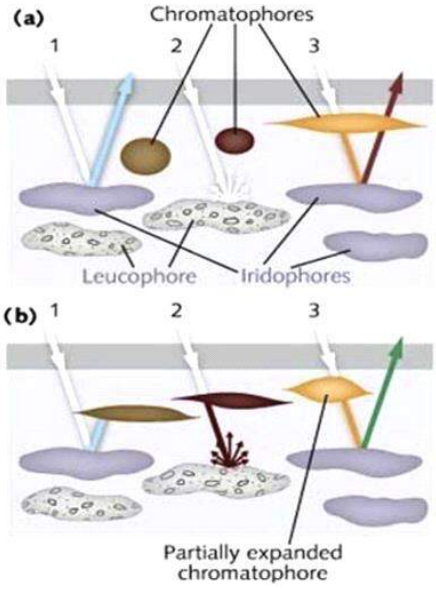
[1] for ga (Maxiteration = 200. nPop = 100. pc = 0.8. pm = 0.3.gamma = 0.2. mu = 0.2.beta = 8). for pso ($\phi 1 = \phi 2 = 2.05$)

بطوریکه هر نوع تکنیک دسته بندی و کلاً تکنیک های به کار رفته در تشخیص آنومالی با رویکرد و استراتژی خاصی از انتخاب ویژگی متناسب بوده و منجر به خروجی بهینه می شوند. این مسئله موجب شده محققان دراین زمینه هر بار و در هر آزمایش رویکردهای مختلف تشخیص را با رویکردهای مختلف انتخاب ویژگی ترکیب نموده و پیشنهاداتی را ارائه دهند. اما نکته ای که وجود دارد این است که بیشتر این رویکردها حتّی با ترکیب در فرایند انتخاب ویژگی نیز بهینه موفق عمل نکرده اند. این ضعف عمده تکنیکهای کاهش ابعاد ساخته انسان است.

طبیعت و رفتار بیولوژیکی موجودات زنده در آن ، علاوه بر اینکه منشاء کافی برای ایده پردازی از الگوهای سیستم دفاعی آنها جهت کاربرد در زمینه های مختلف تشخیص نفوذ هستند بلکه بسیاری از ویژگیهای جانوران وجود دارند که در حل مسائل به کار رفته اند. بنابراین الهام از سیستم طبیعیِ موجوداتِ زنده در طبیعت و تقلید از غریزه و حسّ مشارکتیِ آنان در حلّ مسائل منبعی سرشار از ایده پردازی را برای محققان فراهم ساخته است. از جمله مسئله استراتژی جستجو و انتخاب بهینه ترین ویژگیها در زمینه تشخیص آنومالیِ شبکه می باشد.

۱-۴- استراتژی جستجوی دَه - پا

دَه - پا[1] ، نام جانوری شبیه به هشت پا در دریاست که نور تابیده شده به سطح پوست و ماهیچه های بدن خود را با مکانیزمهای بسیار پیچیده و مخفی در لایه های زیرین پوست خود پردازش نموده و انعکاس می دهد. رنگ ها و الگوهای ده پا بوسیله نور منعکس شده از سه لایه مختلف پوست رخ می دهند. این رفتار تغییر رنگ نور بازتابیده شده ایده ای است که برای حل بسیاری از مسائل بهینه سازی به کار می رود.



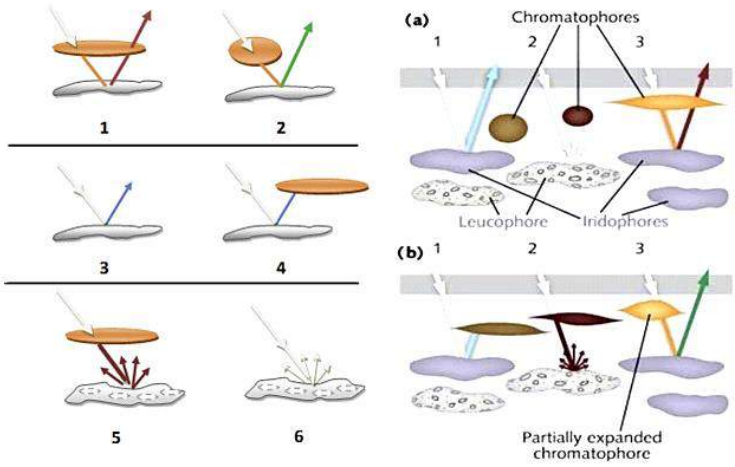
انعکاس از زوایای مختلف موجب می شود تا نور به محض تابش به سطوح مختلف زیر پوست به دلیل اینکه این از ویژگیهای مهم این متد اینست که در فازهای مختلف تکرار الگوریتم ، وابستگی زیادی به تابع ارزیابی و هزینه محاسبه شده فاز قبلی دارد. بدین معنی که از نظر عملکرد و انتخاب بهترین ویژگیها هر فاز به فاز قبلی خود وابسته است.

دَه - پا در زیر پوست خود سه ماهیچه دارد که عمل انقباض و انبساط را انجام می دهند. هر ماهیچه حاوی سلولهای زیادی است. نور وقتی که به این ماهیچه ها می تابد ، بسته به زاویه تابش R و اینکه به کدام نوع ماهیچه [2] در کدام لایه پوستی و کدام بافت سلولی تابیده ، در مجموع شش وضعیت کلی ممکن است برای نور رخ دهد که منجر به تغییر رنگ آن شود.

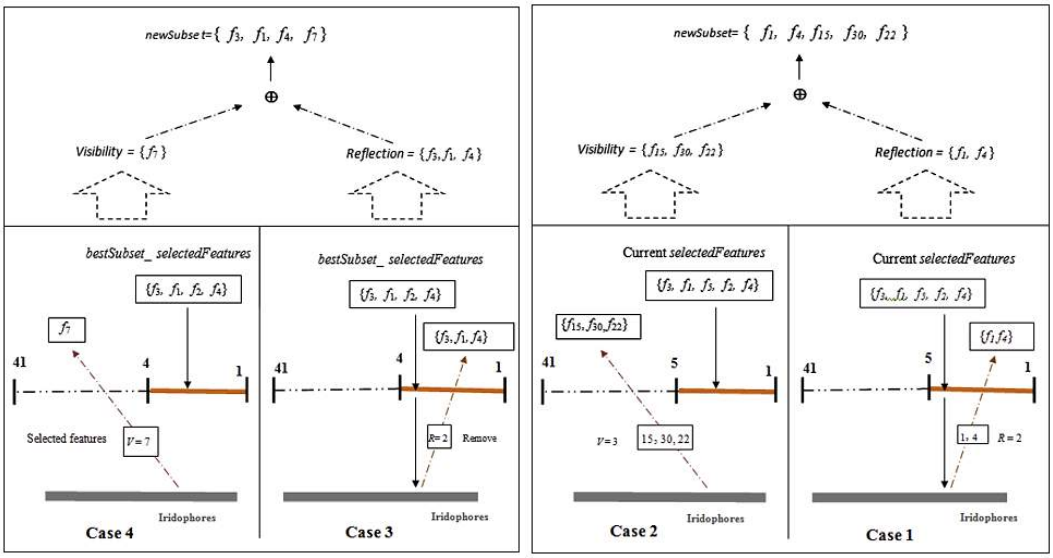
[1] Cuttlefish
[2] Chromatophores, Iridophores, leucophores

دو تابع انعکاس و بازتاب از این سازوکار تقلید می شود. این بافت های سلولی را می توان به عنوان بردارهای ویژگی کاندید در نظر گرفت. در حقیقت ، دو این دو تابع با زاویه تابش و انعکاس V نوری که نمایان می شود ، استراتژی جستجوی بهینه ترین بردار ویژگی را ترتیب می دهد. شکل زیر سه اثر پرتو نور متمایز را نشان می دهد که این جانور می تواند رنگ آمیزی بازتابی را با آنها تغییر دهد.

این وضعیت ها به گونه ای هستند که با هر بار تابش نور ورودی به ماهیچه ها در سطوح مختلف ، سلولی با مختصات مشخص بر روی بافت سلولی ماهیچه انتخاب می شود اما با شیوه های مختلف این کار صورت میگیرد. بطوریکه هر کدام از این سه ماهیچه با دو فرایند انقباض و انبساط، قادرند رنگ یا زاویه بازتابِ نور تابانده شده را تغییر داده و باصطلاح انتخاب ویژگی انجام دهند. سلولهای بازتابنده نور در درون اندامهای بدن جانور مواظبت می شوند.

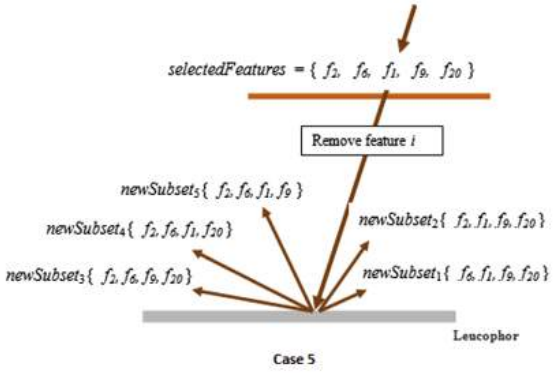


شکل ۳ – اثرِ متفاوتِ بازتاب و انعکاس نور تابیده شده به سه ماهیچه ی زیر پوستِ دَه – پا و انقباض و انبساط ماهیچه ها



شکل ۵ - فرایند تابش و انعکاس نور به ماهیچه های دَه – پا

شکل ۵ فرایند تابش و انعکاس نور در فاز اول را نشان می دهد. مطابق این شکل در هر تکرار این فاز ، از بین ویژگیهای بردار ویژگی انتخاب شده ، به صورت تصادفی ویژگی انتخاب می شود و از بین زیر مجموعه ویژگیهای انتخاب نشده بردار مذکور نیز تعدادی ویژگی طبق رابطه مربوطه می شوند تا زیر مجموعه جدیدی ایجاد گردد. زیر مجموعه ویژگی انتخاب شده از عمل انعکاس نور تابیده شده به سطح ماهیچه لایه اول پوست دَه پا الهام گرفته شده است. زیر مجموعه ویژگی انتخاب شده نیز عمل بازتاب را تداعی می کند. الگوریتم در فاز ۵ جستجوی محلی انجام می دهد.



Algorithm - Cuttlefish Algorithm Ref. [10]

Step 1: Randomly initialize the population of solutions; identify the most feasible solution and the average value of the best solution's points.
Step 2: Generate a new solution using the reflection and the visibility of pattern.
Step 3: Generate new solutions using the reflection of light and visibility of matching pattern.
Step 4: Generate random solution using reflection of incoming light

شکل ۶ – شبه کد الگوریتمِ جستجو و انتخاب ویژگی دَه - پا

۲-۴- آزمایش

ما برای انجام این آزمایش از دَه درصد دادگان نفوذ استاندارد استفاده کردیم. به دلیل حجم بسیار بالای این دادگان و عدم امکان پردازش کل این داده عظیم در نرم افزار مطلب ، در حدود ۰,۴ درصد از این دادگان را برای ایجاد زیر مجموعه یادگیری با دقت خاصی و به شکل تصادفی یکنواخت انتخاب نمودیم. جزئیات استخراج نمونه های تصادفی یادگیری مطابق جدول زیر می باشد.

نوع ترافیک	ظرفیت کل دادگان	ظرفیت بعد از انتخاب
نرمال	۹۷۲۷۸ (٪ ۱۹,۶۹)	۵۶۰
Dos	۳۹۱۴۵۹ (٪ ۷۹,۲۴)	۱۲۰۰
Probe	۴۱۰۸ (٪ ۰,۸۳۲)	۱۶۰
U2R	۵۳ (٪ ۰,۰۱)	۱۰
R2L	۱۱۲۷ (٪ ۰,۲۲۸)	۹۰
جمع کل	۴۹۴۰۲۵	۲۰۲۰

طبق جدول ، پنج درصد ترافیک رندوم انتخاب شده مربوط به حملات از نوع R2L و U2R با سهم ۱۰ به یک، ۸ درصد را حمله شناسایی و مابقی ۸۷ درصد را به ترتیب ۲۸ درصد نرمال و ۶۰ درصد Dos تشکیل می دهند. به نظر

می رسد که این سهمیه بندی در تشکیل دیتاست آزمایش ما ، استاندارد باشد. این دادگان جدید، حاوی حملات زیر است. (جدول ۱) برای آزمون نیز دادگان دیگری دقیقاً با همین اندازه را از کل دادگان موجود انتخاب نمودیم. منتها این بار تعدادی نوع حمله جدید را از دادگان کل استخراج نموده و در داده آزمون گنجاندیم ، حملاتی که در داده یادگیری استخراج شده وجود ندارند. جداول زیر نوع و تعداد حملاتی که به تفکیک در تشکیل هر دو دادگان یادگیری و آزمون به کار رفته اند را نشان می دهد. همانطور که ملاحظه می شود ، در دادگان آزمون تعدادی حمله از نوع R2L و U2R گنجانده شده که در نمونه مشابه یادگیری آن نیست. این بدان جهت است که عملکرد دسته بند یا همان تابع هدف در ارزیابی بهینه ترین زیر مجموعه انتخاب شده در تشخیص حملات ناشناخته بدست آید.

جدول ۱ - زیر مجموعه استخراج شده برای آزمون و یادگیری

Category	Training Data		Test Data	
	Class Labels	Number	Class Labels	Number
Normal	Normal	۵۶۰	Normal	۵۶۰
	Ipsweep	۵۳	ipsweep	۴۳
Probe	Nmap	۷	nmap	۵
	portsweep	۴۰	portsweep	۳۶
	Satan	۶۰	satan	۵۱
			mscan	۱۰
			saint	۱۵
	Back	۷	back	۷
Dos	Land	۱	land	۱
	Neptune	۳۰۱	neptune	۲۸۹
	Pod	۲	pod	۲
	Smurf	۸۸۶	smurf	۸۴۷
	Teardrop	۳	teardrop	۲
U2R			apache2	۴
			mail bomb	۴۲
			processtable	۴
			udpstorm	۲
	buffer overflow	۶	buffer overflow	۲
	load module	۱	load module	۱
	Rootkit	۳	Rootkit	۲
			ps	۱
			httptunnel	۲
			sqlattack	۱
			xterm	۱
	ftp_write	۱	ftp_write	۱
R2L	guess password	۲	guess password	۲
	Imap	۳	imap	۳
	multihop	۲	multihop	۱
	warez client	۸۱	warez client	۶۶
	warez master	۱	warez master	۱
			snmpguess	۳
			snmpgetattack	۷
			send mail	۱
			named	۱
			worm	۱
			xlock	۲
			Xsnoop	۱
New Attacks	۰		۹۸	
Total	۲۰۲۰			

ما ۹۸ نوع حمله جدید در دادگان آزمون گنجاندیم تا در فازهای بعدی پژوهش با متدها تشخیص نفوذ و دسته بند های مختلف آزمایش کنیم این میزان حمله جدید تقریباً در حدود ۴,۸۵ درصد از کل حجم دادگان آزمون ما را

تشکیل می دهند. در نتیجه دو مجموعه آزمون و یادگیری را با هم ترکیب کردیم تا در مجموع ۴۰۴۰ رکورد ترافیک تصادفی بدست آمد. این دادگان نفوذ دارای انواع کلاسهای مختلف و تمامی حملات از هر نوع می باشد که برای آموزش یک شبکه عصبی که به عنوان دسته بند یادگیری در یک متد راپر کافی است. همچنین نرخ های یادگیری ، تصدیق و آزمون را به ترتیب ۶۰ ، ۱۵ و ۲۵ درصد وارد کردیم.

۴-۲-۱- بَهره اطلاعات

بهره اطلاعات به عنوان معیاری کلی برای ارزیابی میزان ارتباط هر ویژگی با کلاس در بردار مربوطه می باشد. این معیار مبتنی بر تئوری اطلاعات است و بستگی به حجم دادگان s دارد. [۲۴]

اگر دادگان نفوذ حاوی m کلاس را در نظر بگیریم بطوریکه S_i نمونه از کلاس I موجود باشد ، در اینصورت میزان بهره اطلاعاتی مورد انتظار در دسته بندی نمونه ها به صورت رابطه زیر است :

$$I(s_1.s_2 \dots .s_m) = -\sum_{i=1}^{m} \frac{s_i}{s} \log(\frac{s_i}{s}) \qquad (۴-۱)$$

یک ویژگی F با مقادیر ممکن $\{f_1.f_2 \dots f_v\}$ می تواند دادگان نفوذ را به v زیر مجموعه به صورت $\{S_1.S_2 \dots S_v\}$ تقسیم می شود که در آن S_j ، زیر مجموعه ای است که مقدار f_j را برای ویژگی F دارد. بنابراین S_j شامل s_{ij} نمونه از کلاس i می باشد. پس آنتروپی ویژگی F به صورت زیر است :

$$E(F) = \sum_{j=1}^{v} \frac{s_{1j} \dots s_{mj}}{s} \times I(S_{1j} \dots S_{mj}) \qquad (۴-۲)$$

بهره اطلاعات ویژگی F :

$$Gain(F) = I(s_1.s_2 \dots .s_m) - E(F) \qquad (۴-۳)$$

نمودار زیر بهره اطلاعات بدست آمده هر ویژگی را در دادگان نفوذ مورد آزمایش ما را نشان می دهد. همچنین در بخش آزمایش از بهره اطلاعات به عنوان یک متد تشخیص نفوذ مبتنی بر تئوری اطلاعات مبتنی بر شبکه عصبی به عنوان تابع ارزیابی استفاده کردیم.

مشخصات پارامترهای ورودی این آزمایش برای دو حالت انتخاب/حذف تصادفی و انتخاب/ حذف مبتنی بر بهره اطلاعات هر ویژگی در فاز جستجوی زیر مجموعه ها ، به صورت جدول زیر می باشد. تابع ارزیابی نخست که همان بَرازندگیِ حاصل از دسته بندِ ما را نشان می دهد منطبِق بر نرخ تشخیص و خطای مثبتِ اشتباهِ تابع ارزیاب (شبکه عصبی) در هر فازِ تکرار الگوریتم بوده و به صورت زیر فرموله می شود. ضمناً خطای بدست آمده از مجموع یادگیری و آزمون شبکه عصبی را نیز به عنوان تابع هزینه در نظر گرفتیم.

در مقاله اصلی [۱۰] ، نویسنده برای تابع هدف از درخت تصمیم [1] استفاده کرده و کاربرد شبکه عصبی را به عنوان یک مسئله باز در این حوزه ، برای کارهای آتی پیشنهاد کرده است. ضمن اینکه با بررسی دقیق الگوریتم اصلی مشاهده شد که این الگوریتم به صورت تصادفی در برخی مشاهدات با مشکلِ بهینه محلی[2] مواجه می شود. ما برای رفع این چالشها سعی کردیم که مفهوم بهره اطلاعات را در فازهایی استفاده کنیم که حذف ویژگی را به صورت تصادفی انجام می دهند. از اینرو بَهره اطلاعات را در فاز آغازین و پنجم به کار بردیم بطوریکه حلقه مرحله پنجم را به طور کلی حذف نموده و به جای آن در این مرحله یک ویژگی با کمترین آنتروپی را حذف نمودیم تا مکانیسمِ این الگوریتم را هدفمند سازیم.

از طرفی بحثِ اینکه کدام نوع شبکه عصبی به عنوانِ دسته بندِ یادگیریِ CFA به کار رود مسئله ای است که در بخش بعد با ارزیابیِ مقایسه ای متد MLP و شبکه ی عمیقِ دولایه به نام Stacked Auto Encoder و با دسته بندی ۵ کلاسه به طور کامل بررسی گردید و نهایتاً شبکه عمیق در آزمایشات بعدی به کار رفت.

در دو آزمایش متفاوت (انتخاب/حذف ویژگی به صورت تصادفی یا مبتنی بر آنتروپی هر ویژگی) از شبکه عصبی به عنوان تابع هدف استفاده نمودیم. نتایج بدست آمده حاکی از آنند که شبکه عصبی می تواند دسته بندیِ به مراتب بهتری را نسبت به درخت تصمیم انجام دهد. همچنین نشان داده شده که استفاده از بهره اطلاعاتِ ویژگی ها به منظور انتخاب/حذف در فازهای مختلف CFA نیز که می تواند به طرز موثری نرخ تشخیصِ دسته بندِ یادگیری را بالا برده و خطاهای کاذب را کاهش دهد و از حذف / انتخاب کورکورانه و صرفاً تصادفیِ ویژگی (های) خاص در هر فاز تکرارِ الگوریتم جلوگیری نماید.

به منظور ارزیابیِ بهترین زیر مجموعه ی ویژگی انتخاب شده رابطه ارزیابیِ زیر را به کار بردیم. به این ترتیب که پس از انتخاب زیر مجموعه ویژگی در هر فاز ، شبکه عصبی بلافاصله به تشکیلِ ماتریسِ درهم ریختگی پرداخته و معیارهای خطا و عملکردِ دسته بندی را محاسبه می نماید. تابعِ فیتنسی که در مقاله ارائه شده به صورت رابطه زیر می باشد.

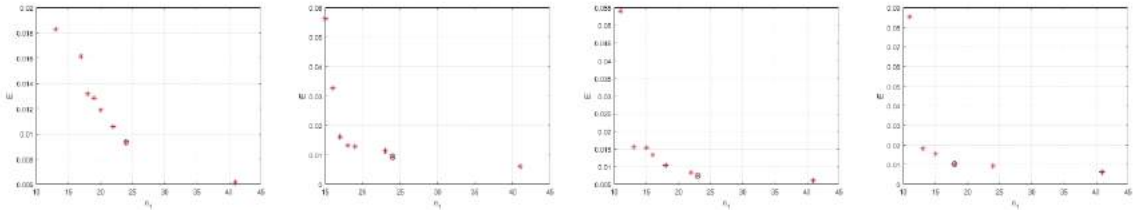
$$Fit = \alpha \times DR + \beta \times (1 - FPR) \qquad (۴-۴)$$

[1] Decision Tree
[2] Local Optimum Problem

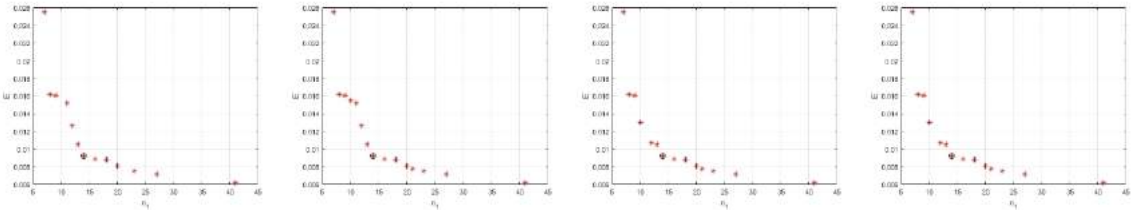
در الگوریتم اصلی ، فاز پنجم حلقه ای را به تعداد ویژگیهای انتخاب شده در AVbestSubset تکرار نموده و هر بار یک ویژگی را حذف می کند تا عملکرد ماهیچه Leocophor در لایه زیرین پوست ده پا را تَداعی کند. ما با بررسی دقیقِ عملکردِ الگوریتم (Trace) در این فاز متوجه شدیم که به جز اتلافِ زمان و افزایشِ پیچیدگیِ فضا و زمانِ الگوریتم هیچ عملکردِ مثبتی ندارد. زیرا از دیدگاه تئوری اطلاعات و بنا به استناد به مقاله [۲۴] صرفاً حذفِ ویژگیِ با کمترین آنتروپی می تواند خروجیِ تابعِ ارزیابی را بهبود بخشد. بنابراین با حذف یک ویژگی که کمترین بهره اطلاعاتی را دارد فاز پنجم اصلاح شد.

جدول ۲ - پیکربندی پارامترهای ورودی

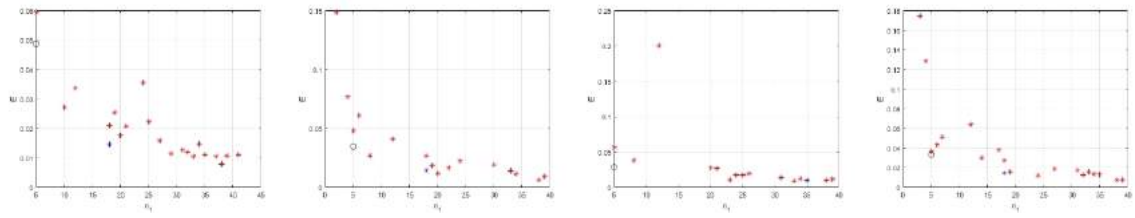
Pop size	Cross over rate (%)	Mutation rate (%)	Parameters
۲۰	۰,۶	۰,۴	NSGA II
Pop size	t		Parameters
۲۰	۱۰		CFA – ANN



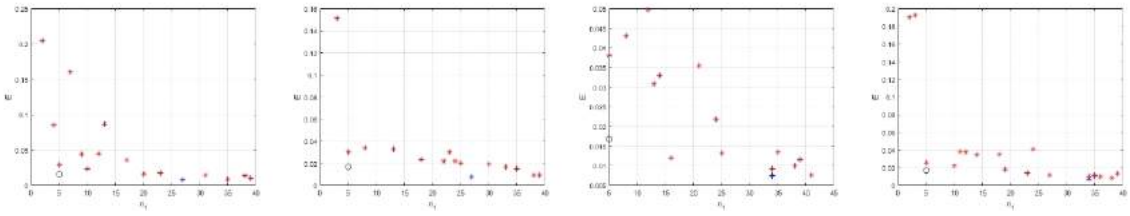
خروجی الگوریتم NSGA II در تکرارهای مختلف (از چپ به راست)



۴ تکرار آخر (از راست به چپ)

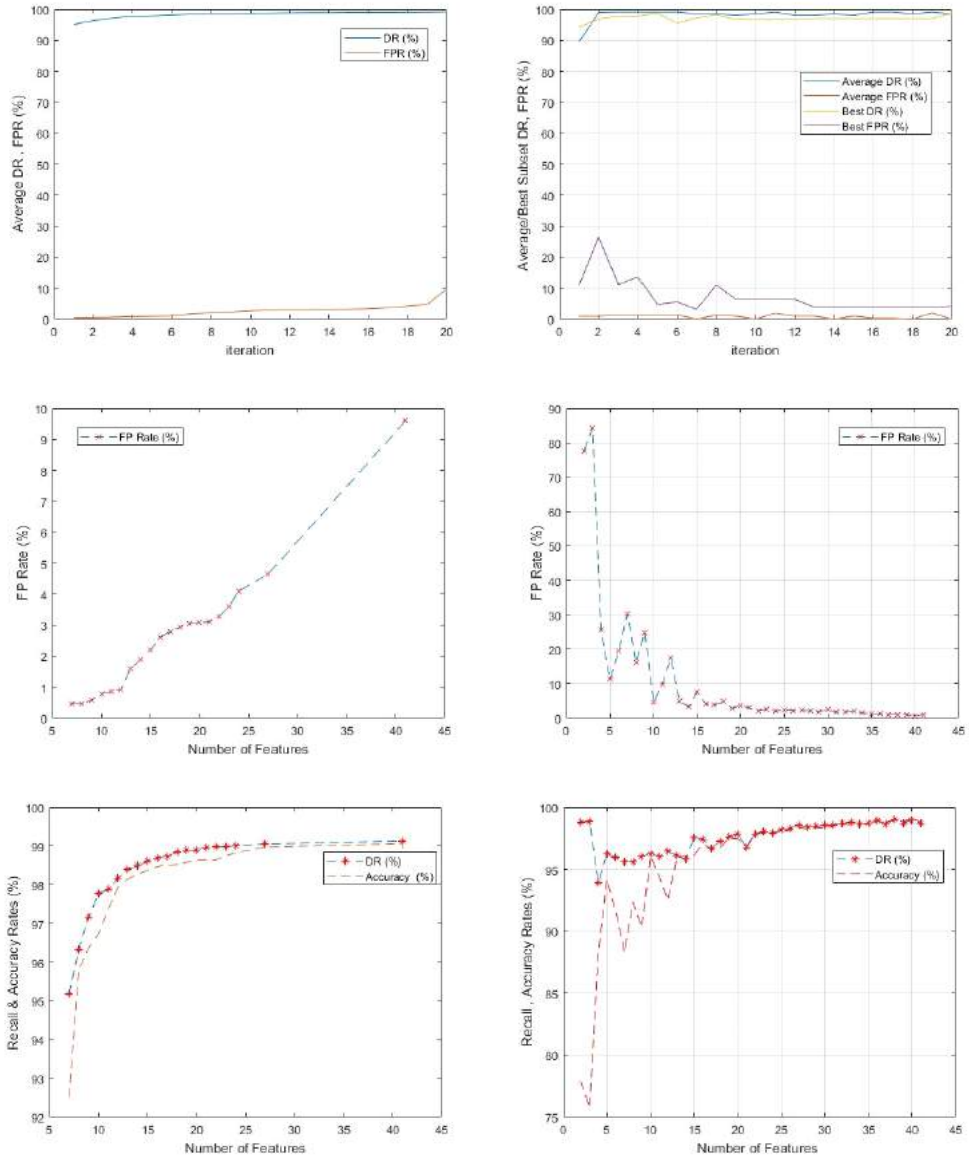


خروجی الگوریتم CFA – ANN در تکرارهای مختلف



۴ تکرار آخر

۳-۴- آنالیز مقایسه ای CFA – ANN و NSGA II



نمودار ۳ – نمودارها ، سمت چپ NSGA II , سمت راست CFA-ANN

جدول ۳- مقایسه بهینه ترین زیر مجموعه بردار ویژگی بدست آمده پس از ۲۰ تکرار

No. of features	*FPR*(%)	*DR*(%)	*ACC*(%)	*Fitness*(%)	*Type*	*Method*
۵	۲,۵۷۸۸	۹۹,۰۷۳	۹۸,۵۹۷	۹۸,۵۷۷۵	*Wrapper*	$Cfa - ANN$
۲ ۳ ۱۲ ۳۲ ۴۰						$Op - subset$
۱۴	۰,۹۴۳۴	۹۸,۹۹۳۳	۹۹,۰۱	۹۹,۰۱۲۳		*NSGA II*
۲ ۳ ۴ ۷ ۱۰ ۱۲ ۱۳ ۱۵ ۲۸ ۲۹ ۳۲ ۳۳ ۳۴ ۴۰						$Op - subset$
۱۶	۱,۸۳۲۵	۹۸,۷	۹۹,۱	۹۹	*Filter*	*Information Gain* (*IG*)
۲ ۳ ۴ ۵ ۱۸ ۱۹ ۲۱ ۲۲ ۲۹ ۳۰ ۳۱ ۳۶ ۳۷ ۳۸ ۳۹ ۴۱						$Op - subset$

مطابق جدول فوق در هر سه مورد ، اندازه لایه مخفی را ۲۵ گرفتیم. متد IG بر خلاف رویکرد های راپر، تکرار نداشته و فقط یکبار اجرا شد. سپس مقادیر آنتروپی بدست آمده از تمام ویژگیها را به ترتیبِ نزولی مرتب سازی نموده و به اندازه ی بهترین تعداد ویژگیهای بدست آمده از دو رویکرد راپر ، از این ترتیبِ نزولی ، ویژگی انتخاب نمودیم و با شبکه عصبیِ آزمون کردیم. همانطور که قبلاً نیز اشاره شد یک مزیّت مهم رویکرد های راپر نسبت به فیلترِ علیرغم زمان بالای پردازشی و حافظه مورد نیاز آنها ، حذف ویژگیهای غیر مرتبط یا اضافی با انجام تکرارها و آزمون های مکرّر مبتنی بر آزمون و خطا در این الگوریتم هاست. از آنجایی که هدف از طراحی متدهای انتخابِ ویژگی بخصوص با رویکرد راپر ، کاربرد آنها در سیستم های هوشمندی همچون سیستم تشخیص آنومالی میباشد که پیوسته با دادگانِ نفوذِ حجیم سرو کار دارند.

با مشاهده ی نمودارها با استفاده از رابطه فیتنس فوق می توان بهترین نقاط را در هر تکرار الگوریتم بدست آورد و با bestSubset و AVbestSubset در همان تکرار مقایسه کرد تا بهترین زیر مجموعه ویژگی نهایی بدست آید. مزیّت رابطه فوق در این است که تمامی معیارهای موثر در انتخاب زیر مجموعه ویژگی بهینه مانند دقّت تشخیص ، خطا و حتّی تعداد ویژگیهای انتخاب شده را نیز لحاظ می کند.

۴-۴- یک بررسی

از جمله نقاط ضعفِ الگوریتمِ اصلی دَه - پا می توان به عدم قابلیت اطمینان به عملکرد آن در طی فازهای مختلف اجرا و انتخاب زیر مجموعه بهینه از ویژگی ها را نام برد. به عبارت دیگر طبق بررسی های انجام شده از نتایج و تحلیل داده های بدست آمده از آزمایشات ، bestSubset در بهینه محلی گیر می کند و تکرار زیادی لازم است تا از آن خارج شود. البته با در نظر گرفتن AVbestSubset و میانگین های بردارهای ویژگی بدست آمده از تمامی فازها ، می توان آنالیز بهتری را انجام داد که نشان می دهد این الگوریتم نیاز به بهبود دارد.

همچنین تصادفی بودن حذف ویژگیها یکی دیگر از نقاط ضعف این الگوریتم است زیرا حداکثر تعداد انتخاب های ممکن برای ایجاد زیر مجموعه بهینه بسیار بیشتر از آنست که در حد و اندازه توان پردازشی و زمان اجرای آزمون و محاسبه برای تمام انتخابها در عمل میسر باشد. بدین معنی که برای انتخاب بهینه ترین بردار ویژگی نهایی در دادگان حجیم می بایست تکرار های زیادی انجام شود که مقرون به صرفه نیست.

۵-۴- جمع بندی

در شبیه سازیِ CFA از بهره اطلاعات به منظور حذف یا انتخاب تصادفی ویژگیهای موجود در فاز جستجو استفاده شده و یک شبکه عصبیِ نیز به عنوان دسته بند جهت آموزش و ارزیابی هزینه زیر مجموعه انتخاب شده در این الگوریتم به کار رفت. همانطور که از نتایج آزمایشات قابل استنباط است استراتژی جستجوی مبتنی بر دَه - پا در فازهای مختلف آزمون و خطاهای متعددی انجام می دهد تا تجربه لازم جهت انتخاب بهینه ترین بردار ویژگی را کسب نماید و به تَبَعِ آن حجم حافظه و زمان پردازش زیادی برای دادگان نفوذ لازم است.

بخش پنجم – ارزیابی مقایسه ای دو شبکه عصبی به منظور کاربرد جهت دسته بندِ یادگیریِ CFA

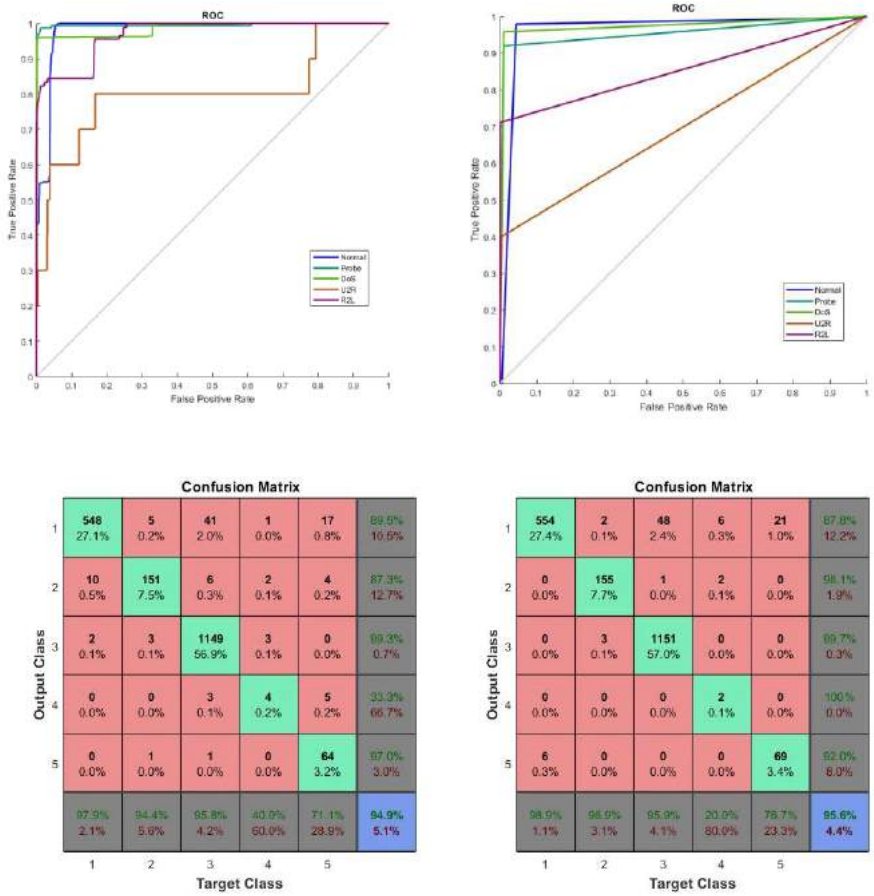
طبق جدول ۴ مشاهده می گردد که بهینه ترین پیکربندی متد MLP در آزمایش چهارم توانسته است بهترین نتایج را حاصل کند ، اما در مقام مقایسه با متد یادگیری عمیق نمی تواند رقابت نماید. بطوریکه در آزمایش ۱۵ بهینه ترین پیکربندی برای یادگیری عمیق ، نتایج آن را بهبود داده است. ما در آزمایشی دیگر اندازه لایه مخفی MLP را از ۲۰ به ۱۰۰ افزایش دادیم تا رقابت MLP را با متد یادگیری عمیق (Stacked Auto Encoder) مشاهده کنیم. نتایج زیر از این آزمایش بدست آمد. همانطور مشاهده می شود، با افزایش لایه مخفی نیز متد MLP باز هم توانایی رقابت با یادگیری عمیق با دو لایه مخفی را ندارد. بنابراین متد یادگیری عمیق فوق بهترین گزینه برای استفاده به عنوان "دسته بند یادگیری" در الگوریتم پیشنهادی انتخاب ویژگی می باشد.

method	run	hidden size		L2 Weight Regularization	Sparsity Regularization	Sparsity Proportion	Decoder Transfer Function	scale data	rec	acc	fpr	fnr	tpr	tnr	f-measure	spec	sen	err	prec	ec
MLP	1	20		train ratio =80 ,validation ratio = 20					94.9315	0.9619	0.5357	5.0685	94.9315	99.4643	1.9749	0.9946	0.9978	0.0381	0.9978	0.9117
	2								94.8630	0.9609	0.7143	5.1370	94.8630	99.2857	1.9735	0.9929	0.9486	0.0391	0.9971	0.9093
	3								93.5615	0.9490	1.6071	6.4384	93.5615	98.3929	1.9660	0.9839	0.9356	0.0510	0.9935	0.8829
	4								95.3425	0.9653	0.3571	4.6575	95.3425	99.6429	1.9764	0.9964	0.9534	0.0347	0.9986	0.9194
	5								94.8630	0.9624	0.1786	5.1370	94.8630	99.8214	1.9777	0.9982	0.9486	0.0376	0.9993	0.9133
AVG-MLP1									94.7123	0.9599	0.389282	5.28768	94.7123	99.32144	1.9737	0.9932	0.9568	0.0401	0.99726	0.90732
MLP	6	25		train ratio = 85,validation ratio =15					94.5205	0.9584	0.7143	5.4795	94.5205	99.2857	1.9734	0.9929	0.9452	0.0416	0.9971	0.9041
	7								94.5890	0.9525	3.0357	5.4110	94.5890	96.9643	1.9553	0.9696	0.9459	0.0475	0.9878	0.8878
	8								94.7945	0.9614	0.3571	5.2055	94.7945	99.6429	1.9763	0.9964	0.9479	0.0386	0.9986	0.9109
	9								94.9315	0.9634	0	5.0685	94.9315	100	1.9792	1	0.9493	0.0366	1	0.9157
	10								95.2055	0.9653	0	4.7945	95.2055	100	1.9792	1	0.9521	0.0347	1	0.9199
AVG-MLP2									94.7542	0.960206	0.82142	5.1918	94.8082	99.17858	1.97268	0.99178	0.94808	0.0398	0.9967	0.90768
MLP	11	41		train ratio =85 ,validation ratio = 15					95.2740	0.9644	0.5357	4.7260	95.2740	99.4643	1.9750	0.9946	0.9527	0.0356	0.9978	0.9170
	12								94.8630	0.9619	0.3571	5.1370	94.8630	99.6429	1.9763	0.9964	0.9486	0.0381	0.9986	0.9120
	13								95.2740	0.9653	0.1786	4.7260	95.2740	99.8214	1.9778	0.9982	0.9527	0.0347	0.9993	0.9197
	14								94.5205	0.9564	1.4286	5.4795	94.5205	98.5714	1.9678	0.9857	0.9452	0.0436	0.9942	0.8987
	15								95.0685	0.9634	0.3571	4.9315	95.0685	99.6429	1.9764	0.9964	0.9507	0.0366	0.9986	0.9152
AVG-MLP3									95	0.96228	0.57142	5	95	99.42858	1.97466	0.99426	0.94998	0.03772	0.9977	0.91252
AVG-MLP									94.8221	0.96079	0.5940	5.1598	94.84	99.30953	1.97368	0.99308	0.95162	0.0392	0.99722	0.90917
St-Auto	1	41	25	0.001/0.001	4/4	0.05/0.1	purelin/ purelin	false	95.1370	0.9649	0	4.8630	95.1370	100	1.9792	1	0.9514	0.0351	1	0.9189
	2	41	25	0.001/0.005	8/4	0.1/0.1	logsig/ logsig	false	95.3425	0.9634	1.0714	4.6575	95.3425	98.9286	1.9575	0.9893	0.9534	0.0366	0.9889	0.9433
	3	41	20	0.005/0.001	4/8	0.05/0.05	purelin/ logsig	true	95.2740	0.9658	0	4.7260	95.2740	100	1.9792	1	0.9527	0.0342	1	0.9538
	4	25	41	0.005/0.005	8/8	0.1/0.1	logsig/ purelin	true	95.2055	0.9589	2.3214	4.7945	95.2055	97.6786	1.9326	0.9768	0.9521	0.0411	0.9762	0.9291
	5	20	41	0.001/0.001	4/4	0.1/0.05	logsig/ logsig	false	95.1370	0.9649	0	4.8630	95.1370	100	1.972	1	0.9514	0.0351	1	0.9525
AVG-St-Auto1									95.2192	0.96358	0.67856	4.7808	95.2192	99.32144	1.9641	0.99322	0.9522	0.03642	0.99302	0.93952
St-Auto	3	41	25	0.001/0.001	4/4	0.05/0.1	purelin/ purelin	true	100	0.7228	100	0	100	0	0.9950	0	1	0.2772	0.5	0
	7	41	25	0.001/0.005	8/4	0.1/0.1	logsig/ logsig	false	95.3425	0.9639	0.8929	4.6575	95.3425	99.1071	1.9611	0.9911	0.9534	0.0361	0.9907	0.9452
	8	41	20	0.005/0.001	8/8	0.05/0.05	purelin/ logsig	true	95.4795	0.9673	0	4.5205	95.4795	100	1.9793	1	0.9548	0.0327	1	0.9558
	9	41	20	0.005/0.001	8/8	0.05/0.05	purelin/ logsig	false	95.4795	0.9673	0	4.5205	95.4795	100	1.9793	1	0.9548	0.0327	1	0.9558
	10	41	41	0.001/0.001	8/8	0.05/0.05	purelin/ logsig	false	95.2055	0.9653	0	4.7945	95.2055	100	1.9792	1	0.9521	0.0347	1	0.9532
AVG-St-Auto2									96.3014	0.91732	20.17858	3.6986	96.3014	79.82142	1.77878	0.79822	0.96302	0.08268	0.89814	0.762
St-Auto	11	41	25	0.005/0.005	8/8	0.1/0.1	logsig/ logsig	false	94.9315	0.9599	1.2500	5.0685	94.9315	98.7500	1.9537	0.9875	0.9493	0.0401	0.9870	0.9375
	12	41	25	0.001/0.001	4/4	0.1/0.1	logsig/ logsig	false	95.4110	0.9634	1.2500	4.5890	95.4110	98.7500	1.9539	0.9875	0.9541	0.0366	0.9871	0.9421
	13	41	20	0.001/0.001	4/4	0.1/0.1	logsig/ logsig	true	95.3425	0.9653	0.3571	4.6575	95.3425	99.6429	1.9719	0.9964	0.9534	0.0347	0.9963	0.9507
	14	41	41	0.005/0.005	8/8	0.05/0.05	purelin/ logsig	false	95.3425	0.9663	0	4.6575	95.3425	100	1.9792	1	0.9534	0.0337	1	0.9545
	15	41	10	0.001/0.001	8/8	0.05/0.05	purelin / logsig	false	95.8219	0.9693	0.1786	4.1781	95.8219	99.8214	1.9757	0.9982	0.9582	0.0307	0.9981	0.9572
AVG-St-Auto3									95.36988	0.96484	0.60714	4.63012	95.36988	99.39286	1.96688	0.99392	0.95368	0.03516	0.9937	0.9484
AVG-St-Auto									95.63016	0.94858	7.15476	4.36984	95.63016	92.84524	1.903253	0.9285	0.9563	0.05142	0.96162	0.8833

جدول ۵ – مقایسه بین بهترین نتایج بدست آمده از آزمایشات جدول ۴

method	MLP	St-Auto
Run	۴ / new	۱۵
hidden size	۱۰۰	۴۱ / ۱۰
L2 Weight Regularization	train ratio =80 =,validation ratio ۲۰	۰,۰۰۱/۰,۰۰۱
Sparsity Regularization		۸/۸
Sparsity Proportion		۰,۰۵/۰,۰۵
Decoder Transfer Function		/purelin
Scale data		false
rec	۹۵,۳۴۲ / ۹۴,۸۶۳۰	۹۵,۸۲۱۹
Acc	۰,۹۶۵۳ / ۰,۹۶۹۰	۰,۹۶۹۳
fpr	۰,۳۵۷۱ / ۰,۷۱۴۳	۰,۱۷۸۶
fnr	۴,۶۵۷۵ / ۵,۱۳۷۰	۴,۱۷۸۱
tpr	۹۵,۳۴۲ / ۹۴,۸۶۳۰	۹۵,۸۲۱۹
tnr	۹۹,۶۴۲ / ۹۹,۲۸۵۷	۹۹,۸۲۱۴
f-measure	۱,۹۷۶۴ / ۱,۹۷۳۵	۱,۹۷۵۷
Spec	۰,۹۹۶۴ / ۰,۹۹۲۹	۰,۹۹۸۲
sen	۰,۹۵۳۴ / ۰,۹۴۸۶	۰,۹۵۸۲
err	۰,۰۳۴۷ / ۰,۰۳۹۱	۰,۰۳۰۷
prec	۰,۹۹۸۶ / ۰,۹۹۷۱	۰,۹۹۸۱
cc	۰,۹۱۹۴ / ۰,۹۰۹۳	۰,۹۵۷۲

نکته ای که وجود دارد اینست که در بهترین حالتها برای دو متد فوق (آزمایش ۴ مربوط به شبکه عصبی و آزمایش ۱۵ مربوط به شبکه عمیق) نرخ منفی اشتباه به ترتیب از ۴,۶۵۷۵ و ۴,۱۷۸۱ پایینتر نمی رود که نشان دهنده عدم شناسایی حملات ناشناخته تعبیه شده در دادگان آزمون توسط دو متد فوق است. با استناد به بخش قبل که اشاره کردیم که دادگان آزمون ایجاد شده در مجموع ۹۸ مورد حمله جدید ناشناخته در خود دارد که از این تعداد ۲۵ مورد از نوع شناسایی ، ۵۲ مورد انکار سرویس ، ۵ مورد U2R و ۱۶ مورد نیز از نوع R2L هستند. بنابراین در آزمایشی مستقل با بردار ۴۰ ویژگی ، با بررسی دقیق و ارزیابی نتایج دسته بندی دو متد فوق با بهترین پیکربندی پارامترهای ورودی (۱۵ و ۴) مشاهده گردید که متد یادگیری عمیق مذکور در بدترین حالت قادر به کشف ۱۵,۳۰ درصد از حملات ناشناخته می باشد که نرخ بسیار پایینی است. (نمودار های زیر)



نمودار ۴ - ماتریس درهم ریختگی و منحنی
(سمت راست متد شبکه عصبی و سمت چپ متد یادگیری عمیق)

پیوست ها

پیوست الف – توضیحات بیشترِ مطالب

بخشِ نخست - اصل دوّمِ امنیّت (دفاع در عمق)

دوّمین اصل از اصول سیزده گانه امنیّت بحث دفاع در عمق را مطرح می کند. طبق این اصل ، امنیّت باید در چندین لایه برقرار شود به صورتی که هر لایه لایه دیگر را پوشش دهد. با رعایت این اصل اگر یک لایه دفاعی با شکست مواجه شود لایه های دیگر می توانند جلوی حملات را بگیرند.

ایده دفاع در عمق با انواع استراتژیهای دفاعی ، مدیریت ریسک را انجام می دهد. اگر یک لایه دفاعی نامناسب باشد و از دسترس خارج شود . امیدواریم که لایه دفاعی دیگر از نقص به صورت کامل دفاع کند. از اینرو ، افزونگی و لایه بندی در امنیت فکر خوبی است.

بخش دوّم - ارزیابیِ مقایسه ایِ سه الگوریتمِ فراابتکاری

در این قسمت ، به منظور بررسی کیفیت دسته بندی نهایی ترافیک شبکه ، عملکردِ برخی از الگوریتم ها و متدهای منتخب در نرم افزار متلب در دادگان ترافیک شبکه UNSW-NB15 ارزیابی مقایسه ای نموده ایم. بدین منظور آزمایشِ زیر که به ترتیب بدون نظارت و با نظارت می باشند انجام شده اند. نتیجه این آزمایش تفاوت میان شیوه حلّ مسئله با دسته بندی باینری ترافیک شبکه در الگوریتمهای ارزیابی شده و همچنین الگوریتم سلولهای دندریت استاندارد (دسته بندیِ متمرکزگِرا و توزیع شده) را آشکارتر می نماید.

۱-۲- آزمایش

برای انجام این آزمایش سه الگوریتم فراابتکاری[1] PSO, GA, BHA انتخاب شدند. روش حل مسئله دسته بندی باینری در هر سه متد تقریباً یکسان بوده و به شیوه ی متمرکزگرا مسئله را حل می کنند. تفاوت آنها در استراتژی بهینه سازی موقعیت های مرکزیتهای یافت شده می باشد که تحت عنوان تابع هدف/فیتنس نیز شناخته می شود.

بنابراین در تمامی الگوریتم های فرا ابتکاری ، تابع ارزیابی فیتنس ، نقش موثری در تعیین بهترین موقعیّتِ مرکزیّتها دارد. بطوریکه در هر فازِ تکرارِ الگوریتم ، موقعیتهای هر نمونه/ ستاره/ ذرّه به عنوان مرکزیّتِ پیشنهادی آزمون شده و خوشه بندی انجام می شود در نهایت تابع فیتنس بر حسب معیار ارزیابی خود که معمولاً یک رابطه وزنیِ با نرخهای تشخیص و خطا میباشد تصمیم میگیرد و هزینه یا فیتنسِ موقعیتهای انتخاب شده را در خوشه بندی نهایی محاسبه و ارزیابی می کند. از این نظر انتخاب تابع فیتنس / هزینه و همچنین تعیین معیار مناسب در این تابع نقش بسیار موثری در عملکرد نهایی خواهد داشت. امّا در مورد مسائلی مانند دسته بندی باینری ترافیک شبکه ، که توزیع داده

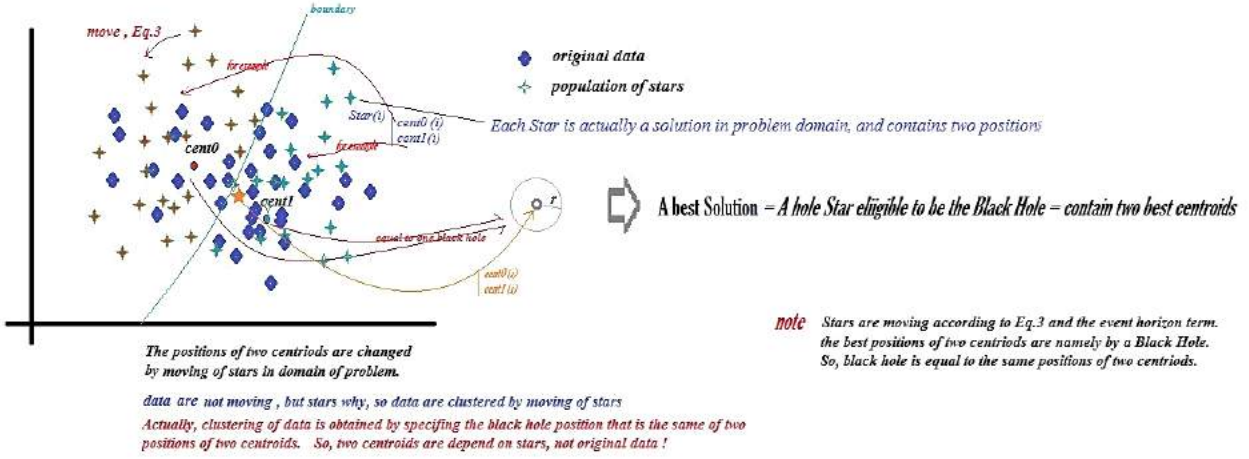
[1] The Black Hole Optimization Algorithm (in Clustering problem)

ها در فضای ابعاد مسئله به گونه ایست که ممکن است به جای دو خوشه ، چندین خوشه آنومالی یا نرمال داشته باشیم پس با ارزیابیِ نتایج آزمایشات اثبات می کنیم که کاربرد الگوریتم های فرا ابتکاری برای حلّ مسئله دسته بندی باینری در تشخیص نفوذ انتخاب مطلوبی نیست.

- ❖ الگوریتم ازدحام ذرّات
- ❖ الگوریتم ژنتیک
- ❖ الگوریتم حفره سیاه

بخش سوّم – الگوریتم سیاه چاله

الگوریتم حفره سیاه ، از پدیده بَلعیدن ستاره ها در فضا توسط سیاه چاله الهام گرفته شده است. این الگوریتم برای اولین بار در سال ۲۰۱۳ در مقاله [۱۳] ارائه گردید. توضیحات مربوط به این الگوریتم در بخش بخش بعدی ارائه شده است. برخی مانند نویسنده مقاله [۴۱] معتقدند که این الگوریتم شکل ساده شده و در واقع حالت سکون " الگوریتم ازدحام ذرّات" است و حتّی در مقالات [۵۶] الهام از بیولوژیک بودن آن مورد انتقاد قرار گرفته شده است. اینفوگرافیک زیر به وضوح سازوکار این الگوریتم جدید را نشان می دهد.

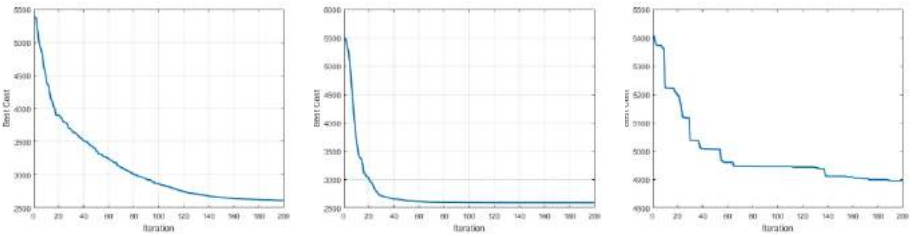


شکل ۱ - اینفوگرافیکِ مکانیسمِ خوشه بندیِ حفره سیاه ، برگرفته از شبیه سازی منتشر شده [۲]

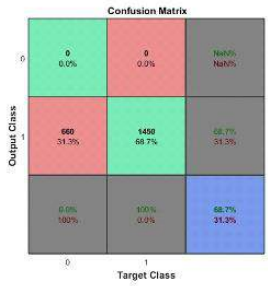
نتایج

نتایج آزمایش نشان می دهد که در تمام آزمون ها ، میزان خطای مثبت اشتباه بسیار بالاست. حتی ما تلاش کردیم و مجدداً پارامترها را تغییر دادیم اما ، میزان خطای مثبت کاهش چشمگیری نیافت. حتی با تغییر تابع فیتنس نیز وضع به همین گونه است. تابع فیتنس برای هر سه الگوریتم ، فاصله درون خوشه ای[1] تعیین شد.

[1] Fitness function: Sum of Within - Cluster Distance (WCD)



نمودار - 1 نتایج الگوریتم به ترتیب از راست با پارامترهای پیشفرض[1]



نمودار ۲ - ماتریس در هم ریختگی برای تمام الگوریتمها با پارامترهای پیشفرض

بخش چهارّم - اهمیّتِ انتخابِ ویژگی (دفاع غیر مستقیم)

در این بخش ضمنِ بیانِ مقدّمه ای در رابطه با انتخابِ ویژگی و اهمیّت آن ، به بررسی استراتژی جستجوی ده پا در بازتاب و انعکاس نور تابیده شده به ماهیچه های پوست این جاندار پرداخته و کاربردِ مکانیسمِ انتخاب ویژگیِ دَه -پا را در انتخاب موثّرترین ویژگیهای دادگان ترافیکِ شبکه ارزیابی نموده ایم.

مهمترین نیازمندیهایی که برای یک سیستم تشخیص نفوذ در شبکه مطرح می باشند عبارتند از افزایش دقت و سرعت فرایند تشخیص در حین یادگیری ، کار با داده حجیم ، دانستن مجموعه ویژگیهای بهینه مرتبط با هر نوع حمله. به منظور افزایش دقت و سرعت طبق تحقیقات انجام شده موثرترین راهکار همان کاهش حجم داده از طریق انتخاب بهترین و بهینه ترین ویژگیها و حتی نمونه هاست. انتخاب و استخراج ویژگی، یک تکنیک، یادگیری ماشینِ ضروری است که در ایجاد سیستم های دسته بندی کارآمد موثر است. وقتی که برای کاهش ویژگی استفاده شود نتیجه آن کاهش هزینه های محاسباتی و دسته بندی بهتر است [۱۰۵]. مسئله ی انتخاب ویژگی تلاش می کند تا زیر مجموعه حداقلی به اندازه $m < n$ را از بین n ویژگی پیدا کند تا بدین طریق عملکردِ دسته بندها را افزایش دهد. در واقع در این وضعیّت ، ابعادِ مسئله کاهش یافته است.

هر نوع حمله ای در شبکه ، از زیر مجموعه ای خاص از ویژگیهای خاص نَشئَت میگیرد. به همین دلیل کاهشِ ابعاد فرایندِ بسیار مهمی می باشد. از طرفی برای انتخابِ بهینه ترین ویژگیهای مرتبط با هر نوع حمله می بایست استراتژی جستجوی مناسبی اتخاذ گردد که در کمترین زمانِ ممکن بتواند ویژگیهای مناسب را انتخاب و حجم مجموعه ی داده را کاهش دهد. به طور کلی رویکردهای مختلفی برای فرایندِ انتخابِ ویژگی در تشخیص نفوذ به کار رفته اند. در این میان، الگوریتمهای مختلفی با رویکردهای فیلتر و راپر و حتی ترکیبی وجود دارند. مسئله اساسی که در زمینه این الگوریتمها وجود دارد انتخاب رویکرد صحیح ، بسته به کاربرد در مسئله دسته بندی در تشخیص نفوذ است.

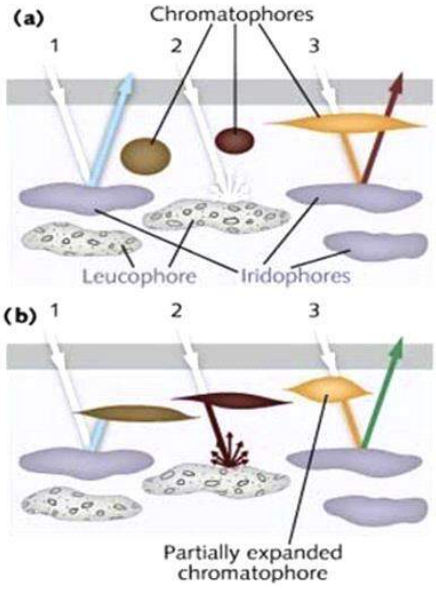
[1] for ga (Maxiteration = 200. nPop = 100. pc = 0.8. pm = 0.3.gamma = 0.2. mu = 0.2.beta = 8). for pso ($\phi 1 = \phi 2 = 2.05$)

بطوریکه هر نوع تکنیک دسته بندی و کلاً تکنیک های به کار رفته در تشخیص آنومالی با رویکرد و استراتژی خاصی از انتخاب ویژگی متناسب بوده و منجر به خروجی بهینه می شوند. این مسئله موجب شده محققان دراین زمینه هر بار و در هر آزمایش رویکردهای مختلف تشخیص را با رویکردهای مختلف انتخاب ویژگی ترکیب نموده و پیشنهاداتی را ارائه دهند. اما نکته ای که وجود دارد این است که بیشتر این رویکردها حتّی با ترکیب در فرایند انتخاب ویژگی نیز بهینه موفق عمل نکرده اند. این ضعف عمده تکنیکهای کاهش ابعاد ساخته انسان است.

طبیعت و رفتار بیولوژیکی موجودات زنده در آن ، علاوه بر اینکه منشاء کافی برای ایده پردازی از الگوهای سیستم دفاعی آنها جهت کاربرد در زمینه های مختلف تشخیص نفوذ هستند بلکه بسیاری از ویژگیهای جانوران وجود دارند که در حل مسائل به کار رفته اند. بنابراین الهام از سیستم طبیعیِ موجوداتِ زنده در طبیعت و تقلید از غریزه و حسّ مشارکتیِ آنان در حلّ مسائل منبعی سرشار از ایده پردازی را برای محققان فراهم ساخته است. از جمله مسئله استراتژی جستجو و انتخاب بهینه ترین ویژگیها در زمینه تشخیص آنومالیِ شبکه می باشد.

۱-۴- استراتژی جستجوی دَه - پا

دَه - پا[1] ، نام جانوری شبیه به هشت پا در دریاست که نور تابیده شده به سطح پوست و ماهیچه های بدن خود را با مکانیزمهای بسیار پیچیده و مخفی در لایه های زیرین پوست خود پردازش نموده و انعکاس می دهد. رنگ ها و الگوهای ده پا بوسیله نور منعکس شده از سه لایه مختلف پوست رخ می دهند. این رفتار تغییر رنگ نور بازتابیده شده ایده ای است که برای حل بسیاری از مسائل بهینه سازی به کار می رود.



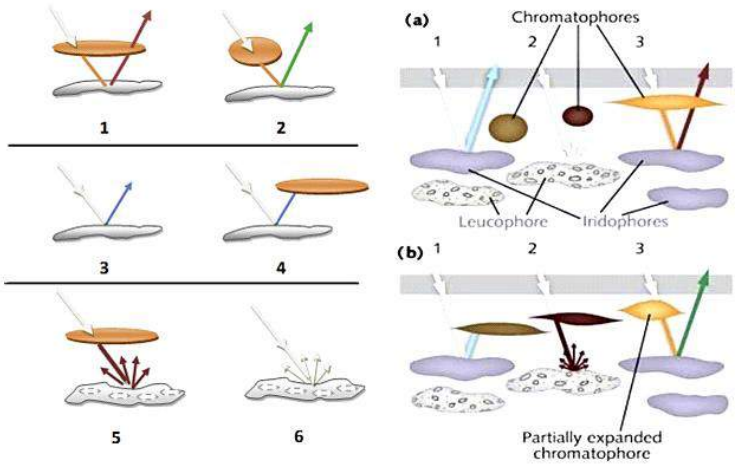
انعکاس از زوایای مختلف موجب می شود تا نور به محض تابش به سطوح مختلف زیر پوست به دلیل اینکه این از ویژگیهای مهم این متد اینست که در فازهای مختلف تکرار الگوریتم ، وابستگی زیادی به تابع ارزیابی و هزینه محاسبه شده فاز قبلی دارد. بدین معنی که از نظر عملکرد و انتخاب بهترین ویژگیها هر فاز به فاز قبلی خود وابسته است.

دَه - پا در زیر پوست خود سه ماهیچه دارد که عمل انقباض و انبساط را انجام می دهند. هر ماهیچه حاوی سلولهای زیادی است. نور وقتی که به این ماهیچه ها می تابد ، بسته به زاویه تابش R و اینکه به کدام نوع ماهیچه [2] در کدام لایه پوستی و کدام بافت سلولی تابیده ، در مجموع شش وضعیت کلی ممکن است برای نور رخ دهد که منجر به تغییر رنگ آن شود.

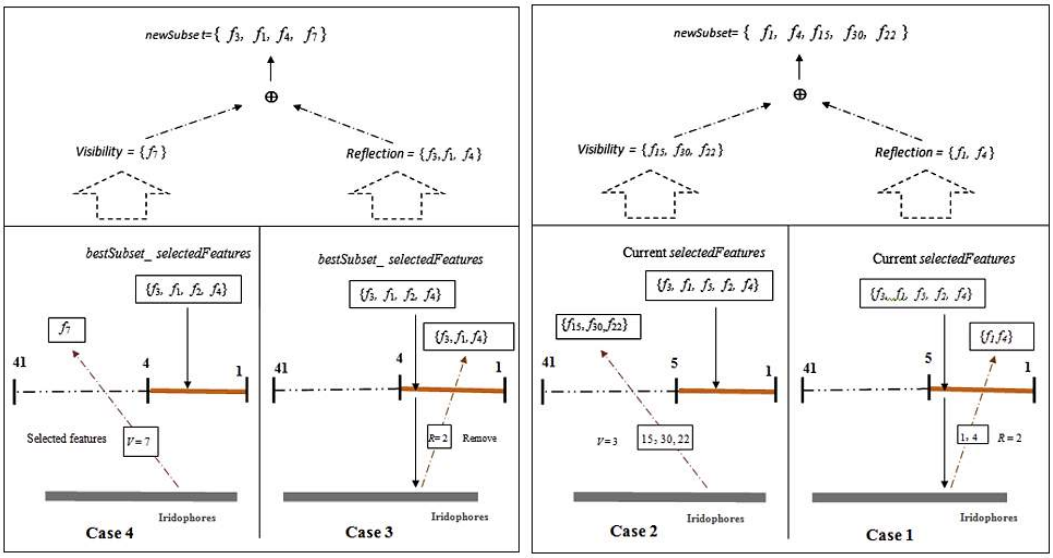
[1] Cuttlefish
[2] Chromatophores, Iridophores, leucophores

دو تابع انعکاس و بازتاب از این سازوکار تقلید می شود. این بافت های سلولی را می توان به عنوان بردارهای ویژگی کاندید در نظر گرفت. در حقیقت ، دو این دو تابع با زاویه تابش و انعکاس V نوری که نمایان می شود ، استراتژی جستجوی بهینه ترین بردار ویژگی را ترتیب می دهد. شکل زیر سه اثر پرتو نور متمایز را نشان می دهد که این جانور می تواند رنگ آمیزی بازتابی را با آنها تغییر دهد.

این وضعیت ها به گونه ای هستند که با هر بار تابش نور ورودی به ماهیچه ها در سطوح مختلف ، سلولی با مختصات مشخص بر روی بافت سلولی ماهیچه انتخاب می شود اما با شیوه های مختلف این کار صورت میگیرد. بطوریکه هر کدام از این سه ماهیچه با دو فرایند انقباض و انبساط، قادرند رنگ یا زاویه بازتابِ نور تابانده شده را تغییر داده و باصطلاح انتخاب ویژگی انجام دهند. سلولهای بازتابنده نور در درون اندامهای بدن جانور مواظبت می شوند.

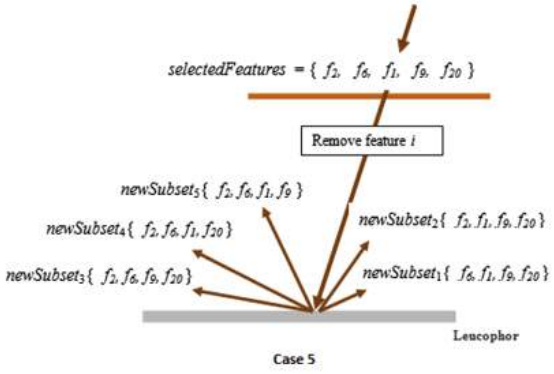


شکل ۳ – اثرِ متفاوتِ بازتاب و انعکاس نور تابیده شده به سه ماهیچه ی زیر پوستِ دَه – پا و انقباض و انبساط ماهیچه ها



شکل ۵ - فرایند تابش و انعکاس نور به ماهیچه های دَه – پا

شکل ۵ فرایند تابش و انعکاس نور در فاز اول را نشان می دهد. مطابق این شکل در هر تکرار این فاز ، از بین ویژگیهای بردار ویژگی انتخاب شده ، به صورت تصادفی ویژگی انتخاب می شود و از بین زیر مجموعه ویژگیهای انتخاب نشده بردار مذکور نیز تعدادی ویژگی طبق رابطه مربوطه می شوند تا زیر مجموعه جدیدی ایجاد گردد. زیر مجموعه ویژگی انتخاب شده از عمل انعکاس نور تابیده شده به سطح ماهیچه لایه اول پوست دَه پا الهام گرفته شده است. زیر مجموعه ویژگی انتخاب شده نیز عمل بازتاب را تداعی می کند. الگوریتم در فاز ۵ جستجوی محلی انجام می دهد.



Algorithm - Cuttlefish Algorithm Ref. [10]

Step 1: Randomly initialize the population of solutions; identify the most feasible solution and the average value of the best solution's points.
Step 2: Generate a new solution using the reflection and the visibility of pattern.
Step 3: Generate new solutions using the reflection of light and visibility of matching pattern.
Step 4: Generate random solution using reflection of incoming light

شکل ۶ – شبه کد الگوریتمِ جستجو و انتخاب ویژگی دَه - پا

۲-۴- آزمایش

ما برای انجام این آزمایش از دَه درصد دادگان نفوذ استاندارد استفاده کردیم. به دلیل حجم بسیار بالای این دادگان و عدم امکان پردازش کل این داده عظیم در نرم افزار مطلب ، در حدود ۰,۴ درصد از این دادگان را برای ایجاد زیر مجموعه یادگیری با دقت خاصی و به شکل تصادفی یکنواخت انتخاب نمودیم. جزئیات استخراج نمونه های تصادفی یادگیری مطابق جدول زیر می باشد.

نوع ترافیک	ظرفیت کل دادگان	ظرفیت بعد از انتخاب
نرمال	۹۷۲۷۸ (٪ ۱۹,۶۹)	۵۶۰
Dos	۳۹۱۴۵۹ (٪ ۷۹,۲۴)	۱۲۰۰
Probe	۴۱۰۸ (٪ ۰,۸۳۲)	۱۶۰
U2R	۵۳ (٪ ۰,۰۱)	۱۰
R2L	۱۱۲۷ (٪ ۰,۲۲۸)	۹۰
جمع کل	۴۹۴۰۲۵	۲۰۲۰

طبق جدول ، پنج درصد ترافیک رندوم انتخاب شده مربوط به حملات از نوع R2L و U2R با سهم ۱۰ به یک، ۸ درصد را حمله شناسایی و مابقی ۸۷ درصد را به ترتیب ۲۸ درصد نرمال و ۶۰ درصد Dos تشکیل می دهند. به نظر

می رسد که این سهمیه بندی در تشکیل دیتاست آزمایش ما ، استاندارد باشد. این دادگان جدید، حاوی حملات زیر است. (جدول ۱) برای آزمون نیز دادگان دیگری دقیقاً با همین اندازه را از کل دادگان موجود انتخاب نمودیم. منتها این بار تعدادی نوع حمله جدید را از دادگان کل استخراج نموده و در داده آزمون گنجاندیم ، حملاتی که در داده یادگیری استخراج شده وجود ندارند. جداول زیر نوع و تعداد حملاتی که به تفکیک در تشکیل هر دو دادگان یادگیری و آزمون به کار رفته اند را نشان می دهد. همانطور که ملاحظه می شود ، در دادگان آزمون تعدادی حمله از نوع R2L و U2R گنجانده شده که در نمونه مشابه یادگیری آن نیست. این بدان جهت است که عملکرد دسته بند یا همان تابع هدف در ارزیابی بهینه ترین زیر مجموعه انتخاب شده در تشخیص حملات ناشناخته بدست آید.

جدول ۱ - زیر مجموعه استخراج شده برای آزمون و یادگیری

Category	Training Data		Test Data	
	Class Labels	Number	Class Labels	Number
Normal	Normal	۵۶۰	Normal	۵۶۰
	Ipsweep	۵۳	ipsweep	۴۳
Probe	Nmap	۷	nmap	۵
	portsweep	۴۰	portsweep	۳۶
	Satan	۶۰	satan	۵۱
			mscan	۱۰
			saint	۱۵
	Back	۷	back	۷
Dos	Land	۱	land	۱
	Neptune	۳۰۱	neptune	۲۸۹
	Pod	۲	pod	۲
	Smurf	۸۸۶	smurf	۸۴۷
	Teardrop	۳	teardrop	۲
U2R			apache2	۴
			mail bomb	۴۲
			processtable	۴
			udpstorm	۲
	buffer overflow	۶	buffer overflow	۲
	load module	۱	load module	۱
	Rootkit	۳	Rootkit	۲
			ps	۱
			httptunnel	۲
			sqlattack	۱
			xterm	۱
	ftp_write	۱	ftp_write	۱
R2L	guess password	۲	guess password	۲
	Imap	۳	imap	۳
	multihop	۲	multihop	۱
	warez client	۸۱	warez client	۶۶
	warez master	۱	warez master	۱
			snmpguess	۳
			snmpgetattack	۷
			send mail	۱
			named	۱
			worm	۱
			xlock	۲
			Xsnoop	۱
New Attacks	۰		۹۸	
Total	۲۰۲۰			

ما ۹۸ نوع حمله جدید در دادگان آزمون گنجاندیم تا در فازهای بعدی پژوهش با متدها تشخیص نفوذ و دسته بند های مختلف آزمایش کنیم این میزان حمله جدید تقریباً در حدود ۴,۸۵ درصد از کل حجم دادگان آزمون ما را

تشکیل می دهند. در نتیجه دو مجموعه آزمون و یادگیری را با هم ترکیب کردیم تا در مجموع ۴۰۴۰ رکورد ترافیک تصادفی بدست آمد. این دادگان نفوذ دارای انواع کلاسهای مختلف و تمامی حملات از هر نوع می باشد که برای آموزش یک شبکه عصبی که به عنوان دسته بند یادگیری در یک متد راپر کافی است. همچنین نرخ های یادگیری ، تصدیق و آزمون را به ترتیب ۶۰ ، ۱۵ و ۲۵ درصد وارد کردیم.

۴-۲-۱- بَهره اطلاعات

بهره اطلاعات به عنوان معیاری کلی برای ارزیابی میزان ارتباط هر ویژگی با کلاس در بردار مربوطه می باشد. این معیار مبتنی بر تئوری اطلاعات است و بستگی به حجم دادگان s دارد. [۲۴]

اگر دادگان نفوذ حاوی m کلاس را در نظر بگیریم بطوریکه S_i نمونه از کلاس I موجود باشد ، در اینصورت میزان بهره اطلاعاتی مورد انتظار در دسته بندی نمونه ها به صورت رابطه زیر است :

$$I(s_1.s_2 \dots .s_m) = -\sum_{i=1}^{m} \frac{s_i}{s} \log(\frac{s_i}{s}) \qquad (۴-۱)$$

یک ویژگی F با مقادیر ممکن $\{f_1.f_2 \dots f_v\}$ می تواند دادگان نفوذ را به v زیر مجموعه به صورت $\{S_1.S_2 \dots S_v\}$ تقسیم می شود که در آن S_j ، زیر مجموعه ای است که مقدار f_j را برای ویژگی F دارد. بنابراین S_j شامل s_{ij} نمونه از کلاس i می باشد. پس آنتروپی ویژگی F به صورت زیر است :

$$E(F) = \sum_{j=1}^{v} \frac{s_{1j} \dots s_{mj}}{s} \times I(S_{1j} \dots S_{mj}) \qquad (۴-۲)$$

بهره اطلاعات ویژگی F :

$$Gain(F) = I(s_1.s_2 \dots .s_m) - E(F) \qquad (۴-۳)$$

نمودار زیر بهره اطلاعات بدست آمده هر ویژگی را در دادگان نفوذ مورد آزمایش ما را نشان می دهد. همچنین در بخش آزمایش از بهره اطلاعات به عنوان یک متد تشخیص نفوذ مبتنی بر تئوری اطلاعات مبتنی بر شبکه عصبی به عنوان تابع ارزیابی استفاده کردیم.

مشخصات پارامترهای ورودی این آزمایش برای دو حالت انتخاب/حذف تصادفی و انتخاب/ حذف مبتنی بر بهره اطلاعات هر ویژگی در فاز جستجوی زیر مجموعه ها ، به صورت جدول زیر می باشد. تابع ارزیابی نخست که همان بَرازندگیِ حاصل از دسته بندِ ما را نشان می دهد منطبِق بر نرخ تشخیص و خطای مثبتِ اشتباهِ تابع ارزیاب (شبکه عصبی) در هر فازِ تکرار الگوریتم بوده و به صورت زیر فرموله می شود. ضمناً خطای بدست آمده از مجموع یادگیری و آزمون شبکه عصبی را نیز به عنوان تابع هزینه در نظر گرفتیم.

در مقاله اصلی [۱۰] ، نویسنده برای تابع هدف از درخت تصمیم [1] استفاده کرده و کاربرد شبکه عصبی را به عنوان یک مسئله باز در این حوزه ، برای کارهای آتی پیشنهاد کرده است. ضمن اینکه با بررسی دقیق الگوریتم اصلی مشاهده شد که این الگوریتم به صورت تصادفی در برخی مشاهدات با مشکلِ بهینه محلی[2] مواجه می شود. ما برای رفع این چالشها سعی کردیم که مفهوم بهره اطلاعات را در فازهایی استفاده کنیم که حذف ویژگی را به صورت تصادفی انجام می دهند. از اینرو بَهره اطلاعات را در فاز آغازین و پنجم به کار بردیم بطوریکه حلقه مرحله پنجم را به طور کلی حذف نموده و به جای آن در این مرحله یک ویژگی با کمترین آنتروپی را حذف نمودیم تا مکانیسمِ این الگوریتم را هدفمند سازیم.

از طرفی بحثِ اینکه کدام نوع شبکه عصبی به عنوانِ دسته بندِ یادگیریِ CFA به کار رود مسئله ای است که در بخش بعد با ارزیابیِ مقایسه ای متد MLP و شبکه ی عمیقِ دولایه به نام Stacked Auto Encoder و با دسته بندی ۵ کلاسه به طور کامل بررسی گردید و نهایتاً شبکه عمیق در آزمایشات بعدی به کار رفت.

در دو آزمایش متفاوت (انتخاب/حذف ویژگی به صورت تصادفی یا مبتنی بر آنتروپی هر ویژگی) از شبکه عصبی به عنوان تابع هدف استفاده نمودیم. نتایج بدست آمده حاکی از آنند که شبکه عصبی می تواند دسته بندیِ به مراتب بهتری را نسبت به درخت تصمیم انجام دهد. همچنین نشان داده شده که استفاده از بهره اطلاعاتِ ویژگی ها به منظور انتخاب/حذف در فازهای مختلف CFA نیز که می تواند به طرز موثری نرخ تشخیصِ دسته بندِ یادگیری را بالا برده و خطاهای کاذب را کاهش دهد و از حذف / انتخاب کورکورانه و صرفاً تصادفیِ ویژگی (های) خاص در هر فاز تکرارِ الگوریتم جلوگیری نماید.

به منظور ارزیابیِ بهترین زیر مجموعه ی ویژگی انتخاب شده رابطه ارزیابیِ زیر را به کار بردیم. به این ترتیب که پس از انتخاب زیر مجموعه ویژگی در هر فاز ، شبکه عصبی بلافاصله به تشکیلِ ماتریسِ درهم ریختگی پرداخته و معیارهای خطا و عملکردِ دسته بندی را محاسبه می نماید. تابعِ فیتنسی که در مقاله ارائه شده به صورت رابطه زیر می باشد.

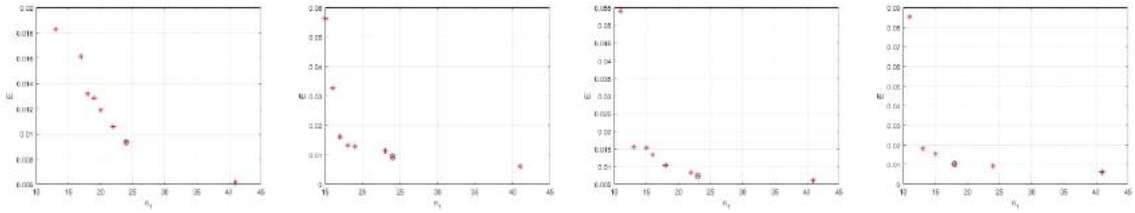
$$Fit = \alpha \times DR + \beta \times (1 - FPR) \qquad (۴-۴)$$

[1] Decision Tree
[2] Local Optimum Problem

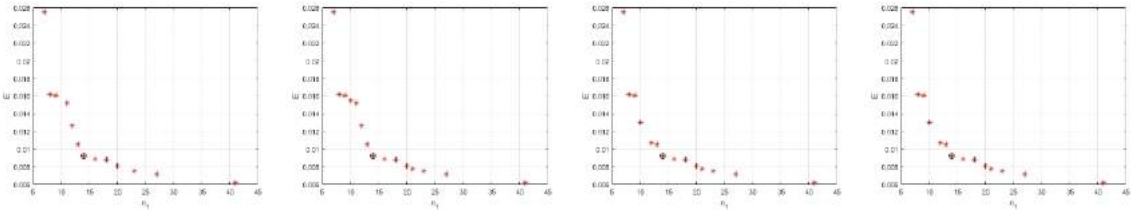
در الگوریتم اصلی ، فاز پنجم حلقه ای را به تعداد ویژگیهای انتخاب شده در AVbestSubset تکرار نموده و هر بار یک ویژگی را حذف می کند تا عملکرد ماهیچه Leocophor در لایه زیرین پوست ده پا را تَداعی کند. ما با بررسی دقیقِ عملکردِ الگوریتم (Trace) در این فاز متوجه شدیم که به جز اتلافِ زمان و افزایشِ پیچیدگیِ فضا و زمانِ الگوریتم هیچ عملکردِ مثبتی ندارد. زیرا از دیدگاه تئوری اطلاعات و بنا به استناد به مقاله [۲۴] صرفاً حذفِ ویژگیِ با کمترین آنتروپی می تواند خروجیِ تابعِ ارزیابی را بهبود بخشد. بنابراین با حذف یک ویژگی که کمترین بهره اطلاعاتی را دارد فاز پنجم اصلاح شد.

جدول ۲ - پیکربندی پارامترهای ورودی

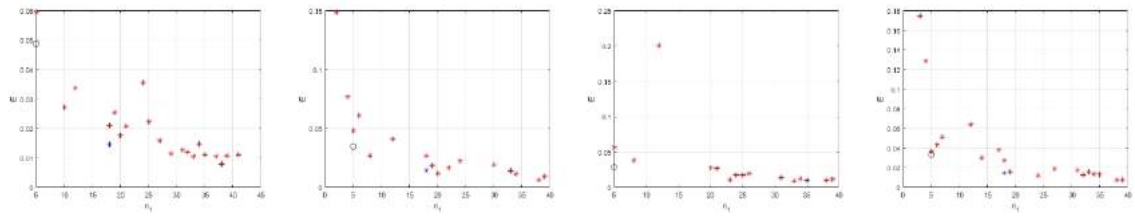
Pop size	Cross over rate (%)	Mutation rate (%)	Parameters
۲۰	۰,۶	۰,۴	NSGA II
Pop size	t		Parameters
۲۰	۱۰		CFA – ANN



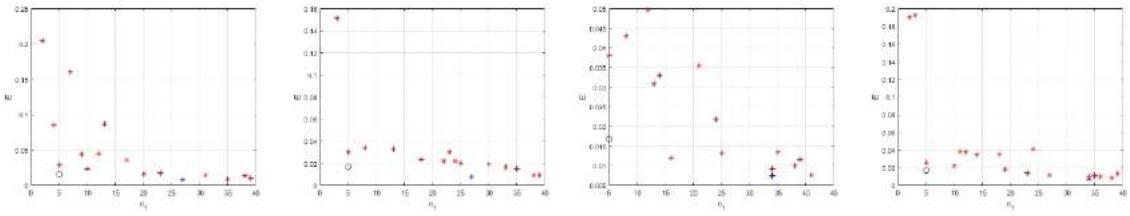
خروجی الگوریتم NSGA II در تکرارهای مختلف (از چپ به راست)



۴ تکرار آخر (از راست به چپ)

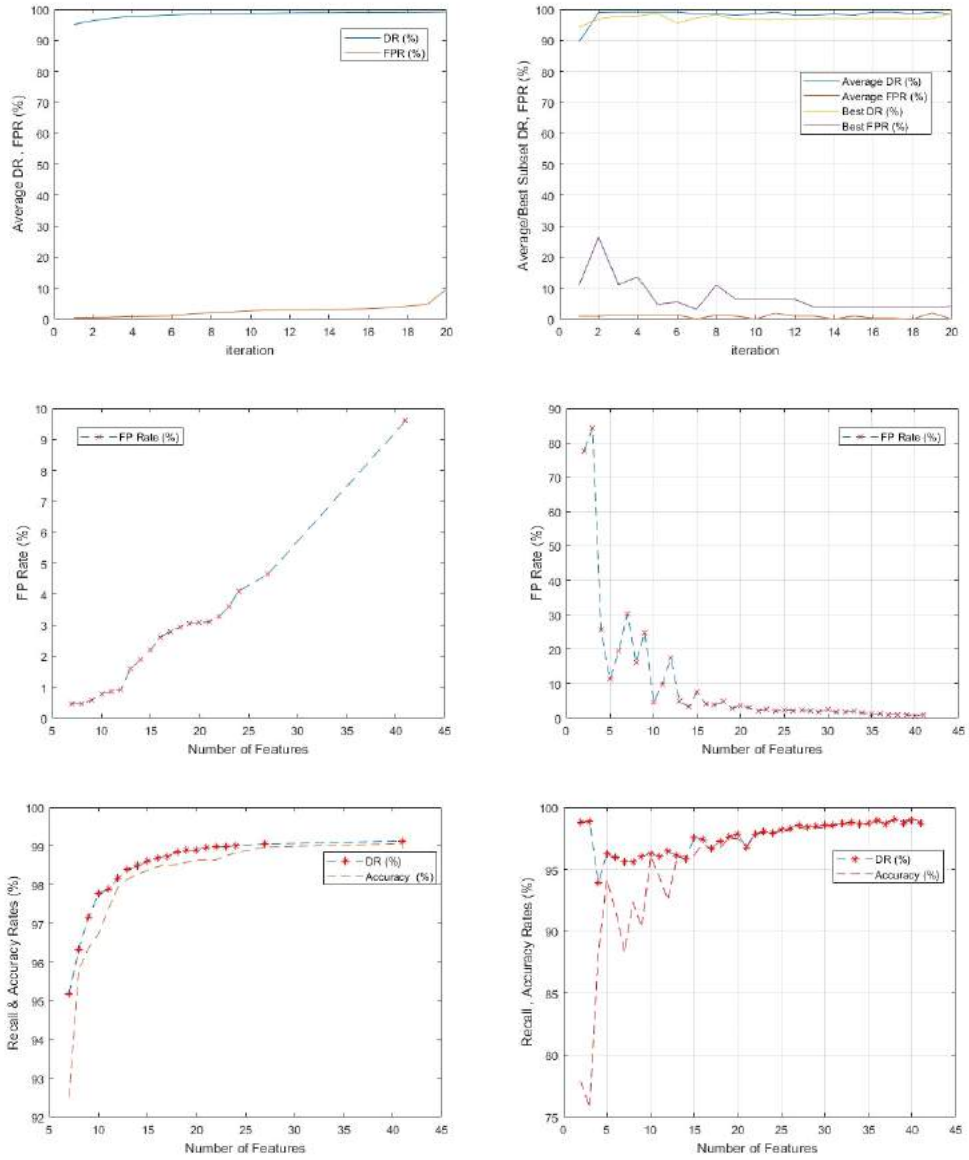


خروجی الگوریتم CFA – ANN در تکرارهای مختلف



۴ تکرار آخر

۳-۴- آنالیز مقایسه ای CFA – ANN و NSGA II



نمودار ۳ – نمودارها ، سمت چپ NSGA II , سمت راست CFA-ANN

جدول ۳- مقایسه بهینه ترین زیر مجموعه بردار ویژگی بدست آمده پس از ۲۰ تکرار

No. of features	*FPR*(%)	*DR*(%)	*ACC*(%)	*Fitness*(%)	*Type*	*Method*
۵	۲,۵۷۸۸	۹۹,۰۷۳	۹۸,۵۹۷	۹۸,۵۷۷۵	*Wrapper*	$Cfa - ANN$
۲ ۳ ۱۲ ۳۲ ۴۰						$Op - subset$
۱۴	۰,۹۴۳۴	۹۸,۹۹۳۳	۹۹,۰۱	۹۹,۰۱۲۳		*NSGA II*
۲ ۳ ۴ ۷ ۱۰ ۱۲ ۱۳ ۱۵ ۲۸ ۲۹ ۳۲ ۳۳ ۳۴ ۴۰						$Op - subset$
۱۶	۱,۸۳۲۵	۹۸,۷	۹۹,۱	۹۹	*Filter*	*Information Gain* (*IG*)
۲ ۳ ۴ ۵ ۱۸ ۱۹ ۲۱ ۲۲ ۲۹ ۳۰ ۳۱ ۳۶ ۳۷ ۳۸ ۳۹ ۴۱						$Op - subset$

مطابق جدول فوق در هر سه مورد ، اندازه لایه مخفی را ۲۵ گرفتیم. متد IG بر خلاف رویکرد های راپر، تکرار نداشته و فقط یکبار اجرا شد. سپس مقادیر آنتروپی بدست آمده از تمام ویژگیها را به ترتیبِ نزولی مرتب سازی نموده و به اندازه ی بهترین تعداد ویژگیهای بدست آمده از دو رویکرد راپر ، از این ترتیبِ نزولی ، ویژگی انتخاب نمودیم و با شبکه عصبیِ آزمون کردیم. همانطور که قبلاً نیز اشاره شد یک مزیّت مهم رویکرد های راپر نسبت به فیلترِ علیرغم زمان بالای پردازشی و حافظه مورد نیاز آنها ، حذف ویژگیهای غیر مرتبط یا اضافی با انجام تکرارها و آزمون های مکرّر مبتنی بر آزمون و خطا در این الگوریتم هاست. از آنجایی که هدف از طراحی متدهای انتخابِ ویژگی بخصوص با رویکرد راپر ، کاربرد آنها در سیستم های هوشمندی همچون سیستم تشخیص آنومالی میباشد که پیوسته با دادگانِ نفوذِ حجیم سرو کار دارند.

با مشاهده ی نمودارها با استفاده از رابطه فیتنس فوق می توان بهترین نقاط را در هر تکرار الگوریتم بدست آورد و با bestSubset و AVbestSubset در همان تکرار مقایسه کرد تا بهترین زیر مجموعه ویژگی نهایی بدست آید. مزیّت رابطه فوق در این است که تمامی معیارهای موثر در انتخاب زیر مجموعه ویژگی بهینه مانند دقّت تشخیص ، خطا و حتّی تعداد ویژگیهای انتخاب شده را نیز لحاظ می کند.

۴-۴- یک بررسی

از جمله نقاط ضعفِ الگوریتمِ اصلی دَه - پا می توان به عدم قابلیت اطمینان به عملکرد آن در طی فازهای مختلف اجرا و انتخاب زیر مجموعه بهینه از ویژگی ها را نام برد. به عبارت دیگر طبق بررسی های انجام شده از نتایج و تحلیل داده های بدست آمده از آزمایشات ، bestSubset در بهینه محلی گیر می کند و تکرار زیادی لازم است تا از آن خارج شود. البته با در نظر گرفتن AVbestSubset و میانگین های بردارهای ویژگی بدست آمده از تمامی فازها ، می توان آنالیز بهتری را انجام داد که نشان می دهد این الگوریتم نیاز به بهبود دارد.

همچنین تصادفی بودن حذف ویژگیها یکی دیگر از نقاط ضعف این الگوریتم است زیرا حداکثر تعداد انتخاب های ممکن برای ایجاد زیر مجموعه بهینه بسیار بیشتر از آنست که در حد و اندازه توان پردازشی و زمان اجرای آزمون و محاسبه برای تمام انتخابها در عمل میسر باشد. بدین معنی که برای انتخاب بهینه ترین بردار ویژگی نهایی در دادگان حجیم می بایست تکرار های زیادی انجام شود که مقرون به صرفه نیست.

۵-۴- جمع بندی

در شبیه سازیِ CFA از بهره اطلاعات به منظور حذف یا انتخاب تصادفی ویژگیهای موجود در فاز جستجو استفاده شده و یک شبکه عصبیِ نیز به عنوان دسته بند جهت آموزش و ارزیابی هزینه زیر مجموعه انتخاب شده در این الگوریتم به کار رفت. همانطور که از نتایج آزمایشات قابل استنباط است استراتژی جستجوی مبتنی بر دَه - پا در فازهای مختلف آزمون و خطاهای متعددی انجام می دهد تا تجربه لازم جهت انتخاب بهینه ترین بردار ویژگی را کسب نماید و به تَبَعِ آن حجم حافظه و زمان پردازش زیادی برای دادگان نفوذ لازم است.

بخش پنجم – ارزیابی مقایسه ای دو شبکه عصبی به منظور کاربرد جهت دسته بندِ یادگیریِ CFA

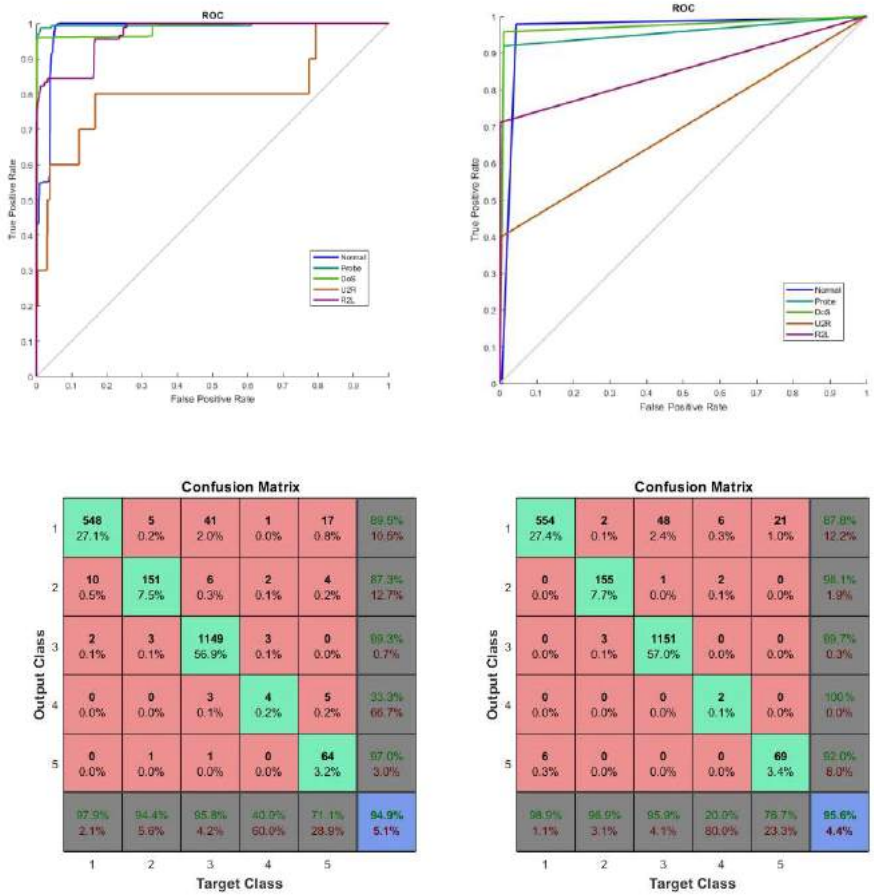
طبق جدول ۴ مشاهده می گردد که بهینه ترین پیکربندی متد MLP در آزمایش چهارم توانسته است بهترین نتایج را حاصل کند ، اما در مقام مقایسه با متد یادگیری عمیق نمی تواند رقابت نماید. بطوریکه در آزمایش ۱۵ بهینه ترین پیکربندی برای یادگیری عمیق ، نتایج آن را بهبود داده است. ما در آزمایشی دیگر اندازه لایه مخفی MLP را از ۲۰ به ۱۰۰ افزایش دادیم تا رقابت MLP را با متد یادگیری عمیق (Stacked Auto Encoder) مشاهده کنیم. نتایج زیر از این آزمایش بدست آمد. همانطور مشاهده می شود، با افزایش لایه مخفی نیز متد MLP باز هم توانایی رقابت با یادگیری عمیق با دو لایه مخفی را ندارد. بنابراین متد یادگیری عمیق فوق بهترین گزینه برای استفاده به عنوان "دسته بند یادگیری" در الگوریتم پیشنهادی انتخاب ویژگی می باشد.

method	run	hidden size		L2 Weight Regularization	Sparsity Regularization	Sparsity Proportion	Decoder Transfer Function	scale data	rec	acc	fpr	fnr	tpr	tnr	f-measure	spec	sen	err	prec	ec
MLP	1	20		train ratio =80 ,validation ratio = 20					94.9315	0.9619	0.5357	5.0685	94.9315	99.4643	1.9749	0.9946	0.9978	0.0381	0.9978	0.9117
	2								94.8630	0.9609	0.7143	5.1370	94.8630	99.2857	1.9735	0.9929	0.9486	0.0391	0.9971	0.9093
	3								93.5615	0.9490	1.6071	6.4384	93.5615	98.3929	1.9660	0.9839	0.9356	0.0510	0.9935	0.8829
	4								95.3425	0.9653	0.3571	4.6575	95.3425	99.6429	1.9764	0.9964	0.9534	0.0347	0.9986	0.9194
	5								94.8630	0.9624	0.1786	5.1370	94.8630	99.8214	1.9777	0.9982	0.9486	0.0376	0.9993	0.9133
AVG-MLP1									94.7123	0.9599	0.389282	5.28768	94.7123	99.32144	1.9737	0.9932	0.9568	0.0401	0.99726	0.90732
MLP	6	25		train ratio = 85,validation ratio =15					94.5205	0.9584	0.7143	5.4795	94.5205	99.2857	1.9734	0.9929	0.9452	0.0416	0.9971	0.9041
	7								94.5890	0.9525	3.0357	5.4110	94.5890	96.9643	1.9553	0.9696	0.9459	0.0475	0.9878	0.8878
	8								94.7945	0.9614	0.3571	5.2055	94.7945	99.6429	1.9763	0.9964	0.9479	0.0386	0.9986	0.9109
	9								94.9315	0.9634	0	5.0685	94.9315	100	1.9792	1	0.9493	0.0366	1	0.9157
	10								95.2055	0.9653	0	4.7945	95.2055	100	1.9792	1	0.9521	0.0347	1	0.9199
AVG-MLP2									94.7542	0.960206	0.82142	5.1918	94.8082	99.17858	1.97268	0.99178	0.94808	0.0398	0.9967	0.90768
MLP	11	41		train ratio =85 ,validation ratio = 15					95.2740	0.9644	0.5357	4.7260	95.2740	99.4643	1.9750	0.9946	0.9527	0.0356	0.9978	0.9170
	12								94.8630	0.9619	0.3571	5.1370	94.8630	99.6429	1.9763	0.9964	0.9486	0.0381	0.9986	0.9120
	13								95.2740	0.9653	0.1786	4.7260	95.2740	99.8214	1.9778	0.9982	0.9527	0.0347	0.9993	0.9197
	14								94.5205	0.9564	1.4286	5.4795	94.5205	98.5714	1.9678	0.9857	0.9452	0.0436	0.9942	0.8987
	15								95.0685	0.9634	0.3571	4.9315	95.0685	99.6429	1.9764	0.9964	0.9507	0.0366	0.9986	0.9152
AVG-MLP3									95	0.96228	0.57142	5	95	99.42858	1.97466	0.99426	0.94998	0.03772	0.9977	0.91252
AVG-MLP									94.8221	0.96079	0.5940	5.1598	94.84	99.30953	1.97368	0.99308	0.95162	0.0392	0.99722	0.90917
St-Auto	1	41	25	0.001/0.001	4/4	0.05/0.1	purelin/ purelin	false	95.1370	0.9649	0	4.8630	95.1370	100	1.9792	1	0.9514	0.0351	1	0.9189
	2	41	25	0.001/0.005	8/4	0.1/0.1	logsig/ logsig	false	95.3425	0.9634	1.0714	4.6575	95.3425	98.9286	1.9575	0.9893	0.9534	0.0366	0.9889	0.9433
	3	41	20	0.005/0.001	4/8	0.05/0.05	purelin/ logsig	true	95.2740	0.9658	0	4.7260	95.2740	100	1.9792	1	0.9527	0.0342	1	0.9538
	4	25	41	0.005/0.005	8/8	0.1/0.1	logsig/ purelin	true	95.2055	0.9589	2.3214	4.7945	95.2055	97.6786	1.9326	0.9768	0.9521	0.0411	0.9762	0.9291
	5	20	41	0.001/0.001	4/4	0.1/0.05	logsig/ logsig	false	95.1370	0.9649	0	4.8630	95.1370	100	1.972	1	0.9514	0.0351	1	0.9525
AVG-St-Auto1									95.2192	0.96358	0.67856	4.7808	95.2192	99.32144	1.9641	0.99322	0.9522	0.03642	0.99302	0.93952
St-Auto	3	41	25	0.001/0.001	4/4	0.05/0.1	purelin/ purelin	true	100	0.7228	100	0	100	0	0.9950	0	1	0.2772	0.5	0
	7	41	25	0.001/0.005	8/4	0.1/0.1	logsig/ logsig	false	95.3425	0.9639	0.8929	4.6575	95.3425	99.1071	1.9611	0.9911	0.9534	0.0361	0.9907	0.9452
	8	41	20	0.005/0.001	8/8	0.05/0.05	purelin/ logsig	true	95.4795	0.9673	0	4.5205	95.4795	100	1.9793	1	0.9548	0.0327	1	0.9558
	9	41	20	0.005/0.001	8/8	0.05/0.05	purelin/ logsig	false	95.4795	0.9673	0	4.5205	95.4795	100	1.9793	1	0.9548	0.0327	1	0.9558
	10	41	41	0.001/0.001	8/8	0.05/0.05	purelin/ logsig	false	95.2055	0.9653	0	4.7945	95.2055	100	1.9792	1	0.9521	0.0347	1	0.9532
AVG-St-Auto2									96.3014	0.91732	20.17858	3.6986	96.3014	79.82142	1.77878	0.79822	0.96302	0.08268	0.89814	0.762
St-Auto	11	41	25	0.005/0.005	8/8	0.1/0.1	logsig/ logsig	false	94.9315	0.9599	1.2500	5.0685	94.9315	98.7500	1.9537	0.9875	0.9493	0.0401	0.9870	0.9375
	12	41	25	0.001/0.001	4/4	0.1/0.1	logsig/ logsig	false	95.4110	0.9634	1.2500	4.5890	95.4110	98.7500	1.9539	0.9875	0.9541	0.0366	0.9871	0.9421
	13	41	20	0.001/0.001	4/4	0.1/0.1	logsig/ logsig	true	95.3425	0.9653	0.3571	4.6575	95.3425	99.6429	1.9719	0.9964	0.9534	0.0347	0.9963	0.9507
	14	41	41	0.005/0.005	8/8	0.05/0.05	purelin/ logsig	false	95.3425	0.9663	0	4.6575	95.3425	100	1.9792	1	0.9534	0.0337	1	0.9545
	15	41	10	0.001/0.001	8/8	0.05/0.05	purelin / logsig	false	95.8219	0.9693	0.1786	4.1781	95.8219	99.8214	1.9757	0.9982	0.9582	0.0307	0.9981	0.9572
AVG-St-Auto3									95.36988	0.96484	0.60714	4.63012	95.36988	99.39286	1.96688	0.99392	0.95368	0.03516	0.9937	0.9484
AVG-St-Auto									95.63016	0.94858	7.15476	4.36984	95.63016	92.84524	1.903253	0.9285	0.9563	0.05142	0.96162	0.8833

جدول ۵ – مقایسه بین بهترین نتایج بدست آمده از آزمایشات جدول ۴

method	MLP	St-Auto
Run	۴ / new	۱۵
hidden size	۱۰۰	۴۱ / ۱۰
L2 Weight Regularization	train ratio =80 =,validation ratio ۲۰	۰,۰۰۱/۰,۰۰۱
Sparsity Regularization		۸/۸
Sparsity Proportion		۰,۰۵/۰,۰۵
Decoder Transfer Function		/purelin
Scale data		false
rec	۹۵,۳۴۲ / ۹۴,۸۶۳۰	۹۵,۸۲۱۹
Acc	۰,۹۶۵۳ / ۰,۹۶۹۰	۰,۹۶۹۳
fpr	۰,۳۵۷۱ / ۰,۷۱۴۳	۰,۱۷۸۶
fnr	۴,۶۵۷۵ / ۵,۱۳۷۰	۴,۱۷۸۱
tpr	۹۵,۳۴۲ / ۹۴,۸۶۳۰	۹۵,۸۲۱۹
tnr	۹۹,۶۴۲ / ۹۹,۲۸۵۷	۹۹,۸۲۱۴
f-measure	۱,۹۷۶۴ / ۱,۹۷۳۵	۱,۹۷۵۷
Spec	۰,۹۹۶۴ / ۰,۹۹۲۹	۰,۹۹۸۲
sen	۰,۹۵۳۴ / ۰,۹۴۸۶	۰,۹۵۸۲
err	۰,۰۳۴۷ / ۰,۰۳۹۱	۰,۰۳۰۷
prec	۰,۹۹۸۶ / ۰,۹۹۷۱	۰,۹۹۸۱
cc	۰,۹۱۹۴ / ۰,۹۰۹۳	۰,۹۵۷۲

نکته ای که وجود دارد اینست که در بهترین حالتها برای دو متد فوق (آزمایش ۴ مربوط به شبکه عصبی و آزمایش ۱۵ مربوط به شبکه عمیق) نرخ منفی اشتباه به ترتیب از ۴,۶۵۷۵ و ۴,۱۷۸۱ پایینتر نمی رود که نشان دهنده عدم شناسایی حملات ناشناخته تعبیه شده در دادگان آزمون توسط دو متد فوق است. با استناد به بخش قبل که اشاره کردیم که دادگان آزمون ایجاد شده در مجموع ۹۸ مورد حمله جدید ناشناخته در خود دارد که از این تعداد ۲۵ مورد از نوع شناسایی ، ۵۲ مورد انکار سرویس ، ۵ مورد U2R و ۱۶ مورد نیز از نوع R2L هستند. بنابراین در آزمایشی مستقل با بردار ۴۰ ویژگی ، با بررسی دقیق و ارزیابی نتایج دسته بندی دو متد فوق با بهترین پیکربندی پارامترهای ورودی (۱۵ و ۴) مشاهده گردید که متد یادگیری عمیق مذکور در بدترین حالت قادر به کشف ۱۵,۳۰ درصد از حملات ناشناخته می باشد که نرخ بسیار پایینی است. (نمودار های زیر)



نمودار ۴ - ماتریس درهم ریختگی و منحنی
(سمت راست متد شبکه عصبی و سمت چپ متد یادگیری عمیق)

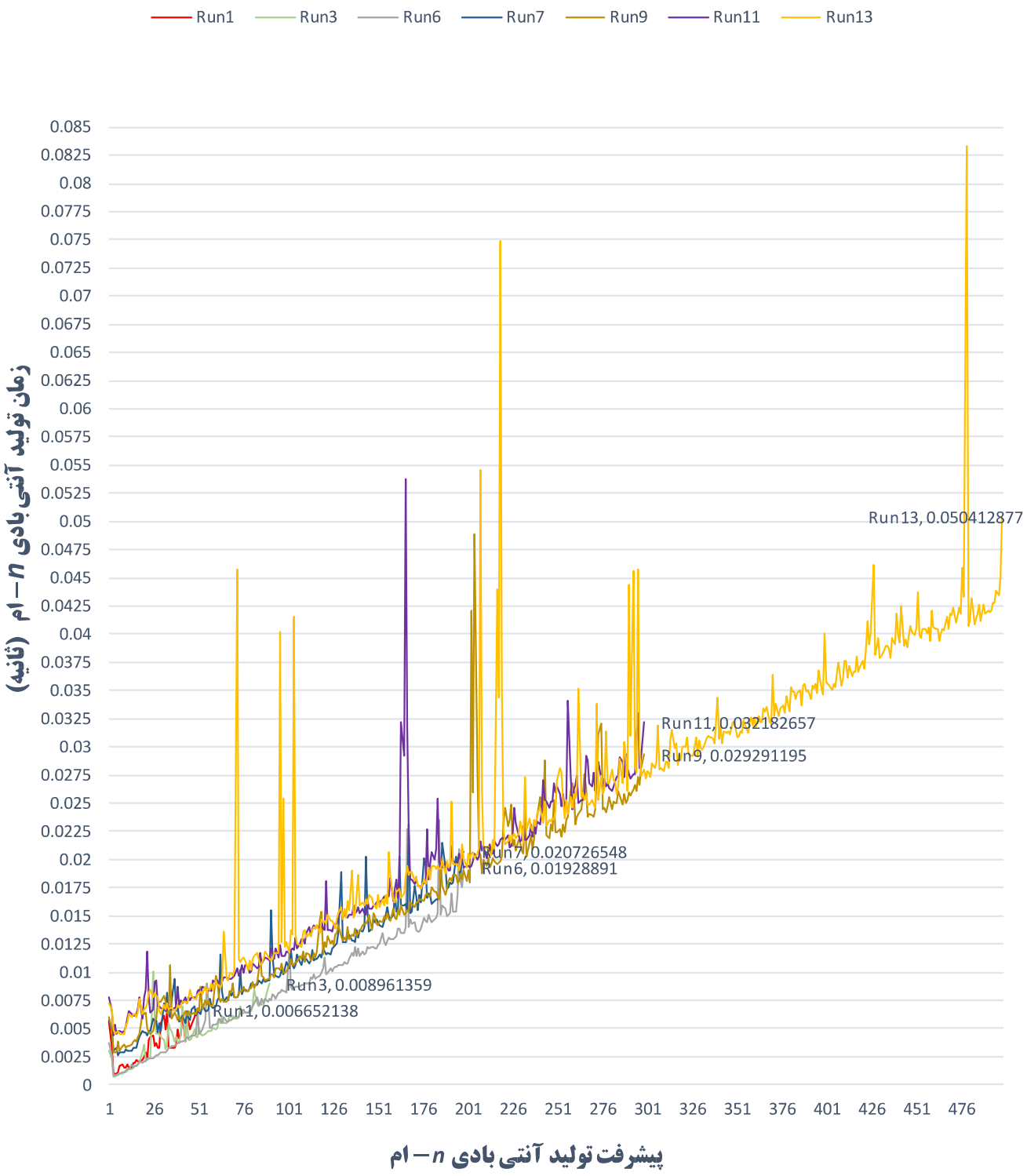
بخش ششم – آزمایشِ زمانِ تولیدِ تشخیص دهنده های غیر خودی در فازِ یادگیری اوّلیه در مکانیسمِ انتخابِ منفی

نمودار های ۵ تا ۷ " زمان لازم برای تولید آنتی بادیهای بالغ بر مبنای پیشرفت تولید آنها " را نشان می دهد. در این آزمایش خاص ، پارامترهای FtNSA [6] به صورت جدول زیر تعیین گردید. از این نمودارها قابل ملاحظه است که با پیشرفت تولید آنتی بادیها ، زمان لازم برای تولید آنها نیز افزایش می یابد. این به دلیل فضای بالای تحت پوشش آنتی بادیهای تولید شده در مراحل اوّلِ اجرای الگوریتم می باشد .

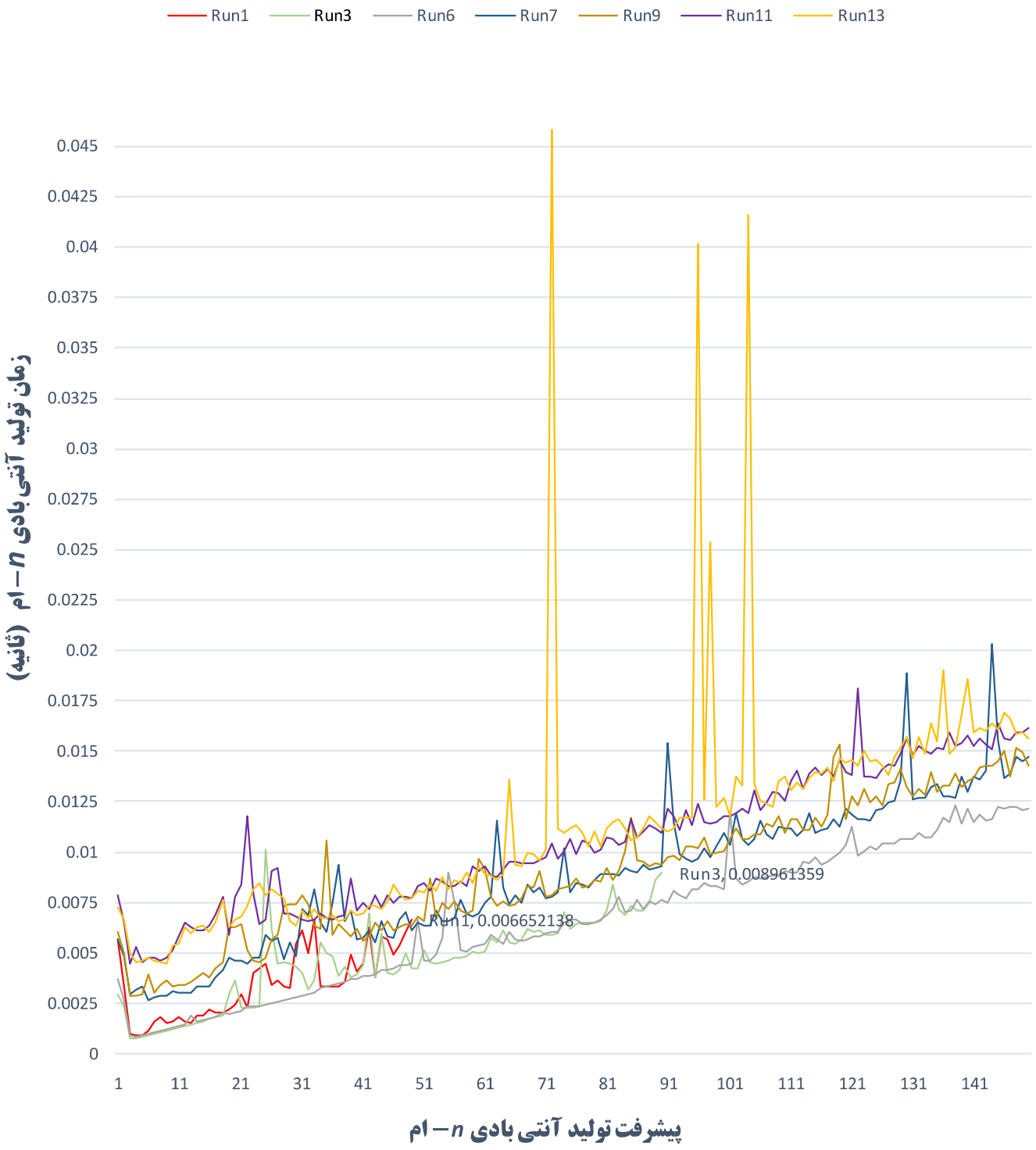
به مرور زمان با پیشرفت تولید آنتی بادیها ، فضای حفره ها به تدریج کاهش یافته و با توجه به اینکه بُردارِ موقعیّتِ آنتی بادیها صرفاً به صورت تصادفیِ یکنواخت تولید میگردد پس جهت بررسی اینکه آیا بردار موقعیّتِ آنتی بادی های جدیداً تولید شده در ناحیه حفره ها قرار دارند یا نه ، زمان زیادی صرف می گردد. این زمان به دلیل کاهش و تنگ تر شدنِ فضای حفره های موجود به تدریج سیر صعودی پیدا نموده و پوشش حفره های باقیمانده نیز به مراتب مشکل تر میشود که لازم است استراتژی خاصی بدین منظور در نظر گرفته شود.

جدول ۶ - تست های مربوط به زمان تولید تصادفیِ آنتی بادیها
صرفاً با در نظر گرفتن پروفایل خودیِ نرمال در فضای ابعادِ مسئله

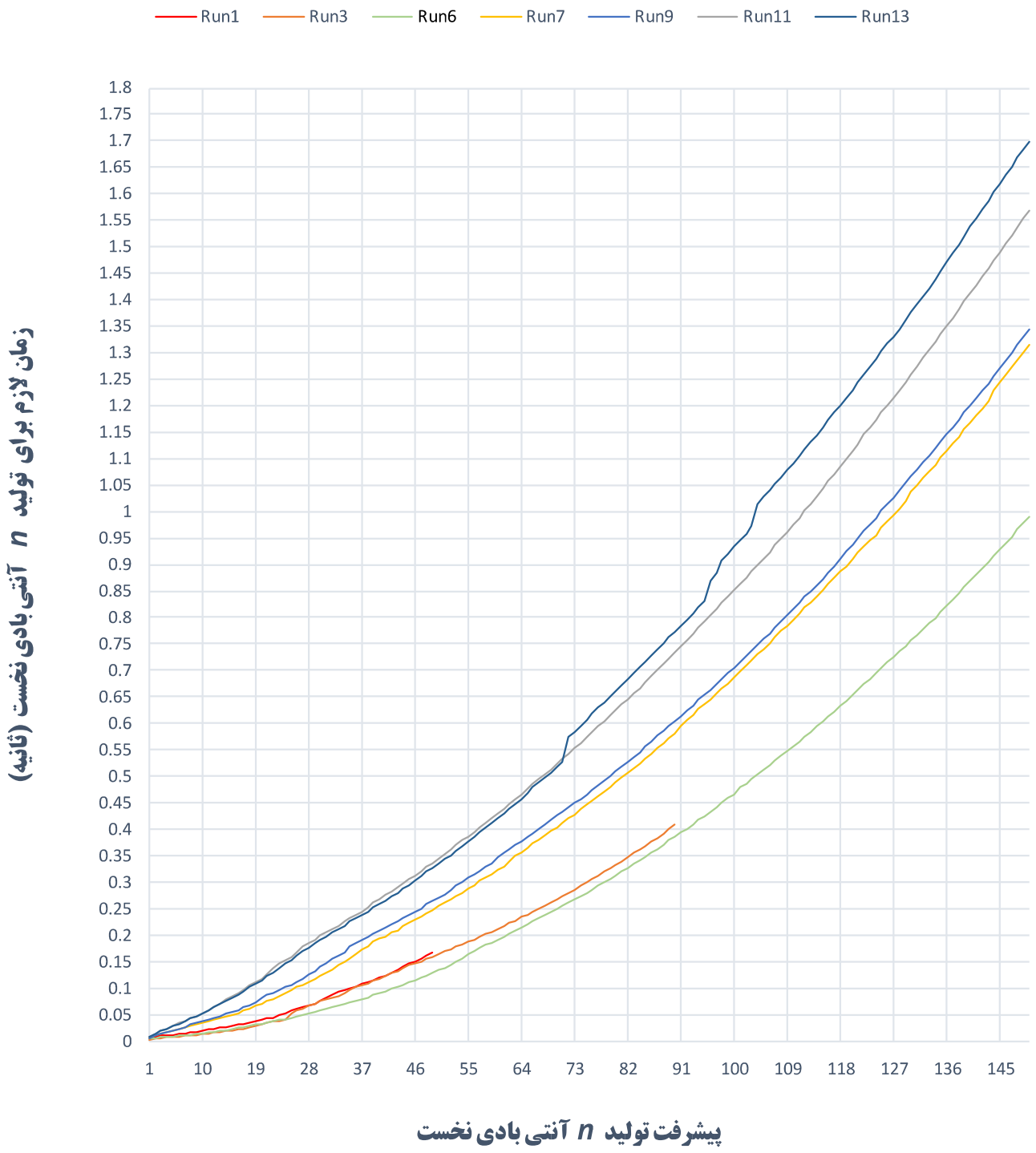
Run	$Self - norm$	rS	$Tdmax$	*Elapsed time of Normal Profile Generation (Seconds)*	*Total Time of Generation of Antibodies (Seconds)*
1	500	0.01	*49*	0.000422	0.1685
2	500	0.01	89	0.000062	0.389986
3	500	0.02	89	0.000184	0.4083
4	1000	0.01	89	0.000187	0.4028
5	1000	0.01	199	0.000406	1.7663
6	1000	0.02	199	0.000177	1.7234
7	5000	0.02	199	0.000238	2.1556
8	5000	001	199	0.000176	2.2093
9	5000	0.02	299	0.000179	4.5570
10	5000	0.01	299	0.000181	4.3484
11	9076	0.02	299	0.000181	5.0383
12	9076	0.01	299	0.000173	5.0262
13	9076	0.02	499	0.000235	12.4744
14	9076	0.01	499	0.000204	12.3084



نمودار ۵ - زمان مورد نیاز برای تولید تصادفی آنتی بادیهای بالغ بر مبنای پیشرفت پوشش ناحیه غیر خودی



نمودار ۶- زمان مورد نیاز برای تولید تصادفی ۱۵۰ آنتی بادی نخست بر مبنای پیشرفت پوشش ناحیه غیر خودی



نمودار ۷ - زمان تجمّعی مربوط به تولید آنتی بادیها صرفاً با اعمال پروفایل خودی-نرمال

بخش هفتم - بررسی نظریِ دلیلِ تاثیرِ متدِ انتخابِ ویژگی در متدهای دسته بندیِ مُتمرکزگِرا مبتنی بر یادگیریِ ماشین

در این قسمت تحلیلی کوتاه راجع به میزان تاثیر متدِ کاهشِ ابعادِ مسئله در بهبودِ دقّت و نرخ های دسته بندی بررسی میکنیم. به طور کلّی استراتژی دسته بند های متمرکز گِرا بر خلاف متدهای دسته بندی ایمنی مصنوعی مبتنی بر فاصله نبوده و از جمعیّت نمونه اوّلیه و یک تابع فانکشن به منظور کشف مرکزیّتها استفاده می کنند. در واقع در این نوع دسته بندها مسئله ی دسته بندی از طریق کشف دقیق ترین مرکزیّتها و با آزمون و خطا در طول فازهای تکرار متعدّد صورت میگیرد. البته یافتن مرکزیّتها خود یک چالش برای این گونه دسته بندها بوده و هست.

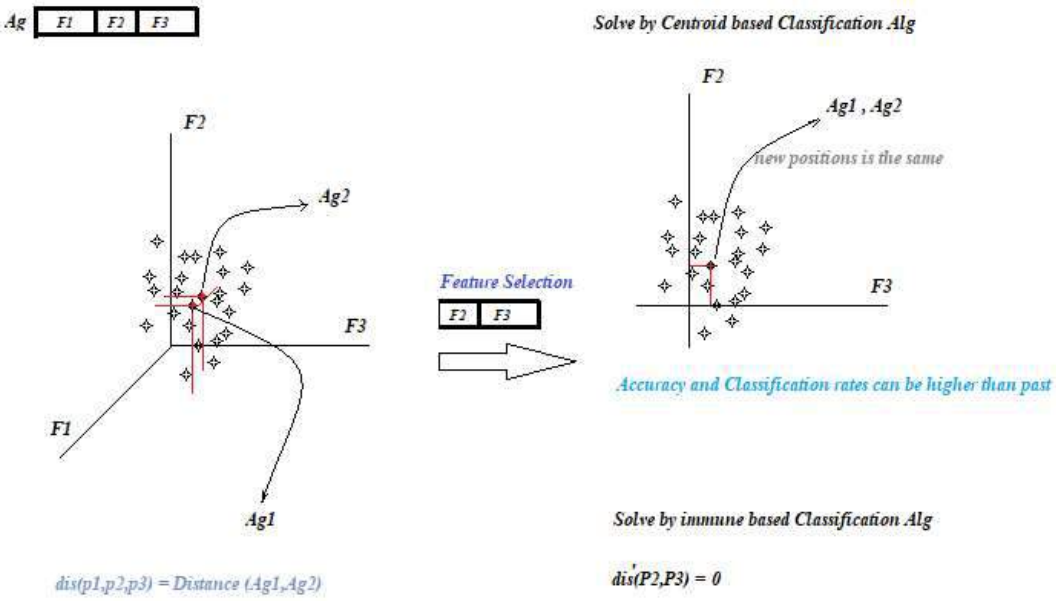
سوال - چرا متد کاهش ابعاد در نتیجه ی دسته بندی متدهای متمرکزگِرا در مقایسه با متدهای ایمنی مصنوعی بسیار موثّرتر است؟

زیرا اگر یک یا چند ویژگی مرتبط / غیر مرتبط حذف گردند ، با کاهش ابعاد مسئله ابعاد مرکزیّتها نیز کاهش یافته و دسته بندی در ابعاد جدید انجام خواهد یافت. بدین صورت مسئله در ابعاد جدیدی باز هم قابل حل خواهد بود و البته سریعتر و شاید دقیق تر. لذا این دقّت و سرعت منجر به کاهش خطای دسته بندی نیز خواهد گردید.

به عبارت دیگر، دسته بندهای متمرکزگِرا جهت حل مسئله ی دسته بندی چندان وابسته به ابعاد نیستند برخلاف متدهای دسته بندی مبتنی بر ایمنی مصنوعی که جهت حل مسئله و تعیین فواصل بین نمونه های خودی-غیرخودی وابستگی بالایی به تعداد ابعاد مسئله دارند. اگر از این تعداد کم شود ، تاثیر این کاهش در محاسبه فواصل محسوس خواهد بود و بر دقّت دسته بندی و خطا تاثیر مستقیم خواهد گذاشت زیرا استراتژیِ پردازش و دسته بندیِ آنها به شکلِ توزیع شده و نامتمرکز می باشد. ما این تحلیل را با ارائه مثالی به صورت اینفوگرافیک زیر واضحتر توضیح داده ایم.

همانگونه که از شکل بالا مشاهده می شود اگر فرض کنیم فضای ابعاد مسئله سه بعدی بوده و بدین ترتیب دو نمونه $Ag_1(a,b,c)$, $Ag_2(a',b,c)$ دارای سه ویژگی باشند تاثیر کاربرد متد انتخاب ویژگی بر عملکرد دو سیستم دسته بندی متمرکزگِرا و ایمنی مصنوعی متفاوت خواهد بود.

اگر فرض کنیم که انتخاب ویژگی صرفاً ویژگی $F1$ را حذف نماید در این صورت ابعاد مسئله از سه به دو کاهش می یابد که با اینکار دو نمونه ی مذکور در فضای جدید بر روی هم می افتند. (شکل زیر سمت راست)



شکل ۷ - مقایسه ی تاثیر انتخاب ویژگی بر دسته بند متمرکزگِرا و دسته بند ایمنی مصنوعی

در نتیجه اگر از متد دسته بندی مبتنی بر جمعیّت نمونه اولیه استفاده شود محاسبه فیتنس مرکزیّتها و یافتن بهترین موقعیت برای آنها در ابعاد دو بعدی حل می گردد ولی در مورد متد ایمنی مصنوعی وضع به گونه دیگری است، با کاهش ابعاد به دلیل آنکه مقادیر این دو نمونه در ابعاد F2, F3 با هم برابر می باشند فاصله ای بین این دو در فضای ابعاد جدید صفر خواهد شد که این مورد خطا را افزایش می دهد و تاثیر نامطلوبی بر عملکرد دسته بند نهایی می گذارد.

در مثال فوق ، اگر تعداد این نمونه ها که دو ویژگی یکسان داشته و صرفاً در چند ویژگی با هم متفاوت اند (مثل ویژگی اوّل در دو نمونه فوق ، $a \neq a'$) زیاد باشد نتیجه ی کاهشِ اَبعاد در دسته بندهای ایمنی مصنوعی بسیار نامطلوب خواهد بود زیرا بر روی هم افتادن نمونه ها و یا کاهش فاصله ی آنها در فضای ابعاد کاهش یافته منجر به کاهش عملکرد دسته بندی نهایی می گردد.

از طرفی به دلیل آنکه در معیار فاصله بین دو نمونه در فضا تمامی ویژگیهای آن دو نمونه در محاسبات بطور مستقیم تاثیر می گذارند و در واقع مَجذور حاصلجمع توان دو های اختلافات میان تمامی ویژگیهای دو نمونه محاسبه می گردند (فاصله ای اقلیدسی) بنابراین حذف یک یا چند ویژگی خاص تاثیر نامطلوبی در کاهش اندازه ی فاصله ی واقعی آن دو نمونه از هم می گذارد. تجربه ی حاصل از آزمایشات مربوط به شبیه سازی های ما در نرم افزار متلب نیز این نکته ظریف را به اثبات می رساند.

بخش هشتُم – ایده ی استفاده از یک خوشه بند جهت تخصیص سیگنالهای ورودی

ایده پیشنهادی بدین صورت است که پس از فاز خوشه بندی ، میانگین فواصل داده های هر خوشه از مرکزیّتِ همان خوشه را اندازه میگیریم. این میانگین به عنوان شعاع دایره ای به مرکزیّتِ همان خوشه در نظر گرفته می شود. در هر دو خوشه ، داده های درون این دایره داده هایی با توزیع استاندارد هستند که فاصله شان از مرکز خوشه به عنوان معیاری برای سنجش درجه سیگنال های امن ، خطر و تحریک به کار می روند. البته ، نواحی خارج از هر دو دایره نیز داده هایی هستند که درجه سیگنال نرمال یا آنومالی برای آنها پایین بوده و در عوض درجه بالایی از سیگنال تحریک را به خود اختصاص می دهند. این داده ها به دلیل نحوه توزیع شان در فضای ابعاد مسئله و عدم قرار گرفتن در ناحیه درون یکی از دو دایره ، معمولاً دارای بیشترین احتمال خطاهای اشتباه هستند. نحوه تخصیص سیگنال در بخش زیر به طور مفصّل تشریح شده است.

پس از فاز خوشه بندی ترافیک شبکه ، نحوه توزیع داده ها و موقعیت خوشه ها در یکی از چهار وضعیت زیر ممکن است قرار گیرند. بسته به اینکه حاصل دسته بندی ما کدام وضعیت باشد ، تخصیص سیگنالهای ورودی خطر باید بر اساس روابط مشخصی فرموله گردد. این روابط بر اساس نحوه و نتیجه ی خوشه بندی متفاوت خواهد بود.

مطابق جدول زیر ما پارامترهای زیر را برای تعیین مقادیر حدود آستانه ، فواصل از مرز و مراکز خوشه ها در نظر گرفتیم. برای تعیین حد آستانه نرمال بودن یا آنومالی بودن از " میانه فواصل داده های هر خوشه از مرکز آن خوشه " استفاده نمودیم. امّا چرا "میانه فواصل " معیار مناسبی نسبت به "میانگین فواصل " جهت تعیین حدود آستانه می باشد ؟

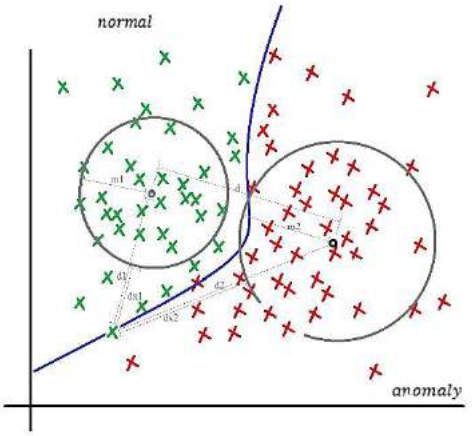
علّت انتخابِ "میانه" ، در این نکته است که توزیع داده های ترافیک شبکه معمولاً از نظم خاصی برخوردار نبوده و غالباً در اکثر دادگان نفوذ مشاهده می شود که نحوه توزیع داده ها در فضای ابعاد مسئله به گونه ای است که پراکندگی داده ها یکنواخت نیست و نوعی بی نظمی مشاهده میشود. بطوریکه حتّی ممکن است داده های نویزی با موقعیت بسیار دورتر از مراکز خوشه ها باشند و علیرغم اینکه به مرز میان دو خوشه نزدیک اند بسته به استراتژی خوشه بند و نزدیکی به مراکز ، به اشتباه در یکی از دو خوشه دسته بندی شوند. در نتیجه در دسته بندی صحیح بسیاری از عوامل موثرند مانند استراتژی خوشه بند در تعیین درست مراکز خوشه ، عدم وجود نویز و نقاط پراکنده در توزیع داده ها در فضای مسئله و عواملی از این قبیل.

بنابراین "میانگین" معیار مناسبی در این گونه موارد نیست چراکه این معیار ، صرفاً میزان تراکم را در نواحی نزدیک به مراکز خوشه را نشان می دهد و داده های دورتر عملاً در تعیین حد آستانه نقش موثری نخواهند داشت. امّا در مورد معیارِ میانه اینطور نیست زیرا در حالت کلّی سنجشِ درستی از نحوه توزیع داده ها ارائه می دهد.

جدول ۷ - پارامترهای مورد نیاز در تخصیص سیگنالهای خطر برای نتیجه خوشه بندی

Formula	Description
$m_1 = Median(X_i.Cent_{norm})$	میانه فواصل داده های نرمال X_i از مرکز خوشه نرمال (حدآستانه نرمال بودن)
$m_2 = Median(X_i.Cent_{anom})$	میانه فواصل داده های آنومالی X_i از مرکز خوشه آنومالی (حدآستانه آنومالی بودن)
$d = dist(Cent_{Normal}.Cent_{Anomaly})$	فاصله اقلیدسی بین مراکز دو خوشه
$d_1 = ABS(d_{x1} - m_1)$	فاصله داده X_i از مرز حدآستانه نرمال
$d_2 = ABS(d_{x2} - m_2)$	فاصله داده X_i از مرز حدآستانه آنومالی
$d_{x1} = dist(X_i.Cent_{Normal})$	فاصله داده X_i از مرکز خوشه نرمال
$d_{x2} = dist(X_i.Cent_{Anomaly})$	فاصله داده X_i از مرکز خوشه آنومالی
d_{max1}	بیشترین فاصله داده نرمال X_i از مرکز خوشه نرمال
d_{max2}	بیشترین فاصله داده آنومالی X_i از مرکز خوشه آنومالی
d_{Far1}	فاصله دورترین داده نرمال X_i از مرکزیت خوشه آنومالی
d_{Far2}	فاصله دورترین داده آنومالی X_i از مرکزیت خوشه نرمال
$O = m_1 - \frac{ABS(m_1 - m_2)}{2}$	فاصله مرکزیت نرمال از میانه ناحیه اشتراکی

- **حالت نخست :** در این وضعیت ، هیچ گونه اشتراکی بین دو خوشه میانگین وجود ندارد. در این حالت برای تخصیص سیگنالهای ورودی بسته به اینکه داده X_i در کدام یک از سه ناحیه درون خوشه ها یا بیرون آن قرار داشته باشد ، روابط زیر را خواهیم داشت.



- روابط تخصیص احتمال سیگنال (۱) :

$$if \quad d > m_1 + m_2 \qquad .then$$
$$if \quad d_{x1} \le m_1$$

$$\rightarrow \begin{cases} P_{SS} = \left(1 - \frac{d_{x1}}{d_{max1}}\right) \times 100 \quad . 0 \le d_{x1} \le m_1 \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \frac{d_2}{d - m_2}\right) \times 100 \quad . d - (m_1 + m_2) \le d_2 \le d - m_2 + m_1 \end{cases}$$

$$else\ if \quad d_{x2} \le m_2$$

$$\rightarrow \begin{cases} P_{DS} = \left(1 - \frac{d_{x2}}{d_{max2}}\right) \times 100 \quad . 0 \le d_{x2} \le m_2 \\ P_{SS} = 0 \\ P_{PAMP} = \left(1 - \frac{d_1}{d - m_1}\right) \times 100 \quad . d - (m_1 + m_2) \le d_1 \le d - m_1 + m_2 \end{cases}$$

$$else\ if \quad Out\ of\ Cluster\ Normal$$

$$if\ \ ABS(d_1 - d_2) \le d - m_2 \quad .\ then$$

$$\rightarrow \begin{cases} P_{SS} = 100 - P_{PAMP} \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \frac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

$$else\ if\ ABS(d_1 - d_2) > d - m_2\ .$$

$$\rightarrow \begin{cases} P_{SS} = 100 - P_{PAMP} \\ P_{DS} = 0 \\ P_{PAMP} = (1 - \frac{d_2}{d_{Far1} - m_2}) \times 100 \end{cases}$$

$$else\ if \quad Out\ of\ Cluster\ Anomaly$$

$$if\ \ ABS(d_1 - d_2) \le d - m_1 \quad .\ then$$

$$\rightarrow \begin{cases} P_{DS} = 100 - P_{PAMP} \\ P_{SS} = 0 \\ P_{PAMP} = \left(1 - \frac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

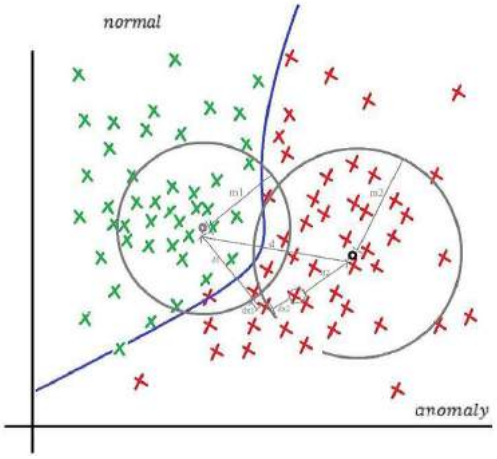
$$else\ if\ ABS(d_1 - d_2) > d - m_1\ .$$

$$\rightarrow \begin{cases} P_{DS} = 100 - P_{PAMP} \\ P_{SS} = 0 \\ P_{PAMP} = (1 - \frac{d_1}{d_{Far2} - m_1}) \times 100 \end{cases}$$

(۱-۸)

- **حالت دوّم** : در این حالت ، بین دو مجموعه خوشه میانگین ، ناحیه اشتراکی[1] به شکل زیر ایجاد می شود. بسته به اینکه نمونه X_i در کدام ناحیه از توزیع داده ها قرار گیرد ، برای تخصیص احتمال سیگنالهای خطر در این حالت چهار رابطه مختلف خواهیم داشت.

[1] joint zone



- روابط تخصیص احتمال سیگنال (۲) :

$$if \quad d \le m_1 + m_2 \quad \& \quad (d > m_1 \quad \& \quad d > m_2) \quad .then$$

$$if \quad d_{x1} \le m_1 \quad \& \quad d_{x2} \le m_2$$

$$\rightarrow \quad \begin{cases} P_{SS} = \left(1 - \dfrac{d_{x1}}{d_{max1}}\right) \times 100 & .0 \le d_{x1} \le m_1 \\ P_{DS} = \left(1 - \dfrac{d_{x2}}{d_{max2}}\right) \times 100 & .0 \le d_{x2} \le m_2 \\ P_{PAMP} = \left(1 - \dfrac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

$$else\ if \quad d_{x1} \le m_1 \quad \& \quad d_{x2} > m_2$$

$$\rightarrow \quad \begin{cases} P_{SS} = \left(1 - \dfrac{d_{x1}}{d_{max1}}\right) \times 100 & .0 \le d_{x1} \le m_1 \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{d_2}{O}\right) \times 100 & . \ \dfrac{O}{2} \le d_2 \le m_1 - \dfrac{O}{2} \end{cases}$$

$$else\ if \quad d_{x1} > m_1 \quad \& \quad d_{x2} \le m_2$$

$$\rightarrow \quad \begin{cases} P_{DS} = \left(1 - \dfrac{d_{x2}}{d_{max2}}\right) \times 100 & .0 \le d_{x2} \le m_2 \\ P_{SS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{d_1}{O}\right) \times 100 & . \ \dfrac{O}{2} \le d_1 \le m_2 - \dfrac{O}{2} \end{cases}$$

$$else\ if \quad d_{x1} > m_1 \quad \& \quad d_{x2} > m_2$$

$$\{Out\ of\ Cluster\ Normal\} \quad \rightarrow$$

$$if\ \ ABS(d_1 - d_2) \le d - m_2\ \ .\ then$$

$$\rightarrow \quad \begin{cases} P_{SS} = 100 - P_{PAMP} \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

$$else\ if\ ABS(d_1 - d_2) > d - m_2\ .$$

$$\rightarrow \quad \begin{cases} P_{SS} = 100 - P_{PAMP} \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{d_2}{d_{Far1} - m_2}\right) \times 100 \end{cases}$$

$$else\ if\quad Out\ of\ Cluster\ Anomaly$$

$$if\ \ ABS(d_1 - d_2) \le d - m_1\ \ .\ then$$

$$\rightarrow \quad \begin{cases} P_{DS} = 100 - P_{PAMP} \\ P_{SS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

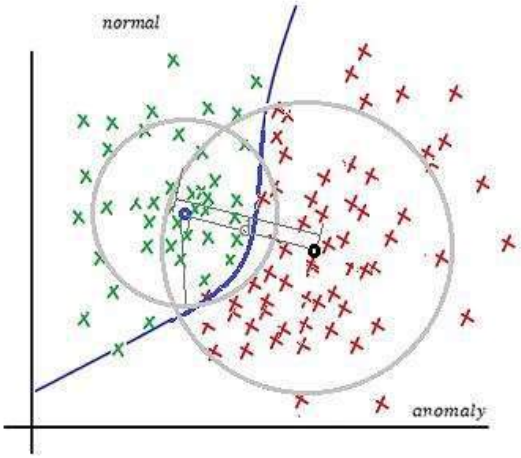
$$else\ if\ ABS(d_1 - d_2) > d - m_1\ .$$

$$\rightarrow \quad \begin{cases} P_{DS} = 100 - P_{PAMP} \\ P_{SS} = 0 \\ P_{PAMP} = (1 - \dfrac{d_1}{d_{Far2} - m_1}) \times 100 \end{cases}$$

(۲-۸)

- **حالت سوّم :**

در این حالت ، بین دو مجموعه خوشه میانگین ، ناحیه اشتراک به شکل زیر ایجاد می شود. بطوریکه فاصله بین دو مرکز خوشه ، اینبار کمتر از شعاع مجموعه بزرگ باشد. در نتیجه بسته به اینکه نمونه X_i در کدام ناحیه از توزیع داده ها قرار گیرد ، برای تخصیص احتمال سیگنالهای خطر در این حالت چهار رابطه مختلف خواهیم داشت.



- روابط تخصیص احتمال سیگنال (۳) :

$$if\ \ d \le m_1 + m_2 \quad \& \quad (d < m_2 \quad or \quad d < m_1) \quad .then$$

$$if \quad d_{x1} \le m_1 \quad \& \quad d_{x2} \le m_2$$

$$\rightarrow \begin{cases} P_{SS} = \left(1 - \dfrac{d_{x1}}{d_{max1}}\right) \times 100 \quad .0 \le d_{x1} \le m_1 \\ P_{DS} = \left(1 - \dfrac{d_{x2}}{d_{max2}}\right) \times 100 \quad .0 \le d_{x2} \le m_2 \\ P_{PAMP} = \left(1 - \dfrac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

$$else\ if \quad d_{x1} \le m_1 \quad \& \quad d_{x2} > m_2$$

$$\rightarrow \begin{cases} P_{SS} = \left(1 - \dfrac{d_{x1}}{d_{max1}}\right) \times 100 \quad .0 \le d_{x1} \le m_1 \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{d_2}{m_1}\right) \times 100 \quad .ABS(m_2 - m_1) \le d_2 \le m_1 \end{cases}$$

$$else\ if \quad d_{x1} > m_1 \quad \& \quad d_{x2} \le m_2$$

$$\begin{cases} P_{DS} = \left(1 - \dfrac{d_{x2}}{d_{max2}}\right) \times 100 \quad .0 \le d_{x2} \le m_2 \\ P_{SS} = 0 \\ P_{PAMP} = \left(1 - \dfrac{d_1}{d}\right) \times 100 \quad . \ m_1 \le d_1 \le d \end{cases}$$

$$else\ if \quad d_{x1} > m_1 \quad \& \quad d_{x2} > m_2$$

$$\{Out\ of\ Cluster\ Normal\} \quad \rightarrow$$

$$if \ \ ABS(d_1 - d_2) \le d + m_1 - m_2 \ \ . \ \ then$$

$$\rightarrow \quad \begin{cases} P_{SS} = 100 - P_{PAMP} \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \frac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

$$else\ if\ ABS(d_1 - d_2) > d + m_1 - m_2\ .$$

$$\rightarrow \quad \begin{cases} P_{SS} = 100 - P_{PAMP} \\ P_{DS} = 0 \\ P_{PAMP} = \left(1 - \frac{d_2}{d_{Far1} - m_2}\right) \times 100 \end{cases}$$

$$else\ if \quad Out\ of\ Cluster\ Anomaly$$

$$if\ \ ABS(d_1 - d_2) \leq d - m_1 \quad .\ then$$

$$\rightarrow \quad \begin{cases} P_{DS} = 100 - P_{PAMP} \\ P_{SS} = 0 \\ P_{PAMP} = \left(1 - \frac{ABS(d_{x1} - dx_2)}{ABS(d_{x1} + d_{x2})}\right) \times 100 \end{cases}$$

$$else\ if\ ABS(d_1 - d_2) > d - m_1\ .$$

$$\rightarrow \quad \begin{cases} P_{DS} = 100 - P_{PAMP} \\ P_{SS} = 0 \\ P_{PAMP} = (1 - \frac{d_1}{d_{Far2} - m_1}) \times 100 \end{cases}$$

(۳-۸)

در این بخش ایده ی استفاده از خوشه بندِ حفره سیاه جهت کشف و تخصیصِ احتمالاتِ سیگنالهای ورودی را از حیث تئوری مورد بررسی قرار داده و روابطی را برای هر یک ارائه نمودیم. متاسفانه آزمایشات [پ-الف-۲] در رابطه با ارزیابی این الگوریتم در دادگان نفوذ شبکه جواب نداد و مورد انتظار نبود. (نمودار ۱ و ۲)

در نتیجه ایده استفاده از یک خوشه بند علیرغم نتایج منفی تست این ایده در قالب خوشه بند سیاه چاله ، نیازمند تحقیقات بیشتر خواهد بود.

بخش نُهُم - یک ایده ، استفاده از دانشِ متخصّصان در کشف و تخصیص سیگنالهای ورودی

به منظور کشفِ سیگنالهای خروجیِ متناظر از قبل باید دادگان نفوذ را مورد تحلیلِ ترافیک قرار داد. بنا به استناد به رساله [48] ، نویسنده توابع سیگنالِ ورودی را برای تولید سه سیگنال خطر [1] ، امن[2] و تحریک

[1] Danger signal (DS)
[2] Safe signal (SS)

شده [1] در مکانیسمِ DCA به ترتیب تعداد بسته های شبکه [2] در ثانیه ، اندازه بسته های شبکه [3] (به بایت) تعداد پیام های خطا در ثانیه[4] در نظر گرفته است. ما در دَه درصد از دادگان نفوذ NSL – KDD ترافیک مربوطه را مورد آنالیز قرار دادیم. [پ–ب–۱]

دادگانِ UNSW – NB15 دارای ویژگیهایی است که آن ویژگیها در NSL – KDD وجود ندارند. به عنوان نمونه به نظر میرسد که ویژگیِ بی ثُباتیِ ارسالی[5] و دریافتی [6] در محاسبه ی سیگنالِ PAMP مفید واقع گردند. از طرفی دو ویژگیِ Sload و Dload که به ترتیب تعداد بسته های شبکه در هر ثانیه را نشان می دهند به نظر میرسد که برای تخصیص به سیگنالِ خطر مناسب باشند. همچنین دو ویژگیِ اندازه بسته های شبکه که به ترتیب smeanz و dmeanz هستند نیز به نظر میرسد که نشانه ای از سیگنالِ امن باشند. ایده ای که به ذهن ما رسید این بود که با الهام از مقاله [48] که به ارائه ی منابعِ احتمالیِ کشفِ سیگنالهای ورودی پرداخته ، بتوانیم ضمن ایده پردازی در جهت تخصیصِ سیگنالها ، خروجیِ دسته بندیِ الگوریتم سلوّلهای دندریت را بهبود ببخشیم. در نتیجه روابطی بر این مبنا بر روی ویژگیهای مذکور اعمال نمودیم تا سیگنالها بدست آیند :

In UNSWNB15 DataSet :

Features of 15,16 'th are

Sload & Dload as network packet per second is suitable for Danger Signal (DS)

DS Signal formula :

$m = \text{mean}(\text{feature.15}, \text{feature.16})$

/* So in this Case ,Median also can be used rather than mean metrics */

loop

if $\text{mean}(Ag(i,15), Ag(i,16)) < m$

$P_{ss} = 0;$

else

$P_{ss} = \text{dist}(\text{mean}(Ag(i,15), Ag(i,16)), m)$

end

end loop

Features of 17,18 'th are

SJitter & DJitter (msec)as network Jitter Rate is suitable for PAMP

PAMP Signal formula :

$m = \text{mean}(\text{feature.17}, \text{feature.18})$

/* So in this Case , Median also can be used rather than mean metrics */

loop

if $\text{mean}(Ag(i,17), Ag(i,18)) < m$

$P_{ss} = 0;$

else

$P_{ss} = \text{dist}(\text{mean}(Ag(i,17), Ag(i,18)), m)$

[1] Stimulated signal (is known as PAMP)
[2] Network packet per second
[3] Size of network packets
[4] Error message per second
[5] SJitter Rate
[6] DJitter Rate

```
                end
        end loop
```

Features of 23,24 'th are
smeanz & dmeanz as network packet size retransmitted is suitable for Safe Signal (SS)
Safe Signal formula :

```
m = mean(feature.23, feature.24)
                /* So in this Case , Median also can be used rather than mean metrics */
loop
        if mean(Ag(i, 23),Ag(i,24)) <  m
                P_ss = 0;
        else
                P_ss = dist(mean(Ag(i,23), Ag(i,24)), m)
        end
end loop
```

شکل ۸ – روابطی بر مبنای دانش متخصصان در تحلیل ترافیک شبکه جهت کشف و تخصیص سیگنالهای ورودی

سپس این روابط را بر روی دادگانِ تستِ نخستِ UNSW – NB15 اِعمال و DCA را با سه رویکردِ تخصیص سیگنالها (رویکردِ فعلی استفاده از دانشِ متخصصان با استفاده از روابط پیشنهادی بالا ، رویکردِ انتخاب ویژگی مبتنی بر بهره اطلاعات و رویکرد سوّم با استفاده از خوشه بندِ حفره سیاه بر مبنای روابط پیشنهاد شده در [پ-الف-۸] مورد ارزیابیِ مقایسه ای قرار دادیم.

قسمتی از نتایج بدست آمده مطابق جدول ۹ نمودار ۴ مربوط به بخش ۳-۱-۸-۱ حاکی از آنست که استفاده از خوشه بندی همچون حفره سیاه که در [پ-الف-۸] ایده ی آن مطرح گردید نمی تواند در تخصیصِ سیگنالهای ورودی مورد استفاده قرار گیرد به دلیل آنکه شیوه ی دسته بندی آن به صورت متمرکزگرا بوده (دلایل در [پ-الف-۷] ذکر شده است) و اصولاً به منظورِ دسته بندی دادگانی با ابعاد بالا مانند تشخیص نفوذِ شبکه ساخته نشده است.

از طرفی ارزیابیِ مقایسه ایِ دو رویکردِ دانشِ متخصّصان (روابط ارائه شده در بالا) و بهره اطلاعات نیز نشان می دهند که خروجیِ دسته بندیِ با رویکردِ نخست نرخ های تشخیصِ مثبت و منفیِ (DR و TNR) تقریباً یکسانی دارند همین مشاهده در مورد دو نرخ منفی و مثبت کاذب نیز صادق بوده و آنها نیز در رِنجِ یکسانی قرار دارند. ضمن اینکه برای رویکردِ دوّم که بهره اطلاعات برای تخصیص سیگنالها به کار رفت نتیجه ی DR بسیار بالا بوده و نتیجه ی خطای منفی کاذب نیز پایین ارزیابی شد که نتایج نسبتاً امیدوار کننده ای است.

از سوی دیگر نتایجِ دو نرخ منفی صحیح و مثبت کاذب نیز چندان مطلوب به نظر نمیرسید بطوریکه FPR در مواردی بالا بود و به همین ترتیب. در نهایت ، به نظر میرسد در جهتِ کشفِ رابطه ای دقیق به منظور کشف و تخصیصِ سیگنالها باید بر روی روابطِ ارائه شده در بالا کار تحقیقاتی بیشتری صورت انجام گیرد. در مجموع نتایجِ بدست آمده در خصوصِ استفاده از دانشِ متخصصان و ایده پردازی و ارائه روابط پیشنهادی نسبت به ایده ی استفاده از خوشه بندِ متمرکزگرایی همچون مکانیسمِ سیاه چاله به مراتب بهتر به نظر میرسد.

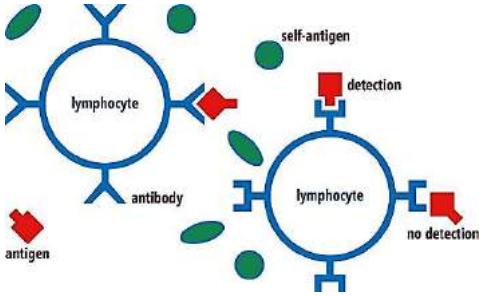
بخش دهُم – شرح جزئیّات مصونیّت های ذاتی و اکتسابی

در این بخش سه واکنش ایمنی به تفصیل بیان شده اند.

۱۰-۱- واکنش ایمنی ویژه

آنتی ژنها بر روی سطح مولکولی شان دارای پروتئینهایی به نام " پِپتید " هستند که به شکل پُرزهایی برآمده هستند. هر آنتی ژنی دارای تعداد زیادی از این پپتیدها هست که همگی یک نوع هستند. بدین معنا که شکلِ سرِ تمام این پُرزها با یکدیگر یکسان هست که می تواند نقطه ضعفی برای یک آنتی ژن در بدن باشند زیرا بدن در مواقع لزوم (شناسایی ویروس/آنتی ژن) ، آنتی بادیهایی را بوسیله سلولهای B ترشح می کند. این آنتی بادیها به عنوان تش.د. های سیستم عمل می کنند و بوسیله شاخکهایشان به آنتی ژنها هجوم آورده و به پُرزهای آنها چِفت و بَست می شوند. (شکل زیر) در حقیقت وجود تعداد زیاد این پرزها بر روی یک آنتی ژن و همچنین ترشّح و تکثیر آنتی بادیها در هنگام شناسایی نفوذ عملاً موجب می شود تا آنتی بادی های زیادی ، یک آنتی ژن خاص را با چسبِشِ شاخکها با پُرزها محاصره نموده و مانع فعالیت آزادانه آن در بدن شوند.

این عملیات بیشتر شبیه کار جاسوسان [1] قلعه است. البته در ادامه خواهیم دید که واکنش هیمورال و در زیر مجموعه آن واکنش ویژه با کمک هم کار شناسایی را همانند جاسوسان در یک قلعه انجام می دهند. بدین ترتیب که شناسایی ویروس ها بوسیله سلولهای naïve B صورت میگیرد و سپس عوامل نفوذ بوسیله آنتی بادیها طی واکنش ویژه محاصره می شوند.



با این عمل ، عملاً مانع فعالیت آنتی ژنیک و تبادل اطلاعات از طریق پروتئینهای بر روی سطح آنتی بادیها و واکنش با سلولهای خودی بدن شده و در نتیجه مانع از آلوده کردن سلولهای خودیِ بدن می شوند. لازم به ذکر است که هر نوع آنتی بادی با توجه به فرمتِ سرِ شاخکهای آن صرفاً می تواند به آنتی ژنی خاص بچسبد.

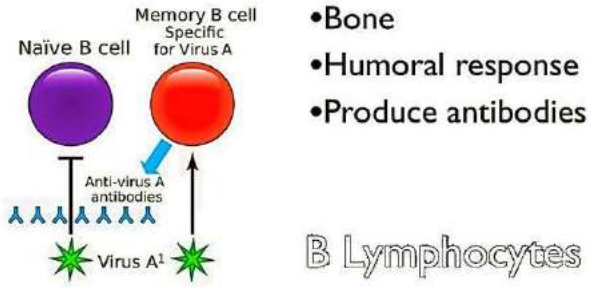
۱۰-۲- واکنش ایمنی هیمورال

سلولهای B شامل دو نوع سلول ساده (naïve B Cell) و سلول حافظه (memory B Cell) می باشند. (شکل ۹) وقتی که نفوذی اتفاق افتاد و ویروسی وارد بدن شد، کار شناسایی ویروس و نوع آن توسط naïve B Cell انجام می شود. این سلولها آنتی ژن را شناسایی نموده و سلّولِ بیِ حافظه را مطلع می کنند

[1] Spies

تا آنتی بادی با الگوهایی نزدیک به آن آنتی ژن را ترشّح کند. پس از مشخص شدن نوع ویروس ، سلول حافظه وارد عمل شده (فعّال سازی) و آنتی بادی متناسب با آن ویروس را جهت مقابله ترشح میکند.

اساساً سلولهای حافظه دو وظیفه مهم را بر عهده دارند یکی به خاطر سپردن اطلاعات مرتبط با حملات (ویژگیهای آنتی ژنها) و دیگری فاز مقابله از طریق ترشّح آنتی بادیها. با ترشّح اولین آنتی بادی به تدریج آنتی بادی ها در بدن تکثیر می یابند تا به سمت اهدافی که قبلاً توسط سلولهای naive B شناسایی شده حمله نموده و آنها را محاصره کنند. (واکنش ویژه)



شکل ۹ – لنفوسیتهای B

با این محاصره عملاً فعالیت آزادانه و وِلگردیِ آنتی ژنها به عنوان عوامل نفوذی در بدن متوقف شده و زمان نابودی آنها فرا میرسد. آنتی ژنی که در دامِ آنتی بادیها اسیر است ، آنقدر در این حالت می ماند تا تحلیل رفته و نابود شود. (این یک روشِ دفعِ آنتی ژن توسط سلّولِ یاخته خوار می باشد در بخش قبل بیان شد) به این نوع رویکردِ ایمنی که مختصِ لنفوسیتهای B می باشد ، واکنش هیومورال می گویند.

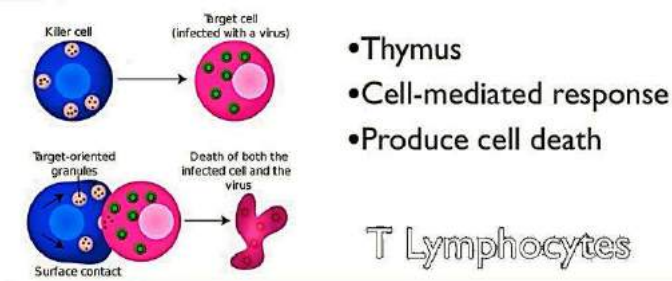
۱۰-۳- مکانیزم ایمنی سلولی (واکنش CM)

مکانیسم ایمنی سلولی به مبارزه با سلولهای آلوده به ویروس و باکتری و سلولهای سرطانی می پردازد. در این روش لنفوسیت های T نقش اصلی را ایفا نموده و آنتی بادی حضور ندارد. سلولهای T عامل نابودی سولهای آلوده و سرطانی بدن به شمار می روند. چالش اصلی آنها بازسازی مجدد و احیای سلولهای زنده و جدید در بدن است. این سلّولهای پروتئینی پس از تولید در مغزِ استخوان، به منظور آموزشهای لازم در غدّه تیموس سازماندهی شده و سپس بالغ و در نهایت وارد بدن می شوند. آنها اساساً به چهار نوعِ زیر تقسیم بندی می شوند که دو نوع اوّلِ آن در این مکانیزم به طور مستقیم نقش دارند. [60]

- سلّولهای T کُشنده یا Killer T Cells

- سلّولهای T کمک رِسان یا Helper T Cells
- سلّولهای T تضعیف کننده[1]
- سلّولهای T خاطره[2]

سلّولهای کُشنده (سیتوتوکسیک ها) توانایی شناسایی و حمله مستقیم را به سلول آلوده یا سرطانی دارند. آنها پس از فعّال شدن و تبدیل به Activated Killer T Cell با ترشّحِ پروتئینی به نام پِرفورین ، منافذی را در این سلّولها ایجاد می کنند که به مرگ آنها و نابودی ویروس[3] (هردو) منجر می شود. شکل زیر ساختار این سلول را بهتر نشان می دهد.



شکل ۱۰ – ساختار یک لنفوسیت کشنده T

سلّولهای کمک رِسان فراوان ترین نوع لنفوسیت های T هستند که به دستگاه ایمنی سلولی کمک می کنند. به نحوی که با عملِ تکثیرِ سلّولی که در بخش های بعدی این پژوهش بیان شده و با رویکرد توزیعی و مشارکتی نقش عمده ای در افزایش سرعت تشخیص و پاسخ به نفوذ بازی می کنند. بطوریکه می توان سلولهای تی را در زُمره ی هر دو مصونیّتِ ایمنی ذاتی و تطبیق پذیر دانست. (شکل ۱۱)

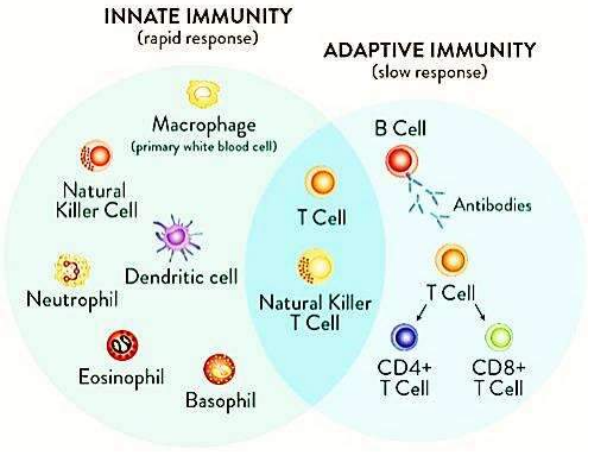
[1] کنترل اعمال سلولهای کشنده و کمک کننده بر عهده سلولهای تضعیف کننده است. این سلولها به سیستم ایمنی خاتمه می دهند و از پاسخ های بیش از حد جلوگیری می کنند.

[2] این سلولها در حال آماده باش می مانند و طریقه مبارزه با ویروسها را یاد می گیرند و به نسل بعدی لنفوسیت های T انتقال می دهند.

[3] دانه های ریزی در درون این سلول وجود دارند که به محض چسبیدن آن ه سلول آلوده ، وارد سلول آلوده شده و موجب نابودی آن می شوند.

این سلولها به محض فعالسازی بوسیله یاخته خواران، نقش اصلی در اتخاذ سیاست امنیتی- دفاعی و تکثیر سلولی ایفا نموده و با بازتولید سلولهای کُشنده T و سلولهای B هردو واکنشِ هیمورال و ایمنیِ سلولی را فعال می کنند تا با عوامل نفوذ مقابله نموده و سلولهای آلوده را از بین ببرند.

عملکرد آنها شبیه به سربازان در هنگام مواجهه با دشمن وارد شده به داخل قلعه است. زیرا سربازان در هنگام مقابله و تیر اندازی با سلاح ضمن از بین بردن دشمن ، طبیعتاً خساراتی نیز به بار می آورند. در دنیای ایمنی زیستی ، اصطلاحاتی وجود دارد از قبیل : Receptor , detector, effector.



شکل ۱۱ - جایگاه سلول تی در میان دو سیستم ایمنی ذاتی و تطبیقی

اولی بیانگر سازوکار پروتئینها و از عوامل داخلی سیستم ایمنی میباشد که رفتاری را در داخل بدن جهت واکنش و مقابله با نفوذ فراهم می کند و با اشتراک اطلاعات نفوذ بین آنها (فعالسازی لنفوسیت) موجب ایجاد تحولاتی نیز می شوند. مثل فعالسازی لنفوسیتهای تی و بی حافظه. تش.د. به عنوان مکانیزم لایه دوّمِ دفاعی عوامل نفوذ را شناسایی نموده و سپس لنفوسیتها را نیز آگاه می کنند تا در مرحله بعدیِ دفاع بتوانند اثر گذار باشند.

۱۰-۴- نَبردی در دو جِبهه

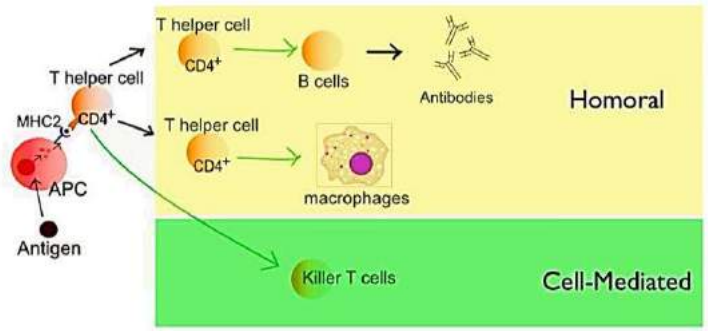
در این بخش سناریوی کلّیِ نحوه شناسایی و مقابله با عامل غیر خودی (آنتی ژنِ نفوذی) و نقش سلول یاخته خوار و سلول کمک رسان بیان شده است. مطابقِ [60] آنچه که در پروسه ی واکنش مربوط به دو مصونیّت در بدن در هنگام بیماری رخ می دهد به صورت زیر تشریح شده است :

۱- آنتی ژن به عنوان عامل نفوذیِ غیر خودی وارد بدن شده و بوسیله یاخته خوار دُرُشت مولکول ، شناسایی شده و بلافاصله بلعیده می شود. این مقابله کننده گان در بدن به نام APC نیز شناخته می شوند[۱] که به محض مشاهده یک نوع آنتی ژن نمونه هایی از هر آنتی ژن را بلعیده و نگه داری می کنند تا در موقع رویارویی با یک لنفوسیتِ تی ، ضمن از بین بردن آنتی ژن نمونه ای از مورد شناسایی شده را برای آنالیز در اختیار لنفوسیت قرار دهند. شبیه به عملکرد یک سرباز در قلعه است زمانی که مهاجمی را دستگیری می کند و تحویل می دهد.

۲- APC در درون خود ، آنتی ژن را به منظور بررسی منجمد نموده و دانه های ریزی را ترشح می کند که ماده شیمیایی MHC2 نامیده می شود. این ماده شیمیایی عامل نابودی آنتی ژن منجمد شده در درون سلول یاخته خوار می باشد. سپس پسماند های آنتی ژن متلاشی شده از طریق منافذ ریزی از درون سلول یاخته خوار به بیرون هدایت می شوند. نکته ای که وجود دارد اینجاست که یاخته خوار مقداری از آنتی ژن را به عنوان نمونه در مَنفَذ خود حفظ می کند تا بعداً به منظور شناسایی و ارائه گزارش به سلولهای لنفوسیت به کار رود.

۳- لنفوسیت T کمک رسان از راه می رسد و با متصل شدن به منفذ APC ، نمونه موجود در منفذ را دریافت می کند تا فعال شود.

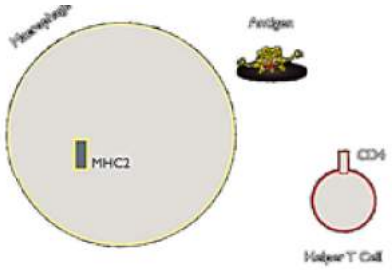
۴- پس از فعال سازیِ سلولهای تیِ کمک رسان ، این سلول به سمت سلولهای B حافظه و تیِ کُشنده و یاخته خوار می رود تا با چسبیدن، آنها را نیز فعال (خبردار) کند. به عبارت دیگر رخداد دفعِ آنتی ژن و مقابله با آن توسط یک سلول تیِ کمک رسان می بایست در حافظه سیستم (سلول B) به خاطر سپرده شود تا موقع رویارویی بعدی با این نوع تهدید ، سیستم به سرعت با تکثیر آنتی بادیها و فعال سازی سایر لنفوسیتهای مشابه جهت مقابله با آن برنامه ریزی موثر و دقیق و سریعی داشته باشد.



شکل ۱۲– نقش سلول تی در واکنشهای هیمورال و ایمنی سلولی

[۱] سلولهای نمونه برداری از آنتی ژنها (Antigen Presenting Cells) : این سلولها به محض مشاهده یک آنتی ژن آن را تحویل یاخته خواران می دهند تا شناسایی نمایند.

۵- سلولهای بیِ فعّال شده و تیِ کُشنده ، تکثیر یافته و به بافتهای حاوی آژ.غ.خد. گُسیل می شوند.
۶- پس از تکثیر کافی نوبت به مقابله با سلولهای آلوده و ویروسها فَرامی رسد. این دو نوع لنفوسیت هر یک به شیوه خود واکنش نشان می دهند. (شکل بالا)



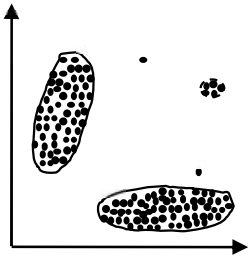
شکل ۱۳ - سناریو فرضی ورود ویروس به بدن و فرایند مقابله سیستم ایمنی بیولوژیک با آن [۱]

لنفوسیتها که در اصل گونه هایی از گلبولهای سفید خون هستند همواره در خون جریان داشته و بوسیله رگهای بدن به تمام بافتها انتقال می یابند. بنابراین آنها در همه جا حضورِ پُر رنگ دارند. به این ترتیب با وجود این توانایی ها در سیستم پیشگیری و مقابله با نفوذ در بدن ، دیگر نیازی به کنترلِ متمرکز از سوی مغزِ جاندار نخواهد بود و سیستم ایمنی به صورت مستقل با هوش و عملکردی کاملاً توزیع شده و مشارکتیِ خود به فعالیتهای سازماندهی شده خود در بدن می پردازد.

بخش یازدهُم – انواع آنومالی

[۱] جهت نمایش انیمیشن، سند را در نرم افزار واژه پردازِ آفیس نسخه ی ۲۰۰۷ به بالا باز کرده و دو بار کلید کنید.

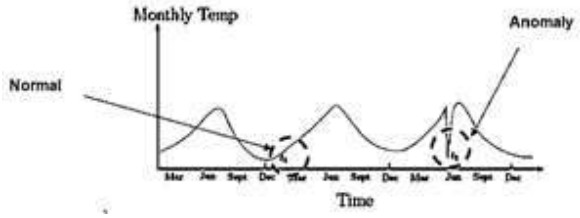
۱۱-۱- آنومالی نقطه ای [1]



مطابق شکل وقتی که یک نمونه داده از الگوهای نرمالِ دادگان (خوشه های اصلی) منحرف می شود ، می تواند به عنوان یک نقطه آنومالی شناخته شود. به عبارت دیگر مطابق شکل رو به رو ، دو نمونه داده [2] دور افتاده از سه خوشه در واقع آنومالی های نقطه ای هستند زیرا متعلق به هیچ کدام از آنها نیستند. (در داخل هیچ یک از خوشه ها قرار نمی گیرند).

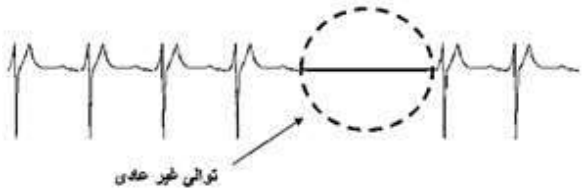
۱۱-۲- آنومالی ضِمنی [3]

وقتی که یک نمونه داده در موقعیتی خاص به یکباره به شکل غیر عادی نسبت به سایر موقعیتهای قبلی رفتار کند. برای مثال مطابق نمودار زیر که سابقه و الگوی توزیع داده ها را در طول زمان نشان می دهد، دو ناحیه t2 و t1 را مشاهده می کنید که در ناحیه t2 منحنی ، انتظار شیبی ملایم مانند موقعیت مشابه آن در t1 می رود اما موقعیت t2 ناگهان مشابه t1 رفتار نمی کند یعنی نسبت به آن موقعیت آنومالی دارد. به این نوع آنومالی شرطی نیز می گویند. برای مثال مصرف هزینه بالا در طول یک هفته جشن و عید و مراسم تعطیل میتواند به شکل یک آنومالی از نوع ضمنی به نظر بیاید. زیرا این هزینه در این دوره زمانی نسبت به سایر روزهای مشابه سال بالاتر است.



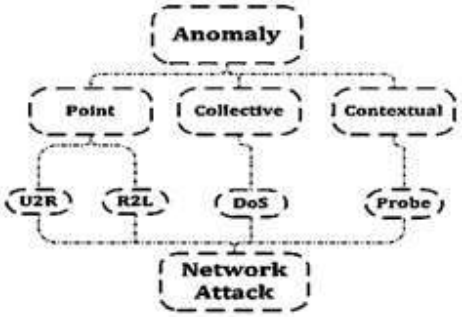
۱۱-۳- آنومالی تجمّعی [4] (به هم پیوسته)

مطابق شکل زیر ، وقتی توده ای از نمونه های داده ای نسبت به کل دادگان به شکل آنومالی و غیر عادی رفتار می کنند به آن آنومالی جمعی گفته می شود. به عبارت دیگر توده ای پیوسته از الگوی توزیع داده ها دارای این خصوصیت باشند که در یک مختصات دو بعدی از زمان ، بر روی منحنی تشکیل توده ای غیر عادی را نسبت به سایر نقاط منحنی بدهند. یک نقطه بر روی این ناحیه به تنهایی نشان دهنده آنومالی نیست.



[1] Point Anomaly
[2] Data Instance
[3] Contextual
[4] Collective anomaly

نکته مهم : از بین دسته بندی های فوق ، تنها آنومالی نوع اول مستقل از زمان می باشد (ارتباطی با زمان ندارد) و دو نوع جمعی و ضمنی در بستر زمان رخ می دهند. در ارتباط با مسائل تشخیص نفوذ ، رفتار آنومالی می تواند به هر یک از این سه نوع رخ دهد.



مثلاً حملات ذاتا توزیع شده مانند DoS/DDoS در واقع نمایانگر آنومالی از نوع جمعی هستند زیرا در این نوع حملات در یک دوره زمانی مشخص درخواست ارتباط بسیار زیاد با یک سرور/ سیستم هدف [1] از نقاط مختلف در شبکه تشکیل توده ای از رفتارهای غیر عادی مشابه را می دهند. لازم به ذکر است حملات از نوع جمع آوری اطلاعات (شناسایی) مانند nmap و portsweep در نهایت می توانند منجر به آنومالی از نوع ضمنی گردند.

سایر حملات انفرادی که به نُدرت اتفاق می افتند و سیستم هدف را تحت تاثیر قرار می دهند مانند R2L و U2R در دادگانِ NSL – KDD نیز از نوع آنومالی نقطه ای شناخته می شوند. [۲۳] شکل فوق این تقسیم بندی را بهتر نشان می دهد.

بخش دوازدهُم – معیارهای ارزیابی خروجی دسته بندی

در این بخش ضمن معرّفی معیارهای مختلف تشخیص و کارائی دسته بندی به ارزیابی و تحلیل روابطی در این خصوص پرداخته ایم.

جدول ۸ – انواع معیارهای ارزیابی تشخیص نفوذ

نوع	معیار	شرح	فرمول
معیارهای تشخیص	نرخ مثبت اشتباه False Positive Rate.	از بین کل تعداد داده های نرمال در مجموعه داده تست ، درصد داده هایی که اشتباهاً حمله تشخیص داده شده ولی در اصل نرمال بوده اند. به این معیار FAR نیز می گویند.	$FPR = \frac{FP}{\text{کل تعداد داده نرمال درداده تست}} \times 100$
	نرخ منفی اشتباه False Negative Rate.	از بین کل تعداد حملات در مجموعه داده تست ، درصد داده هایی که به اشتباه صحیح و نرمال تشخیص داده شده اند ولی در اصل حمله بوده اند.	$FNR = \frac{FN}{\text{کل تعداد حملات}} \times 100$
		TPR=1-FNR	
	نرخ منفی صحیح True Negative Rate.	TNR =1-FPR	

[1] Target

	نرخ تشخیص Detection Rate	$DR = 1 - \frac{FP}{\text{کل تعداد حملات در داده تست}} \times 100$	$DR = \frac{TP}{\text{کل تعداد حملات درداده تست}} \times 100$
	Sensitivity حساسیت	**میزان تشخیص داده غیر نرمال را نشان می دهد.** به این ، معیار کامل بودن[1] نیز گفته میشود. این معیار از بین تمام نمونه های داده ای اعم از مجموع نمونه های مثبت تشخیص داده شده ای که واقعاً مثبت بوده اند بعلاوه نمونه هایی که به اشتباه منفی تشخیص داده شده و واقعاً مثبت بوده اند ، کسری از نمونه هایی را که به درستی مثبت تشخیص داده شده اند را نشان می دهد. بنابراین نشان دهنده نرخ کامل بودن است	$Sensitivity = \frac{TP}{TP + FN}$
	Specificity	میزان تشخیص داده نرمال را نشان می دهد.	$Specificity = \frac{TN}{TN + FP}$
	صحّت Accuracy	به این معیار صحت گفته می شود. این معیار در واقع تعیین کننده میزان تشخیص صحیح نرمال و غیر نرمال را نسبت به کل حالتها نشان می دهد. این معیارها برای ارزیابی و مقایسه انواع دسته بند ها حائز اهمیت اند. معیار خطای دسته بندی از رابطه زیر بدست می آید .این رابطه دقیقاً بر عکس معیار Accuracy در بالا می باشد. کمترین مقدار برابر صفر (بهترین کارائی)و بیشترین مقدار برابر یک (کمترین کارائی) می باشد.	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
	نرخ خطا Error Rate	$ER = \frac{FN + FP}{TP + FP + FN + TN} = 1 - Accuracy$	
معیارهای کارائی	معیار دقت Precision Rate.	این معیار از بین تمام نمونه های مثبت کشف شده (کل نفوذها اعم از مثبتهای درست یا اشتباه) ، کسری از نمونه های داده ای را که در دسته بندی ، مثبت (نفوذ) قلمداد شده و واقعاً نیز مثبت بوده اند را نشان می دهد. به این معیار دقت تشخیص گفته می شود.	$PR = \frac{TP}{TP + FP}$
	Recall	این معیار قسمت گم شده معیار دقت است به عبارت دیگر این معیار درصد نمونه های واقعی حمله که بوسیله دسته بندی پوشش داده شده و تشخیص داده شده اند را نشان می دهد. این معیار معادل با نرخ تشخیص (DR) است و رابطه هردو Recall و DR یکسان است.	DR
	F-Measure	مقدار FM به معنای میانگین هارمونیک (همساز) Recall و Precision می باشد. این رابطه وقتی که فقط یک معیار صحت (accuracy) به عنوان معیار ارزیابی مطلوب در نظر گرفته شده ، پیشنهاد می شود. بنابراین زمانی که معیارهای دیگری به غیر از دقت ، در نتیجه نهایی تاثیر داشته و موثر می باشند ، استفاده از FM پیشنهاد نمی شود.	$FM = \frac{2}{\frac{1}{PR} + \frac{1}{Recall}}$

این روابط به ارائه نتایج دقیقتر در ارزیابی دسته بندی کمک می کنند. واضح است که هر چه قدر معیارهای (FP و FN) پایین تر و همچنین میزان نرخ تشخیص و نرخ صحت یا بالاتر باشد عملکرد (کارائیِ) دسته بندی

[1] Completeness

بهتری را برای IDS ها به دنبال خواهد داشت. همچنین اثر بخشی [1] یک IDS با دو معیار کامل بودن و صحت آن تعیین می شود. با توجه به این نکته که در مسائل دسته بندی ممکن است بین تعداد نمونه های دسته های مختلف توازنی برقرار نباشد، ممکن است یک دسته دارای نمونه های خیلی بیشتر از دسته دیگر باشد، در نتیجه مدل نهایی به سمت دسته با بیشترین نمونه سوق پیدا میکند. بنابراین دسته دارای تعداد نمونه کم عملاً تأثیر چندانی در بهبود یا عدم بهبود کارایی نخواهد داشت .

میتوان نتیجه گرفت که معیار Accuracy در مجموعه داده هایی که دارای دسته های نامتعادل با تعداد مختلف نمونه، معیار مناسبی نمیباشد. تکنیکهای مختلفی برای کاهش این دو نرخ FP و FN ارائه شده است. از جمله تکنیکهای دسته بندی می توان به تکنیک تولید قواعد استنتاجی [2] ، شبکه های عصبی [3] ، ماشین بردار پشتیبان [4] اشاره کرد. معمولاً این نرخ ها در کنار یکدیگر نقش ایفا می کنند و تغییر محسوس هر یک به تنهایی نمی تواند در کارائی تشخیص نفوذ موثر باشد. بنابراین روابط بین این حالت ها توسط محققان به صورت زیر تعریف شده و در بررسی ها و ارزیابی ها مورد ملاک قرار میگیرند. [۲۴]

۱۲-۱- تحلیل روابط

معیارهای سنجش استاندارد که برای ارزیابی مسائل تشخیص نفوذ از سوی جامعه تحقیقاتی امنیت پذیرفته شده و به کار می روند عبارتند از [۲۴]:

الف) Recall (Detection Rate)

ب) False Alarm (false positive) rate

ج) ROC Curve : رابطه بین نرخ تشخیص (DR / Recall) و هشدار خطای کاذِب (FAR) .

نمونه ای از نحوه ارزیابی تشخیص آنومالی را با این منحنی در شکل زیر مشاهده می کنید. از طرفی تعیین IDS بهتر با این معیارها زمانی که بخواهیم دو سیستم را سبک و سنگین کنیم. برای مثال اگر در پژوهشی برای IDS اوّل مقادیر $TPR = 0.8$ و $FPR = 0.1$ بدست آمده باشد در حالیکه در IDS دوم $TPR = 0.9$ و $FPR = 0.2$ بدست بیاید، اگر فقط با این دو معیار بخواهیم کار ارزیابی و سنجش این که کدام IDS عملکرد بهتری داشته را انجام دهیم این کار کار سختی خواهد بود.

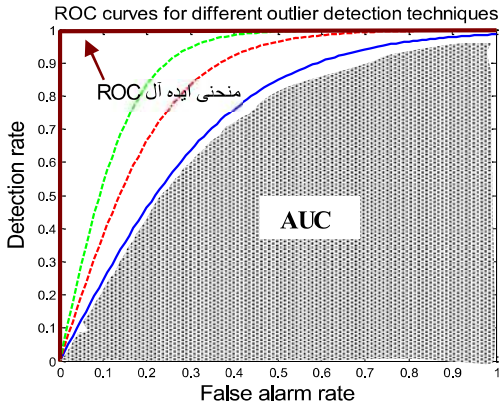
[1] Effectiveness
[2] Inductive rule generation
[3] Neural networks
[4] Support vector machine

در این مواقع منحنیِ RoC برای سبک و سنگین کردن [1] این دو IDS به کار می آید اما این معیار به تنهایی می تواند به ما بگوید که کدام IDS در بیشتر موارد بهتر عمل می نماید؟

برای حل این مشکل معیار جدیدی مبتنی بر علم تئوری اطلاعات از سوی محققان در سال ۲۰۰۶ پیشنهاد شد که معیار CID یا استعداد تشخیص نفوذ نام گرفت. [۲۷]

این معیار دارای ویژگیهای مهمی است زیرا تمام جنبه ها و ابعاد مهم پتانسیل یک سیستم تشخیص نفوذ را مد نظر قرار می دهد. برای یک IDS مقادیر پارامترهای عملیاتی مانند Base Rate, FPR, FNR بسیار اهمیت دارد.

نحوه تعیین این پارامترها بهمراه توضیحات هریک در جدول بالا بیان شده اند.



شکل ۱۴– یک نمونه از منحنیِ RoC و ناحیه زیر آن (AUC) و کاربرد آن در ارزیابی سیستم تشخیص

طبق این جدول معیار B که احتمال وجود نفوذ [2] را در تشخیص نشان می دهد ، مقدار این پارامتر در اغلب مقایسات به دلیل اینکه این احتمال در دسترس نیست به صورت پیشفرض مقداری بسیار کوچک فرض می شود. (معمولاً 10^{-5}) همچنین در بعضی آزمایشات نمی توان بدون این معیار (CID) کار مقایسه را انجام داد. به عنوان مثال اگر IDS اوّل دَه درصد حمله بیشتری را تشخیص بدهد و IDS دوّم نیز بتواند دَه درصد هشدارهای اشتباهش را پایین بیاورد ، کدامیک بهتر عمل نموده اند؟

بدیهی است که یک جواب این خواهد بود که تقریباً هر دو یکسان عمل نموده اند زیرا به اندازه افزایش TP (نرخ تشخیص نفوذ صحیح) به همان اندازه FP (نرخ تشخیص نفوذ نادرست) کاهش یافته است ولی این ساده لوحانه است که بدون در نظر گرفتن سایر معیارها و روابط موجود ، صرفاً معنای این دو را یکسان در نظر بگیریم.

برای حل این مسئله معیار CID برای ارزیابی و انتخاب بهترین پیکربندی یک IDS به کار می رود. فرمول ساده این معیار با رویکرد تئوری اطلاعات به صورت زیر است. [۲۷]

$$C_{ID} = \frac{I(X;Y)}{H(X)} = \frac{H(X)-H(X|Y)}{H(X)} \qquad (۱-۱۲)$$

[1] Trade-off
[2] B = P(I)

در این رابطه $I(X; Y)$ نرخ اطلاعات متقابل x و y می باشد. H(X) نیز آنتروپی X بوده و $H(X|Y)$ آنتروپی شرطی X به شرط Y می باشد که به صورت زیر بدست می آید.

$$H(X) = -\sum_x P(x)\log\big(P(x)\big) = -B\log(B) - (1-B)(\log(1-B)) \qquad (۲-۱۲)$$

بنابراین معیار CID به شکل زیر بدست می آید [۲۷] :

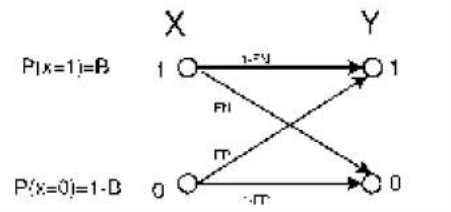
$$C_{ID} = -B(1-\beta)\log(PPV) - B(1-\beta)\log(1-NPV) - (1-B)(1-\alpha)\log(NPV) - (1-B)\alpha\log(1-PPV)$$

(۳-۱۲)

جدول ۹ – پارامترهای مختلف به کار رفته در ارزیابی های سیستم تشخیص نفوذ ، برگرفته از [۲۷]

Term	Equivalent Terms from IDS Literature	Meaning
FP, or α	$P(A\|\neg I)$	False positive rate. The chance that there is an alert, A, when there is no intrusion, $\neg I$.
TP	$(1-\beta)$, $P(A\|I)$	True positive rate. The chance the there is an alert, A, when there is an intrusion, I.
FN, or β	$P(\neg A\|I)$	False negative rate. The chance there is no alert, $\neg A$, when there *is* an intrusion, I.
TN	$(1-\alpha)$, $P(\neg A\|\neg I)$	True negative rate. The chance there is no alert, $\neg A$, when there is *no* intrusion, $\neg I$.
PPV	"Bayesian detection rate", $P(I\|A)$	Positive predictive value. The chance that an intrusion, I, is present when an IDS outputs an alarm, A.
NPV	$P(\neg I\|\neg A)$	Negative predictive value. The chance that there is no intrusion, $\neg I$, when an IDS does not output an alarm, $\neg A$.
B	$P(I)$	Base rate. The probability that there is an intrusion in the observed audit data.

Base Rate (نرخ مبنا): مطابق شکل زیر ، X همان وجود نفوذ است پس اگر فرضاً در مرحله یادگیری، نفوذ موجود باشد [۱] و سیستم تشخیص فرضاً در مرحله تست وجود نفوذ را درست تشخیص دهد این مقدار برابر با B=1-FN خواهد بود. در غیر اینصورت اگر نفوذ اشتباهاً تشخیص داده شد این مقدار معادل FN خواهد بود.



شکل ۱۵ - مدل تشخیص نفوذ مبتنی بر تئوری اطلاعات

به همین ترتیب برای P(x=0) (یا احتمال عدم وجود نفوذ) نیز مطابق شکل فوق می توان معیار FP را در نظر گرفت.

[1] $P(x = 1)$

PPV (نرخ تشخیص بیزین):

$$P(I|A) = \frac{P(I,A)}{P(A)} + \frac{P(I)P(A|I)}{P(I)P(A|I)+P(\neg I)P(A|\neg I)} \quad (۴-۱۲)$$

NPV (مقدار پیش بینی نفوذ[۱]):

$$P(\neg I|\neg A) = \frac{(1-B)(1-\alpha)}{(1-B)(1-\alpha)+B\beta} \quad (۵-۱۲)$$

از طرفی طبق روابط تئوری اطلاعات در بالا و اطلاعات موجود در جدول ۹ ، مقادیر NPV و PPV به معیارهای FP و TP و B وابسته اند. پس با جایگذاری این پارامترها با معادلهای شان ، این احتمالات به صورت زیر بدست می آیند :

$$PPV = \frac{B.TP}{B.TP+\bar{B}.FP} \quad , \quad NPV = \frac{(1-B).(1-FP)}{(1-B).(1-FP)+B.FN} \quad (۶-۱۲)$$

بنابراین طبق دو رابطه فوق و اطلاعات جدول ۹ ، فرمول معادله CID بر حسب ۳ معیار پایه تشخیص نفوذ (FP,FN,TP) به عنوان پارامتر ، به صورت زیر نتیجه گیری و محاسبه می شود که می تواند در ارزیابی ها به کار رود :

$$C_{ID} = -B(1-FN)\log\frac{B.TP}{B.TP+\bar{B}.FP} - B(1-FN)\log\left(1-\frac{(1-B).(1-FP)}{(1-B).(1-FP)+B.FN}\right) - (1-B)(1-FP)\log\frac{(1-B).(1-FP)}{(1-B).(1-FP)+B.FN} - (1-B)FP\log(1-\frac{B.TP}{B.TP+\bar{B}.FP}) \quad (۷-۱۲)$$

نکته : در این فرمول مقدار پارامتر احتمال تشخیص نفوذ یا B را می توان بر حسب آنچه که در شکل مربوط به BR در بالا اشاره شد بر حسب نرخ های FN و FP محاسبه کرد. به منظور پیش بینی نتایج تشخیص نیز می توان معیار دیگری را در آزمایشات استفاده کرد. بنابه استناد به [۲۸] محقق به منظور ارزیابی نتایج مقایسه عملکرد چند تابع کِرنل و الگوریتم انتخاب/استخراج ویژگی به کار رفته در دسته بندِ SVM از شاخص ضریب همبستگی علاوه بر دو معیار DR , FAR استفاده کرده است.

[۱] زمانی که یک سیستم تشخیص نفوذ هیچ آلارمی ندارد این معیار بیانگر شانس آن است که هیچ نفوذی وجود نداشته باشد.

رابطه ضریب همبستگی به شکل زیر می باشد. مقدار cc بین ۱ و ۱- بوده که خروجی ۱ برای cc نشان دهنده این نکته است که نتایج پیش بینی رفتار توسط رویکرد الگوریتمیک به کار رفته در سیستم تشخیص کاملاً با حالت واقعی آن سازگار است. مقدار ۱- نیز نشان دهنده آن است که میزان پیش بینی رویکرد مربوطه تصادفی بوده و با واقعیت ناسازگار است. به این ترتیب با این معیار می توان عملکرد سیستم تشخیص نفوذ را پیش بینی نمود. هر چه قدر این مقدار به ۱ نزدیک تر باشد میزان پیش بینی رفتار الگوریتم/رویکرد به کار رفته در تشخیص نفوذ دقت بالاتری خواهد داشت و به واقعیت نزدیکتر خواهد بود.

$$cc = \frac{TP*TN-FP*FN}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}} \qquad (۸-۱۲)$$

cc در مواقعی کاربرد دارد که به دلیل حجم بالای دادگان ، توان آزمایش و تست درصد بالایی از داده ها امکان پذیر نباشد به این ترتیب با ارزیابی الگوریتم پیشنهادی با انتخاب درصد کمی از دادگان به صورت تصادفی می توان با محاسبه ضریب همبستگی پیش بینی کرد که این الگوریتم در آینده با مجموعه داده بزرگتری با این خصوصیات چه گونه رفتار خواهد نمود. مشخص است که این ضریب می تواند به پیش بینی رفتار آنومالی به سیستم تشخیص و مدیر امنیّت شبکه کمک بسیاری بنماید.

بخش سیزدهُم – تشریح چهار مکانیسمِ ایمنی مصنوعی

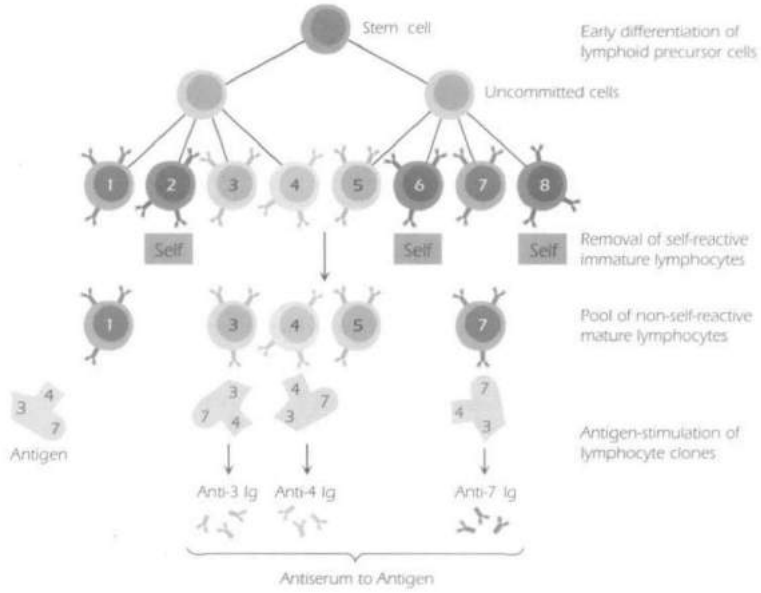
این بخش به توصیف چهار مکانیسمِ ایمنی مصنوعی اختصاص دارد.

۱۳-۱- مکانیسم انتخاب منفی NS

یکی از وظایف مهم غدّه تیموس در بدن سازماندهی و تولید لنفوسیتهای تیِ بالغ و آموزش دیده ، می باشد. این آموزش بدین گونه است که سلولهای تیِ نابالغ پس از تولید در مغز استخوان ، بلافاصله جهت طی یک دوره آموزش و اصطلاحاً "دوره تَمایز" به غدّه تیموس ارسال می شوند. غدّه تیموس تمامی انواع و اشکال سلولهای خودی را در خود ذخیره دارد. بنابراین در این دوره کوتاه مدت [1] ، لنفوسیتهای نابالغ آزمونی را می گذرانند که بیاموزند که نباید به این سلولهای خودی برخورد نموده و نباید به آنها بچسبند. زیرا در صورت برخورد ، حذف شده و به بدن راه پیدا نمی کنند. فقط آن دسته از لنفوسیتهایی که با هیچ یک از سلولهای خودی برخورد ننموده باشند اصطلاحاً بالغ شده و به بدن ترشح می شوند. پس از ترشّح شدن از غدّه نیز تعدادی از آنها ممکن است به شدّت به آنتی ژنهای خودی بچسبند که این موجب مرگ کنترل شده [2] آنها می

[1] Antigen tolerization
[2] Apoptosis

شود و یا اگر به آنتی ژن خودی نچسبند [1] به این اصطلاحاً حد آستانه تحمل مصونیّت[2] می گویند. هر چه این حدآستانه تحمل بالا باشد بهتر است.



شکل ۱۶ – نگاهی جامع بر فرایند انتخاب تکثیری و چگونگیِ بدست آمدنِ حدّ آستانه تحمّلِ مصونیّت یک آنتی ژن [62]

شکل بالا برگرفته از [62] نحوه ی بدست آمدنِ این حدآستانه تحمّل را برای مصونیّت به خوبی به تصویر کشیده است. مهم ترین جنبه ی فرایند انتخاب تکثیری در دستیابی به مصونیّتِ اکتسابی آنست که سیستم را سریعاً به تعادل می رساند. تئوری انتخاب تکثیری در بخش بعد بطور مفصّل تشریح شده است.

اصولاً فرایند های انتخاب منفی و تکثیری مکمّل یکدیگرند و به طور خلاصه می توان گفت که هر آنتی بادی به هر میزان که توانایی شناسایی آنتی ژنهای غیر خودی را کسب نموده و به آنتی ژنهای خودی واکنش نشان ندهد به همان میزان تکثیر می یابد. در حالیکه آنتی بادیهایی که با خودی ها می چسبند و به مرگ دچار می شوند تعادل سیستم را با اخلال مواجه می کنند. در نتیجه با فرایند انتخاب تکثیری همروند با انتخاب منفی، مشاهده شده که از تعداد آنتی بادیهایی که قابلیّت شناسایی آنتی ژنهای غیر خودی را دارند افزوده شده و از تعداد آنتی بادیهای ناموثّر کاسته شده است و این یعنی تعادل. در واقع وجود فرایند انتخاب تکثیری و پتانسیل ذاتی آن در برقراری نوعی میزان سازی خود، عاملی برای تعادل فراهم می کند.

[1] Binding
[2] Immunological tolerance

مطابق شکل ، آنتی بادیهای مترشّح از لنفوسیتهای بیِ ۱ و ۲ و ۶ و ۷ و ۸ به دلیل آنکه الگوهای خودی را بایند نموده اند از تکثیر آنها جلوگیری شده و به مرور زمان از تعداد آنها کاسته می شود. در حالیکه آنتی بادیهای مربوط به لنفوسیتهای ۳ و ۴ و ۵ با کوچکترین تحریکی از سوی آنتی ژنهای غیر خودی ، با هر واکنشی سریعاً تکثیر می شوند.

هدف ، مقاومت در برابر سلولهای خودی است. لنفوسیتهای تی بالغ باید بتوانند پتانسیل بالقوّه ی خود را برای تشخیص آنتی ژنهای ناشناخته که هیچ عکس العملی با سلولهای خودی ندارند را افزایش داده و بر عکس در برابر خودی ها صبور باشند و با آنها برخورد نکنند. [۲۰]

از طرفی اگر برای مدت زمانی معیّن ، یک لنفوسیت هیچ برخورد موفّقی [1] را انجام نداده باشد سنّ آن بالا رفته و می میرد. به این ترتیب لنفوسیت های جدید با تغییرات ژنتیکی (جَهش [2]) و اعمال اصلاحات تصادفی فرمتِ پپتیدهای بر روی سطح آنها توسط غدّه تیموس ، به منظور جایگزینی با لنفوسیت های مُرده باز تولید می شوند. همچنین به سلولهایی که با موفقیت در تیموس بالغ و اصلاح شده اند و پس از ترشح در بدن نیز آنتی ژنهای غیر خودی و سلولهای آلوده به ویروس را شناسایی می کنند سلولهای تهاجمی [3] نیز می گویند.

لَنفوسیتی که با آنتی ژنها برخورد موفّق نموده باشد تبدیل به یک لنفوسیت حافظه می شود که بعداً تحت فرایند انتخاب تکثیری قرار میگیرند و با عمل تکثیر در بدن برای ایجاد پاسخ های سریع استفاده می شوند. در نتیجه موفّقیتِ انتخاب منفی بستگی زیادی به موفقیت نرخِ تولید تش.د.بغ.م. ها دارد که ضمن زنده ماندن پس از بلوغ، قادر به تشخیص غیر خودیها نیز باشند. [52]

۱۳-۱-۱- الگوریتم انتخاب منفی NSA

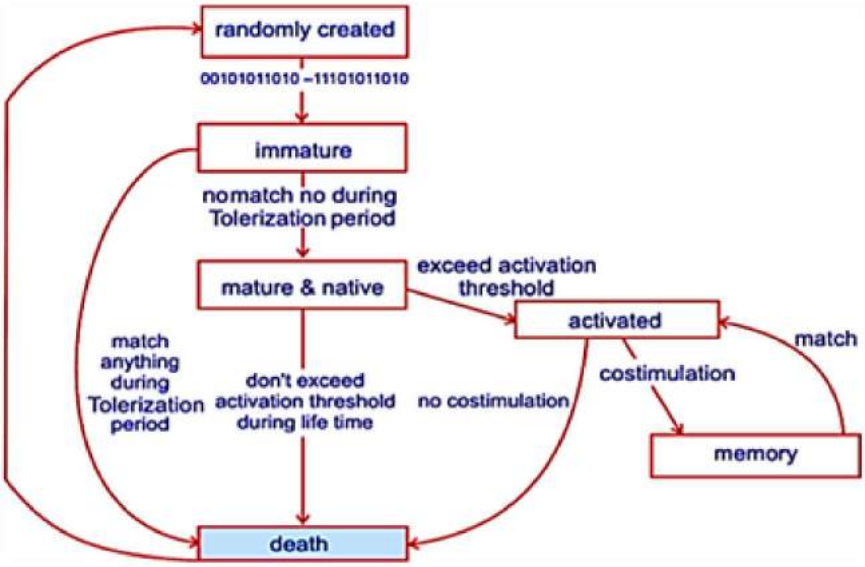
به منظور تقلید از مفهوم انتخاب منفی و استفاده از مزیتهای آن جهت کاربرد در سیستم تشخیص نفوذ ، چارچوبی برای الگوریتم انتخاب منفی برای اولین بار توسط فورِست و همکاران در سال ۱۹۹۴ پیشنهاد گردید. [۲۱] هر سلولِ تی در دیواره سلّولیِ خود ویژگیهای خاصی دارد که اجازه می دهد با آنتی ژنی خاص برخورد کند و فعّال شود. این فعّال سازی همانطور که در شکل ۱۳ در [پ-الف-۱۰] (کلیپ) مشاهده گردید باکمک بیگانه خواران صورت میگیرد. [۱۸]

فورِست و همکاران رشته های باینری را به عنوان تش.د. ، تولید و به آنتی ژنها و لنفوسیتها نسبت دادند، سپس متدی به نام $r - continues\ bit$ را به عنوان حد آستانه جهت محاسبه میزان تطبیق و سازگاری دو رشته ارائه کردند که در تشخیص آنومالی به کار رفت. [۸] در واقع فاز تشخیص بدین صورت بود که یک رشته

[1] Match
[2] Mutation
[3] Invasion T cells

تش.د. با یک رشته آنتی ژن تنها در صورتی می توانند تطبیق بخورند (match / binding) که هر دو رشته در حداقل کاراکترهایی یکسان در امتداد بی وقفه ای از r بیت مشترک باشند.

شکل زیر فرایند انتخاب منفی را نشان می دهد که این دانشمندان جهت پیشنهادِ الگوریتمِ NS ارائه دادند. [۵۰] از عمده کاربردهای انتخاب منفی در تشخیص ویروس ها و بد افزارهایی است که در سیستم اقدام به دستکاری داده ها می نمایند. بنابراین می توانند در HIDS ها [۱] استفاده شوند.



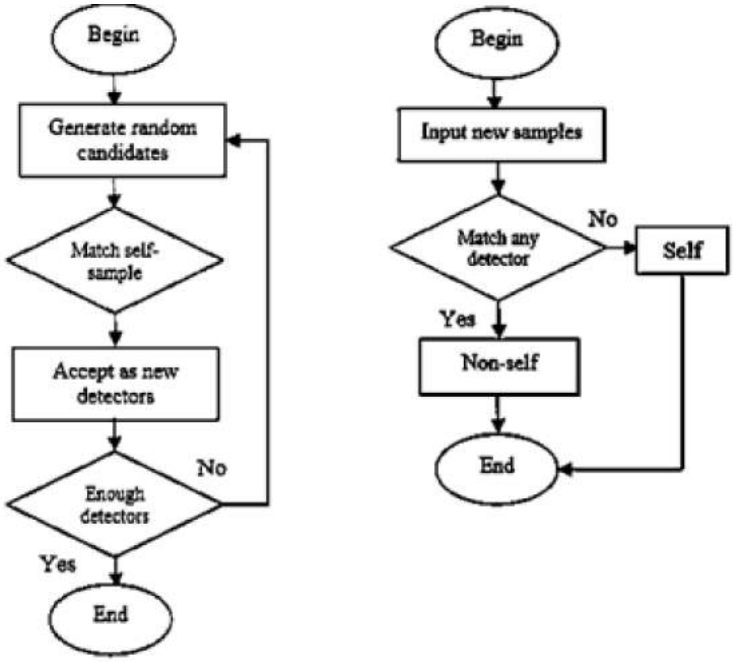
شکل ۱۷ – الگوریتم انتخابِ منفیِ پیشنهاد شده از سوی فورست و همکارانش برای فرایند انتخاب منفی

سود و کدِ NSA در [۶۰] ارائه شده است.

نقطه آغاز این الگوریتم ، تولید یک سری از رشته های خودی (Self Samples) می باشد که وضعیّتِ نرمالِ سیستم را توصیف می کنند. سپس یک سری از تش.د. های بالقوّه (P) نیز تولید شده و با هر یک از اعضای مجموعه ی این رشته ها (الگو های خودی Self) چک می شوند. اگر حداقل یکی از رشته های الگوی خودی با یک تش.د. عضوِ مجموعه ی P از حد آستانه تعریف شده فراتر رفته و در نتیجه این وابستگی به هم تطبیق بخورند حذف می شود ولی در غیر اینصورت به بلوغ رسیده و به مجموعه تش.د. های نهایی بالغِ D اضافه می شود. چَرخه ی تولید تش.د. ها در گِراف زیر بیان شده است.

[۱] سیستم های تشخیص نفوذِ مبتنی بر میزبان

از جمله مسائل مهم این الگوریتم زمان بالای محاسباتی برای مقایسه تش.د. های خودی از غیر خودی[1] و همچنین تولید تش.د. های بغ.م است که بتوانند از آن حد آستانه لازم ($r - continues - bit$) عبور کنند.



شکل ۱۸ – فرایند تولید تش.د. ها (سمت چپ) و فرایند مانیتورینگ (سمت راست)

۱۳-۲- مکانیسم انتخاب تکثیری [2]

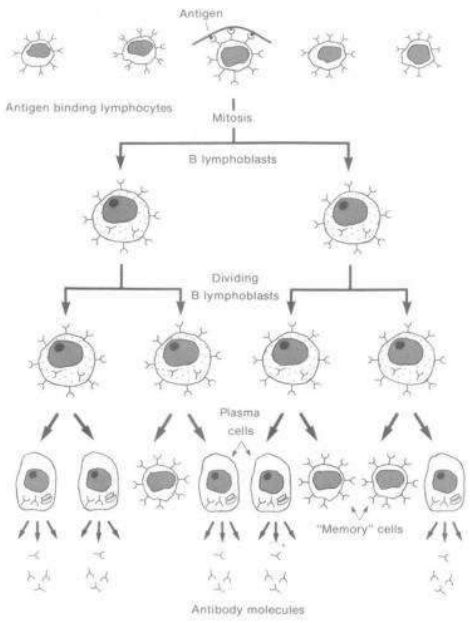
مطابق تئوری انتخاب تکثیری که توسط Burnet در سال ۱۹۹۵ پیشنهاد گردید سیستم ایمنی با فرایند انتخابی روبه روست که نتیجه ی آن ، اِنقیاد [3] با آنتی ژن ها و فعالسازی لنفوسیتهاست. مطابق مطالب بخش قبل شناسایی یک آنتی ژن از سوی لنفوسیتها به نوعی سیگنال آماده باش برای آنهاست که منجر به فعال سازی سلّولهای B و T می شود و سایر لنفوسیتهای T کُشنده و B حافظه را نیز فعّال می کند. زمانی که لنفوسیتهای B حافظه جهت برخورد با آنتی ژن آماده می شوند برای هر سلّولِ بی احتمال این تطبیقِ الگو مقدار مشخصی است که به آن درجه ی وابستگی یا اِنقیادِ وابستگی می گویند. [۶۰]

همانطور که قبلاً نیز بیان گردید هر لنفوسیت حافظه از نوعِ بی بر روی سطح خود تنها از یک نوع الگوی آنتی بادی (نوع شاخک) و به تعداد زیاد از آن الگو دارد که همگی برای تطبیق خوردن با یک نوع آژ.غ.خد.

[1] Affinity time
[2] Clonal Selection
[3] Bind

خاص ساخته شده اند. بنابراین برای لنفوسیتها به عنوان عاملان کُنش گَرِ دفاعیِ تطبیقی پذیرِ بدن بسیار مهم است که هم سریعتر و هم به دقّت تمامی این آنتی ژنها را بایند کرده و نابود سازند. در بخشِ مربوط به تشریحِ انتخاب منفی نیز اشاره گردید که در نگاشت این مفهوم به امنیت شبکه متد r – continues bit تعریف شده که میزان وابستگی و یکسان بودنِ r بیتِ متوالی را در هر رشته آنتی بادی و رشته آنتی ژن نشان می دهد (البته این مفهومِ تطبیق الگو در هر دو نوع لنفوسیتِ تی و بی در برابر آنتی ژنها وجود دارد). اگر در تطبیقِ الگو این میزانِ وابستگی بین دو رشته در حداقل r بیتِ متوالی به اندازه ی کافی بالا باشد بهتر بوده و در نتیجه تکثیرِ آنتی بادی با الگوی مربوطه نیز بالاتر خواهد بود.

اصولاً هدف فرایند انتخاب تکثیری، تکثیر آن دسته از لنفوسیتهای بیِ فعّال شده ای است که میزان وابستگی شان به یک آنتی ژن غیر خودیِ خاص از بقیه ی لَنفوسیتهای موجود به مراتب بیشتر باشد. (شکل ۱۹ و مرحله ۴ در شکل ۲۰) کلاً حضور یک آنتی ژن در مقابل یک لنفوسیت پدیده ای تصادفی است. پس میزان وابستگی نیز پدیده ای تصادفی خواهد بود. قاعدتاً هر چه این میزان وابستگیِ آنتی ژنیک (قدرت چسبندگی یا تطبیق الگو) در آنتی بادیها نسبت به یک آنتی ژنِ غیر خودی خاص بیشتر باشد عمل تکثیر چنین آنتی بادی هایی می تواند عملکردِ تشخیص را بهبود ببخشد. این فرایند شبیه یک فرایندِ دسته بندی به نظر می آید زیرا تمام غیر خودی ها از خودی ها تفکیک می شوند.



شکل ۱۹ - نگاهی ساده به فرایند انتخاب تکثیری ، برگرفته از [62]

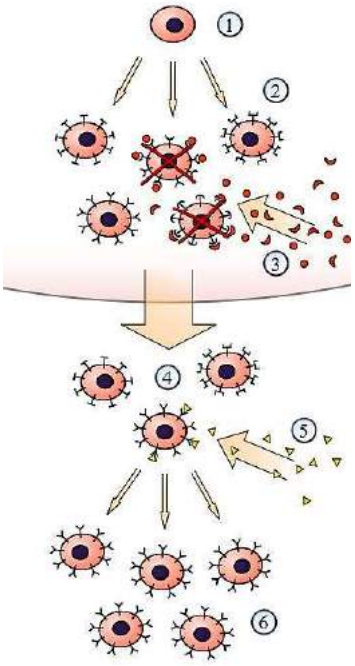
روابطِ ساده زیر که برگرفته از گزارش فنّیِ [62] می باشند تعداد کلان (تکثیر) هایی که برای هر آنتی بادی ایجاد می شوند را در الگوریتمِ CLONALG نشان می دهد. این الگوریتم بهبود یافته ی الگوریتم CSA می باشد. تمام پیاده سازی های مربوط به الگوریتم های انتخاب تکثیری فوق و همچنین الگوریتم های انتخاب منفیِ ARIS Series , Immunos Series در این گزارش فنّی در نسخه ی ۳,۶ نرم افزار داده کاویِ Weka ارائه شده اند.

$$\text{NumClones} = \left\lfloor \frac{\beta.N}{i} \right\rfloor + 0.5 \qquad (۱-۱۳)$$

$$\text{Nc} = \sum_{i=1}^{n} \left\lfloor \frac{\beta.N}{i} \right\rfloor + 0.5 \qquad (۲-۱۳)$$

که در آن β فاکتورِ تکثیر بوده و N نیز اندازه ی استخر آنتی بادی و i نیز اولویت جاری آنتی بادی در بین ترتیبی از رشته ها می باشد بطوریکه : $i \in [1, N]$

مجموع تعداد تکثیرها متناسب با هر آنتی ژن می باشد بدین ترتیب رابطه ی دوّم مجموع تکثیرها را برای تمام آنتی ژنها محاسبه می نماید.



شکل ۲۰ – فرایند انتخاب تکثیری

۱۳-۲-۱- پدیده بلوغ وابستگی [1]

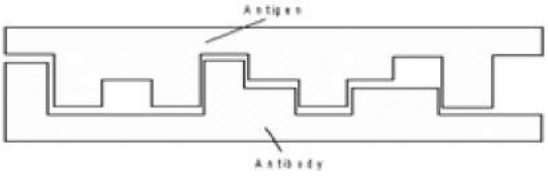
با فرایند تکثیرِ سلّولهای B که بالاترین حد وابستگی به آژ.غ.خد. را دارند (شکل ۲۰ – مرحله ۵) ، عملاً میانگین درجه وابستگی کلّ سیستم دفاعی در مجموع افزایش می یابد. زیرا به محض تطبیق آنتی بادی با آژ.غ.خد. (زرد رنگ)، این نوع لنفوسیت با همان نوع آنتی بادی بر روی سطحش تکثیر پیدا می کند (مرحله ۶) به این پدیده " بُلوغ وابستگی " گویند. به عبارت دیگر ، در طول این فرایند میزان وابستگی کل لنفوسیت های B عملاً با تکثیر آنها افزایش می یابد.

در طول این فرایند همینطور لنفوسیت هایی نیز هستند که به دلیل عدم تطبیق با هیچ کدام از آنتی ژنها می میرند (مرحله ۴ مواردیکه تطبیق نخورده اند) و همچنین لنفوسیتهایی نیز هستند که به دلیل تطبیق با آنتی ژنهای

[1] Affinity maturation

خودی بدن ازبین می روند. پس دو نوعِ مرگ داریم. [۶۰] (مرحله ۳)

۱- ابتدا مجموعه ای از الگوها (آنتی ژنها) شناسایی می شوند (مجموعه ی S). در این الگوها ممکن است خودی و غیر خودی نیز حضور داشته باشند. (عوامل قرمز و زرد رنگ شکل ۲۰)



شکل ۲۱ - درجه ی اِنقیاد (وابستگی)

۲- مجموعه ای تصادفی از لنفوسیتها با آنتی بادیهای متفاوت با الگوی تصادفی تولید می شوند. (مجموعه A) مرحله ۲ شکل فوق.

۳- حَلقه : برای تمام الگوها S_i ، درجه وابستگی آن (آمار تطبیق الگو با آزمون و خطا) با هر عضو از A تعیین می شوند.

۴- زیر مجموعه ای از A با حداکثر درجه وابستگی مشخص شده و هر یک به تناسب درجه وابستگی ، تکثیر می یابند. (مرحله ۵ و ۶) بنابراین تعدادِ Clone ها برای هر آنتی بادی به درجه وابستگی آن با عضوِ S_i وابسته است.

برای مثال در مرحله ی ۶ در شکل فوق لنفوسیتِ بالغ به ۵ لنفوسیت مشابه تولید مثل شده و این میزان تکثیر به علت درجه ی وابستگیِ آنتی بادیها (شاخکهای بر روی لنفوسیت) و دانه های زرد می باشند.

۵- A با حاصل این تکثیر تغییر نموده و یک کپی از لنفوسیتهای دارای پُتانسیلِ تطبیقِ الگو (لنفوسیتهای بیِ دارای آنتی بادیهای با سر دو شاخه V در شکل فوق) به عنوان تش.د. های دارای پتانسیل بالقّوه به مجموعه M منتقل می شوند.

۶- لنفوسیتهای دارای آنتی بادیهایی با کمترین حد وابستگی (آنتی بادیهای دارای سر مستطیلی و سطح صاف در شکل فوق) نسبت به اعضای S نیز از مجموعه ی A حذف شده و به جای آنها لنفوسیتهایی با آنتی بادیهای تصادفی تولید و در مجموعه A جایگزین می شوند. در انتخاب تکثیری ، دو ویژگی مهمِ فرایند بلوغ وابستگی در سلولهای B که در کتابِ تیمیس و دیکاسترو [1] نیز بدان اشاره شده ، به شرح زیر می باشد که می تواند از جنبه محاسباتی مورد بهره برداریِ تحقیقاتی مدّ نظر قرار گیرد.

[1] Timmis and de Castro

- تکثیرِ سلّول های B متناسب با وابستگی به آنتی ژنی که با آن بایند شده اند صورت میگیرد. بنابراین تطبیقِ الگوی بیشتر ، تکثیرِ بیشتری را در پی دارد.
- جهش هایی که آنتی بادی یک سلّولِ B متحمّل آن شده است نسبتِ عکس با درجه ی وابستگی با آنتی ژنی دارد که لنفوسیت بدان چسبیده است. بدان معنی که یک عمل جهش ژنتیکی از نظر الگو و شکل شاخکها در آن دسته از آنتی بادیها اتفاق می افتد که وابستگی شان کمتر بوده و نیازمند تغییر الگو می باشند.

استفاده از این دو ویژگی مهم باعث شد تا دِکاسترو و وُن زوبِن [1] الگوریتمِ Cloning را پیشنهاد دهند. ویژگیهای مهم انتخاب تکثیری عبارتند از [۵۰] :

حذف تکثیرهای خود واکنشی (مرحله ۳ شکل فوق) ، تکثیر و تمایز میان لنفوسیتهای بالغ ، محدودیت یک الگو به یک سلّولِ متمایز و حفظ این الگو بوسیله اولاد تکثیری (مرحله ۶) و در نهایت گسترش تغییرات ژنتیکی در الگوی آنتی بادیهای گوناگون بواسطه شکلی از بلوغ. سودو کدِ الگوریتم انتخاب تکثیری در [61] ارائه شده است.

۱۳-۳- تئوری خطر DT

رشته ایمنی شناسی به تدریج از سال ۱۹۸۵ تاکنون رُشد پیدا کرده است. یک اصل مهم ایمنی شناسی این است که سیستم ایمنی به حضورِ موجودیتهای خارجی (غیر خودی) پاسخ داده و به میزبان (خودی) واکنشِ خاصی را نشان ندهد. (در بخش قبل از این رفتار تحت عنوان حد آستانه قابل تحمل مصونیّت اشاره شده است)

در سال ۱۹۹۴ مت زینگِر [2] تئوریِ خطر را پیشنهاد کرد که در میان ایمنی شناسان به شدّت مشهور شد. [۴۵] تئوری خطر در سالهای اخیر به رشته ایمنی شناسی افزوده شده و از این نظر نسبتاً جدید است. این تئوری می گوید ، طی رخدادی در بدن مثل ایجاد زخم در یک بافت سلّولی APC ها به محض مشاهده آنتی ژن ها ، بوسیله آلارمهایی فعّال شده و با سیگنالهایی محرّک [3] لنفوسیتهای T کمک رِسان را نیز فعّال می کنند که متعاقباً پاسخهای ایمنی متناسب - واکنش های تطبیقیِ هیومورال و ایمنیِ سلولی را ترتیب دهند.

وقتی بافت بدن زخمی می شود سلّولهای آسیب دیده بواسطه آنکه مُرده اند توانایی انتشار سیگنالهایی را در بدن دارند ، این سیگنالها توسط سلّولهای ایمنیِ ذاتیِ مشخص شناسایی می شوند که سلولهای دندریت [4] نام

[1] de Castro and Von Zuben
[2] Matzinger
[3] Co – Stimulatory
[4] Dendritic cells

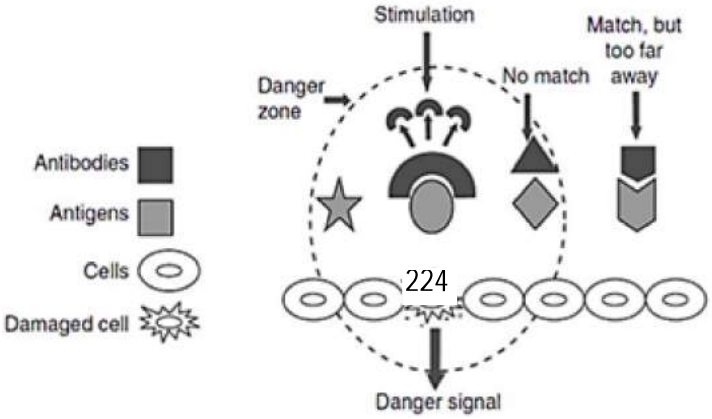
دارند که به نظر می رسد این سلّولها متناسب با نوعِ سیگنال سه مُدِ " بالغ" ، " نارس"، " نیمه بالغ " به خود می گیرند. (شکل ۲۲)

سلّولِ دِندریت در مُدِ نارَس ، آنتی ژنهایی را بهمراه سیگنالهای خطر و امن از محیط اطراف خود دریافت می کند. مثل سیگنالهای از نوع PAMP که کار شناسایی الگوی حملات را انجام می دهند و سیگنالهای منتشر شده و ارسال شده از سیتوکینهای[1] فَساد انگیز در بدن. سلّولِ دِندریت قادر است تا با استفاده از این سیگنالها جمع بندی کند که آیا محیط امن است یا نا امن.

اگر سیگنال امن باشد ، سلول دندریت به مُدِ نیمه بالغ در آمده و به محض ظهور آنتی ژن در برابر سلولهای تی، دِندریت مسبّبِ ایجادِ قابلیتِ حدّ تحمّل برای سلولهای تی می شود یعنی به آنها آماده باش لازم را می دهد و می گوید که صبر کنند و منتظر ظهور آنتی ژن باشند. (وضعیت زرد) امّا اگر دِندریت سیگنال را به عنوان خطر ارزیابی کند سلّول به مُدِ بالغ در آمده و موجب می شود تا سلّولِ تی نیز در برابر آنتی ژن واکنش انفعالی به خود بگیرد. (وضعیت قرمز) [۶۰]

تئوری خطر چگونگی اندازه گیری وضعیت امن یا خطر در سلّولها را در سیستم ایمنی پردازش و مدیریت می نماید. بنابراین دانشمندان ایمنی شناسی با همکاری دانشمندان علوم کامپیوتر در جستجوی راهی برای مُدلسازی ساختمانِ خطر هستند که بتواند در بهبود AIS استفاده گردد.

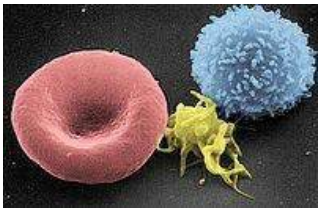
این برای بهبود سیستم های تشخیص آنومالی در شبکه نیز بسیار مفید می باشد. دو الگوریتم شناخته شده به نام های TLR [2] و دیگری الگوریتم DC بر مبنای تئوری خطر توسعه پیدا کرده اند. [۵۰]



شکل ۲۲ - تئوری خطر

[1] Messenger proteins
[2] Tolk– like receptor

۱۳-۳-۱- الگوریتم DC : گرین اسمیت [1] و همکاران ، DCA را بر مبنای تئوری خطر ارائه نمودند. جدول زیر نگاشت مفاهیمِ تئوری خطر را به حوزه محاسبات و علوم کامپیوتر نشان می دهد.



DCA توانایی ترکیب سیگنالهای متعدد را برای تشخیص (ارزیابی) وضعیت جاری سیستم دارد. همبستگی بین سابقه محیط [2] و آنتی ژن به عنوان مبنایی برای تشخیص آنومالی در این الگوریتم استفاده می شود.

شکل ۲۳ - سلول دندریت با زرد رنگ در میکروسکوپ الکترونی

پس می توان تئوری خطر را نوعی اتخاذ استراتژی پیشگیرانه در چَرخه ی دفاع نامید زیرا این تئوری بیانگر اقدامات پیش بینی خطر بوسیله سیگنالهای رخدادها و تعیین وضعیت جاری امن یا ناامن در سیستم می باشد نه صرفاً تشخیص خودی از غیر خودی.

جدول ۱۰ – مفاهیم تئوری خطر

سیگنال	وضعیت دندریت	مفهوم	تشابه در بیولوژیک	از دیدگاه محاسباتی
PAMP	نارس (قبل از دریافت سیگنال)	حضور آنومالی	شاخص حضور میکروبها	پیام خطا در هر ثانیه
سیگنال خطر **DS**	بالغ (وضعیت قرمز)	امکان یا عدم وجود آنومالی	شاخص خسارت بافتهای سلولی	بسته شبکه در هر ثانیه
سیگنال امن **SS**	نیمه بالغ (وضعیت زرد)	عدم وجود آنومالی	شاخص بافت سلولی سالم	اندازه بسته های شبکه

ضمن اینکه اشاره گردید که ترکیبِ سوابقِ سیگنالهای خطرِ ارسالی از قسمتهای مختلف بافت های سلّولیِ بدن، نتیجه گیری نهایی برای تعیین وضعیت جاری سیستم و تشخیص آنومالی را منجر خواهد شد که مصداق کاملِ پیشگیرانه بودن است. سود و کدِ DCA در [61] ارائه شده است.

۱۳-۴- تئوری شبکه ایمنی مصنوعی

[1] Green smith
[2] Context

در سال ۱۹۷۴ ، نیل کی جِرنه [1] تئوری شبکه ی ایمن را برای بهبودِ ویژگیهای سیستمِ ایمنی ارائه کرد. فرضیه وی این بود که سیستم ایمنی به شکل شبکه ای از آنتی بادیها و ضد آنتی بادیهاست که idiotypic network نامیده می شود که در این شبکه همدیگر را شناسایی می کنند. [۵۰]

تئوری پیشنهادی وی این بود که هر گیرنده ی لنفوسیتِ مناسب (آنتی بادی / Receptor) ، باید طی پروسه آزمون و خطا از زیر مجموعه ی شایسته ترین گیرندگان انتخاب گردد. وِی نتیجه گرفت که سیستم ایمنی در حضور آنتی ژن های غیر خودی باید رفتار یا کُنش مناسب را ابتدا طی فعل و انفعالاتی با خودش امتحان نموده و پس از این فعل و انفعالات رفتار ایمنی مناسب شناسایی شود مانند حد آستانه قابل چشم پوشی دامنه تغییرات [2] در تطبیق الگوها و اینکه حافظه چه قدر باشد.

۱۳-۴-۱- الگوریتم شبکه ایمنی مصنوعی (IN)

کار این الگوریتم این است که ابتدا با تولید یک سری الگوها و آنتی بادیها به شکل تصادفی دو مجموعه می سازد. سپس ، پارامترهایی را به عنوان ورودی الگوریتم به منظورِ حدود آستانه لازم برای تکثیر تعیین می کند. سپس ابتدا به بررسی درجه وابسنگی مجموعه آنتی بادیهای تصادفی با الگوهای تصادفی خودی S_i پرداخته و از بین آنها آنتی بادیهایی (N_i) را شناسایی و به زیر مجموعه ی مستقل A هدایت می کند. مجموعه A در این مرحله حاوی آنتی بادیهایی است که بالاترین حد وابستگی را با الگوهای مجموعه S دارند.

در مرحله بعد هر عضوِ A متناسب با حد وابستگی اش با الگوها ، تکثیر شده و به این ترتیب تعداد زیادی فرزند تولید می شوند که از این میان تعداد h فرزند انتخاب شده و به مجموعه تکثیرهای حافظه C انتقال می یابند. مجموعه C حافظه دارای h تا از فرزندانی است که دارای بالاترین حد وابستگی با الگوها می باشند. در مرحله بعد از میان اعضای مجموعه C آن دسته از آنتی بادیهای تکثیر یافته که میزان وابستگی شان با آنتی ژن کمتر از حد آستانه لازم تعیین شده ct هست را حذف می کند. به این ترتیب مجموعه C تعدیل می شود.

سپس تمام مابقی تکثیرهای موجود در C که دو مرحله آزمون را تاکنون طی کرده اند با موفقیت به مجموعه تش.د. های N انتقال می یابند. اعضای N قادر هستند تمام الگوهای ناشناخته را شناسایی کنند. البته این پتانسیل بالقوّه ی آنهاست و در عمل بالفعل سازی آن کار مشکلی می باشد. این چرخه ادامه می یابد تا زمانی که تعداد کافی در N تولید شوند.

به عنوان فاز پایانی این الگوریتم از بین اعضای مجموعه N آن دسته از آنتی بادیهایی که درجه وابستگی شان از حد آستانه لازم تعیین شده برای شبکه (nt) کمتر باشد حذف می شوند و به جای آنها آنتی بادیهای

[1] Niels K.Jerne
[2] Tolerance

تصادفی به تعداد a تا تولید و جایگزین می شوند. این تعداد مجددا وارد پروسه غَربال گریِ الگوریتم فوق می شوند تا تعدیل شده و مجدّدا وارد مجموعه N شوند. [61] سو دو کدِ IN درهمین مرجع ارائه شده است.

این الگوریتم بسیار پویاست بطوریکه مشاهده می کنید که در طی پروسه ای طولانی ابتدا آنتی بادی هایی که سیستم ایمنی به عنوان گیرنده تولید می کند آنها را با الگوهایی تصادفی تولید شده خودش از حیث حد وابستگی می سنجد سپس تعداد مشخصی از بالاترین این میزان را از بین آنها انتخاب و در مراحل بعد با آنتی ژنها می سنجد و بعد از این تعداد مجددا تعدادی که دارای کمترین حد وابستگی از حد آستانه مشخص شده هستند حذف و تعدیل شده و بدین ترتیب وارد مجموعه نهایی تش.د. های واجد شرایط جهت تشخیص و دسته بندی الگوهای ناشناخته می شوند.

به عبارت دیگر این الگوریتم علیرغم پیچیدگی بالایی که نسبت به الگوریتم های نسلِ اولِ سیستم ایمنیِ مصنوعیِ (CS و NS) دارد توانسته است تا حدّ بالایی از دو نیازمندی حافظه پشتیبان (بافر) و حدّ تحمّلِ مورد نیاز را برای سیستم ایمنی جهت شناسایی آنتی ژنها و آنتی ژنهای ناشناخته را ارائه دهد. بطوریکه در صورت پیاده سازی صحیحِ این الگویتم به نظر نمی رسد هیچ موقع سیستم ایمنی با مشکل کمبود تش.د. های کافی مواجه شود ضمن اینکه این میزان تش.د. علاوه بر دارا بودن بالاترین حد وابستگی جهت تشخیص آنتی ژن های غیر خودی ، مُوازنه [1] آن نیز در تولید آنتی بادی حفظ خواهد شد.

بخش چهاردهُم - تحلیل نظریِ حالات ممکن برچسب زنی خطوط دفاعی

مطابق با آنچه که بیان شد ، خطوط دفاعی اوّل و دوّم هر یک پتانسیلها و چالشهایی دارند. اگر الگوریتمهای خطوط دفاعی در کنار یکدیگر بتوانند سیستمی را تشکیل بدهند مسلّماً نتیجه ی ارزیابی ترافیک مشکوک شبکه در خط دفاعی نخست می بایست در نتیجه ی عملکرد خط دفاعی دوّم تاثیر گذار باشد. به بیان ساده تر از دیدگاه بیولوژیک پیامد همکاری و مشارکت دو زیر سیستم مص.ذت. (سازوکار سلولهای دندریت) و زیر سیستم مص.تط. (مکانیسمِ انتخاب منفی در لنفوسیتها) در تعیین سرنوشت یک آنتی ژن بسیار موثّر و کارآمد می باشد.

تعیین سرنوشت یک آنتی ژن و چیستیِ آن ، یک پروسه ای طولانی است که سیستم ایمنی بدن در واقع با آن زندگی می کند. بنابراین شبیه سازی چنین ساختاری با این پیچیدگی نوعی حیات مصنوعی را بوجود می آورد که منابع پردازشی وسیعی را نیز می طلبد. کاری که ما در این پژوهش انجام دادیم و تا حدودی در حد چند سیکل اوّلیه توانستیم به ارزیابی نحوه حیات مصنوعی یک ترافیک مشکوک شبکه بپردازیم ، اینکه سیستم تشخیص با ترافیک شبکه دقیقاً چه کار می کند؟

[1] Trade – off

در دنیای ایمنی بیولوژیک با رخداد پدیده ی نفوذ ، عامل آنتی ژنیک پس از ورود به بافت های سلولی بدن در صدد آسیب رساندن به اهداف مشخصی در داخل سلولهاست. آثار این تخریب به صورت سیگنال به سلولهای دندریت مجاور آشکار می شوند و این سلولهای دندریت هستند که تصمیم گیرنده نهایی هستند و بسته به سیگنال دریافتی از سلولهای آسیب دیده / سالم ، سیگنال خروجی متناسب و بهنگام را تولید نموده و سریعاً واکنش نشان میدهند (در این حالت ممکن است اشتباهی نیز رخ داده باشد و سلولهای دندریت به دلیل ارزیابی نادرست وضعیت سلول ، آنرا به اشتباه محصور نمایند). سلّولهای دندریت پس از تولید سینگالهای خروجی و قرنطینه نمودن نمونه های مشکوک و مشخص نمودن بافت های سلولی آسیب دیده یا در شرف آسیب ، مص.تط. بدن را نیز فعال نموده و نمونه ای از عامل آنتی ژنیک مشکوک در قرنطینه را به لنفوسیتهای T می سپارند. بدین ترتیب این لنفوسیتها نیز فعّال شده و سلولهای B را نیز فعّال می کنند تا با تولید و تکثیر آنتی بادیها ، تطبیق الگوی آنتی ژن- آنتی بادی متناسبی را از خود نشان دهند.

این همکاری مابین دو مص.ذت. و مص.تط. در بدن ، همان ایده ای است ما از آن الهام گرفتیم و از ابتدای کار پژوهشی در این پایان نامه به دنبال شبیه سازی این ایده بودیم. هدف اصلی ما همانطور که قبلاً نیز گفته شد شبیه سازی این حیات مصنوعی بود. حیاتی که در آن این دو واکنش با هم همکاری لازم را داشته باشند و در شناسایی نفوذ، مکمّل یکدیگر باشند.

سوال - یک ترافیک مشکوک شبکه پس از ورود به سیستم همانند آنتی ژنی که به بافتهای سلّولی بدن نفوذ کرده در نهایت چه سرنوشتی خواهد داشت ؟

پاسخ این پرسش از سه دیدگاه نظری ، تجربی و فلسفی قابل بررسی است. ما در این قسمت با تحلیلی جامع، احتمالات تمامی حالاتهایی (مثبت/منفی صحیح ، مثبت/ منفی کاذب) که ممکن است در تست های خطوط رخ بدهند را به صورت نظری تحلیل و بررسی نموده ایم. نتیجه این تحلیل گام مهمّی است که نشان می دهد ایده پیشنهادی چه پتانسیل هایی دارد و اینکه در واقع مکمّل سازی دو مص.ذت. و مص.تط. در کنار هم در قالب یک سیستم تشخیص نفوذ شبکه تا چه حد موثّر بوده و به بهبود نرخ های عملکرد دسته بندی کمک خواهد نمود. در جدول زیر تمامی حالات ممکن الوقوع در برچسب زنی توسط سیستم پیشنهادی را از دیدگاه تئوری بدقّت بررسی نموده و مزیّت مکمّل بودن این خطوط در کنار یکدیگر (میزان موثّر بودن ایده ی ترکیب دو خط دفاعی) در کاهش نرخ های خطا و بهبود عملکرد کیفیّت دسته بندی را بررسی نموده ایم.

جدول ۱۱ - تحلیل تمامی حالات ممکن الوقوع در برچسب زنی توسط سیستم پیشنهادی

rank	Actual	DCA	Conf	NSA	Conf	Hybrid	Probability of Occurance by experiences of tests	Final Label	Output Situation

1	KN	N	TN-1K	KN	TN-KK	Sure	51.868611	TN-KK	Semi-NSD
2	KN	N	TN-1K	UN	TN-KU		51.868611	TN-KU	Semi-D-TN / Semi-H-TN
3	KN	N	TN-1K	KA	FP-KK	Not-Sure	3.001389	-	Semi-ASD
4	KN	N	TN-1K	UA	FP-KU		3.001389	-	Semi-D-TA / Semi-H-TA / Semi-MD-TA / Semi-MD-TN
5	KN	A	FP-1K	KN	TN-KK		42.661389	-	Mat-NSD
6	KN	A	FP-1K	UN	TN-KU		42.661389	-	Mat-D-TN / Mat-H-TN
7	KN	A	FP-1K	KA	FP-KK	Sure	2.468611	FP-KK	Mat-ASD
8	KN	A	FP-1K	UA	FP-KU		2.468611	FP-KU	Mat-D-TA / Mat-H-TA / Mat-MD-TA / Mat-MD-TN
9	UN	N	TN-1U	KN	TN-UK		51.868611	TN-UK	Semi-NSD
10	UN	N	TN-1U	UN	TN-UU		51.868611	TN-UU	Semi-D-TN / Semi-H-TN
11	UN	N	TN-1U	KA	FP-UK	Not-Sure	3.001389	-	Semi-ASD
12	UN	N	TN-1U	UA	FP-UU		3.001389	-	Semi-D-TA / Semi-H-TA / Semi-MD-TA / Semi-MD-TN
13	UN	A	FP-1U	KN	TN-UK		42.661389	-	Mat-NSD
14	UN	A	FP-1U	UN	TN-UU		42.661389	-	Mat-D-TN / Mat-H-TN
15	UN	A	FP-1U	KA	FP-UK	Sure	2.468611	FP-KU	Mat-ASD
16	UN	A	FP-1U	UA	FP-UU		2.468611	FP-KU	Mat-D-TA / Mat-H-TA / Mat-MD-TA / Mat-MD-TN
17	KA	N	FN-1K	KN	FN-KK		1.355075	FN-KK	Semi-NSD
18	KA	N	FN-1K	UN	FN-KU		1.355075	FN-KU	Semi-D-TN / Semi-H-TN
19	KA	N	FN-1K	KA	TP-KK	Not-Sure	15.394925	-	Semi-ASD
20	KA	N	FN-1K	UA	TP-KU		15.394925	-	Semi-D-TA / Semi-H-TA / Semi-MD-TA / Semi-MD-TN
21	KA	A	TP-1K	KN	FN-KK		6.734925	-	Mat-NSD
22	KA	A	TP-1K	UN	FN-KU		6.734925	-	Mat-D-TN / Mat-H-TN
23	KA	A	TP-1K	KA	TP-KK	Sure	76.515075	TP-KK	Mat-ASD
24	KA	A	TP-1K	UA	TP-KU		76.515075	TP-KU	Mat-D-TA / Mat-H-TA / Mat-MD-TA / Mat-MD-TN
25	UA	N	FN-1U	KN	FN-UK		1.355075	FN-UK	Semi-NSD
26	UA	N	FN-1U	UN	FN-UU		1.355075	FN-UU	Semi-D-TN / Semi-H-TN

27	UA	N	FN-1U	KA	TP-UK	Not-Sure	15.394925	-	Semi-ASD
28	UA	N	FN-1U	UA	TP-UU		15.394925	-	Semi-D-TA / Semi-H-TA / Semi-MD-TA / Semi-MD-TN
29	UA	A	TP-1U	KN	FN-UK		6.734925	-	Mat-NSD
30	UA	A	TP-1U	UN	FN-UU		6.734925	-	Mat-D-TN / Mat-H-TN
31	UA	A	TP-1U	KA	TP-UK	Sure	76.515075	TP-UK	Mat-ASD
32	UA	A	TP-1U	UA	TP-UU		76.515075	TP-UU	Mat-D-TA / Mat-H-TA / Mat-MD-TA / Mat-MD-TN

در جدول بالا نویسه های اختصاری انگلیسی بکار رفته اند. در جدول ۲۰ از [پ–ب–۳] این نویسه های اختصاری را شرح داده ایم. هنگام تست نمونه ها در خط نخست دفاعی (الگوریتم سلولهای دندریت) هشت وضعیت ممکن [1] (از نظر صحّت یا اشتباه بودن برچسب زنی) وجود دارد. این هشت وضعیت در کنار حالات خروجی تست خط دوّم دفاعی مجموعاً شانزده حالت ممکن را تشکیل می دهند که ما این شانزده وضعیّت را وضعیّتهای تصمیم گیری Sure – notSure نامیده ایم . در شکل های ۲۴ و ۲۵ تمام پلاتهای خروجی مربوط به این شانزده وضعیّت ممکنِ تصمیم گیری در دادگان یادگیری به خوبی ترسیم شده اند. نکته ای که در خصوص این پلات ها وجود دارد آنست که ابتدا دادگان یادگیری به دو زیر مجموعه ی ASD و NSD تقسیم شده و سپس تش.د. ها در فضای ابعاد مسئله تولید شدند. نمونه تست فرضی با رنگ قرمز در این پلاتها مشخص شده اند. به محض اینکه موقعیتِ آنتی ژن (نمونه تست قرمز رنگ) در فضای تحت پوشش حداقل یک تش.د.بغ قرار بگیرد ، آن تش.د. (ها) تبدیل به تش.د.بغ.م.حظ می شوند که با رنگ زرد نمایش داده شده اند.

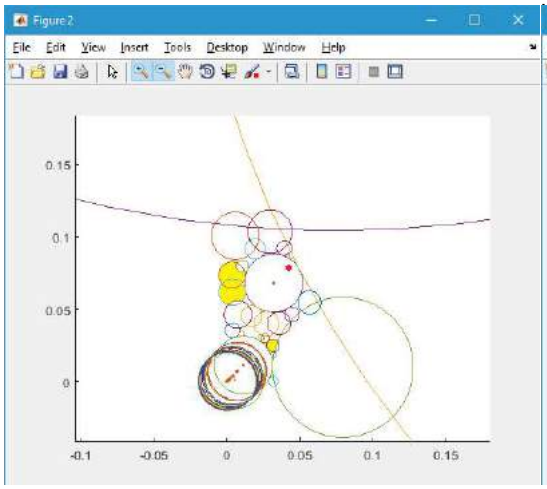
معمولاً برچسب نمونه ها به صورت نرمال یا آنومالی می باشد. اگر برچسب واقعی کلاس نمونه ها را همانند آنچه که در مقاله [50] اشاره شده به صورت چهار کلاس نرمال شناخته شده / نرمال ناشناخته (جدید) / آنومالی شناخته شده (حمله امضاء شده) / آنومالی ناشناخته (حمله جدید) در نظر بگیریم بدین ترتیب وضعیت های ممکن به تفکیک به صورت جدول زیر خواهند بود.

جدول ۱۲ – تمام حالات ممکن برچسب زنی خطوط دفاعی (به تفکیک)

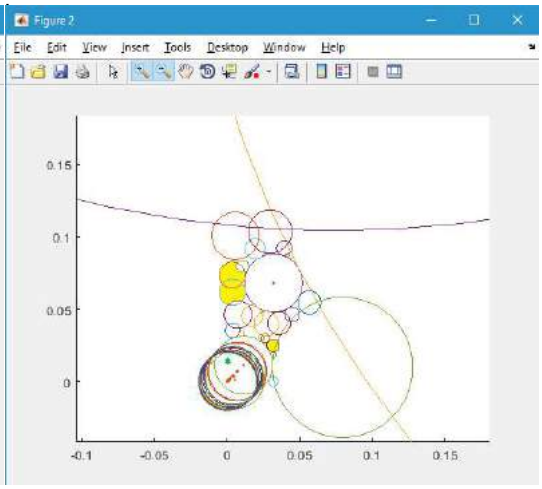
Actual Class Label	DCA's assigned Class Label	NSA's assigned Class Label
KN	N	KN
KN	A	UN
UN	N	KN

[1] $TN-1K/FP-1K/TN-1U/FP-1U/FN-1K/TP-1K/FN-1U/TP-1U$

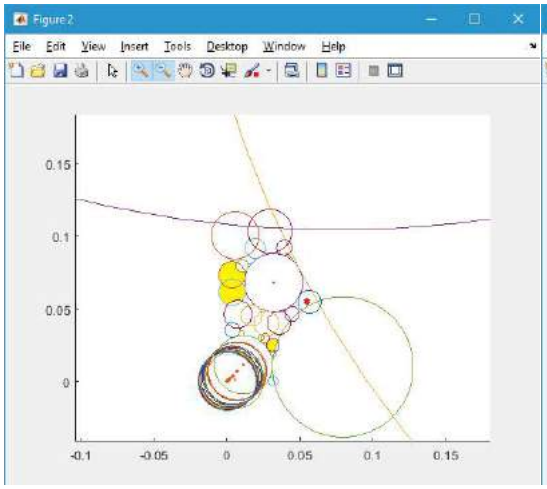
UN	A	UN
KA	N	KN
KA	A	UN
UA	N	KN
UA	A	UN
KN	-	KA
KN		UA
UN		KA
UN		UA
KA		KA
KA		UA
UA		KA
UA		UA



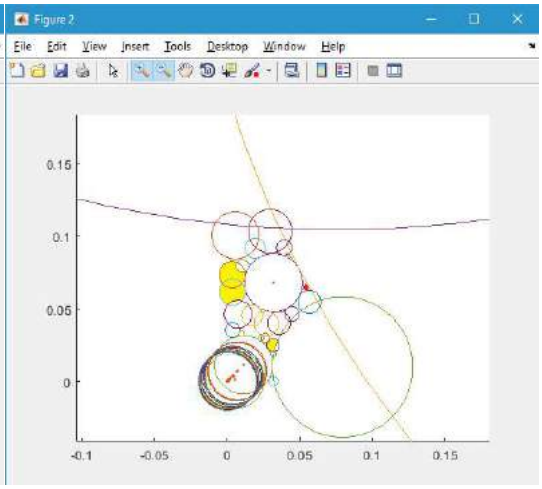
Mat - ASD



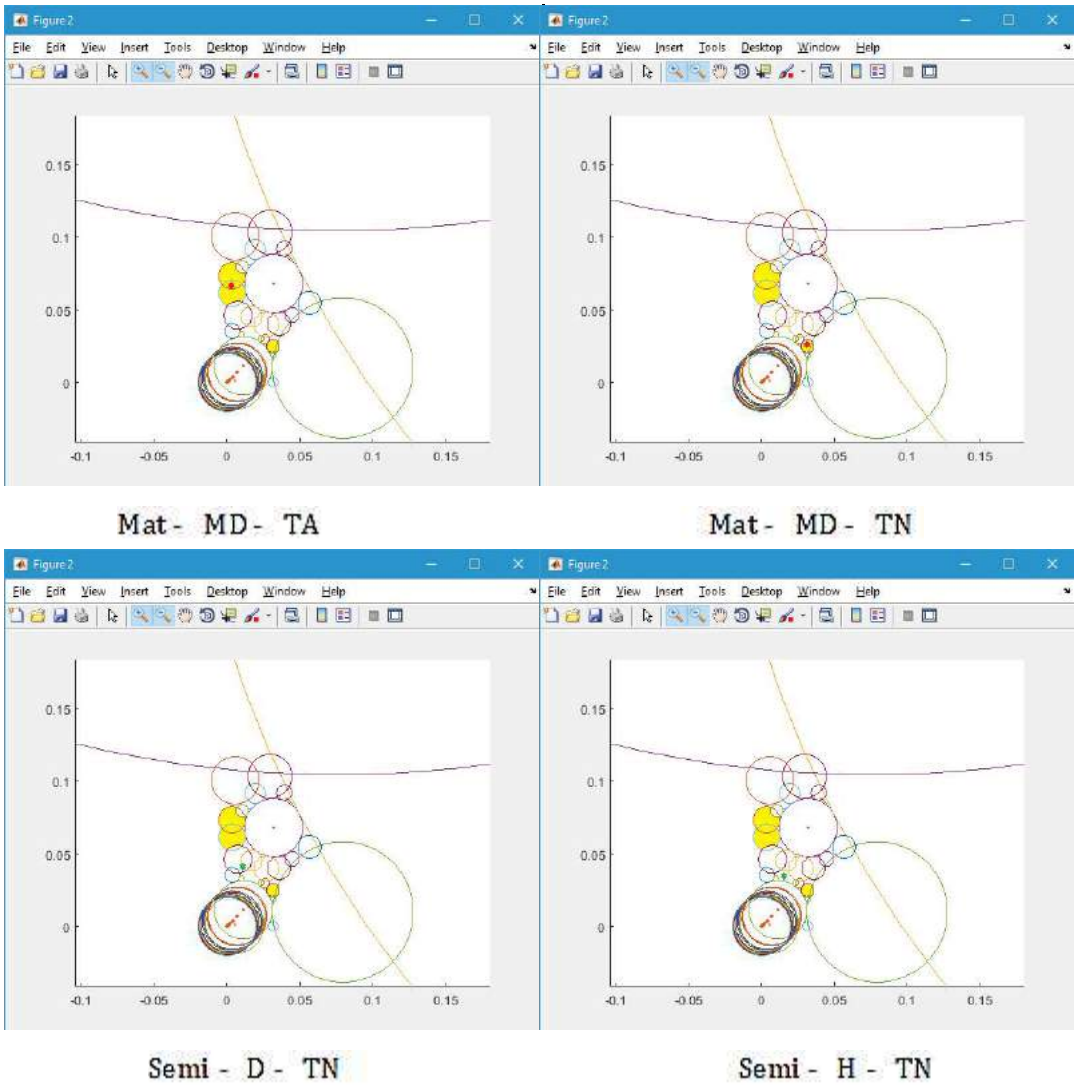
Semi - NSD



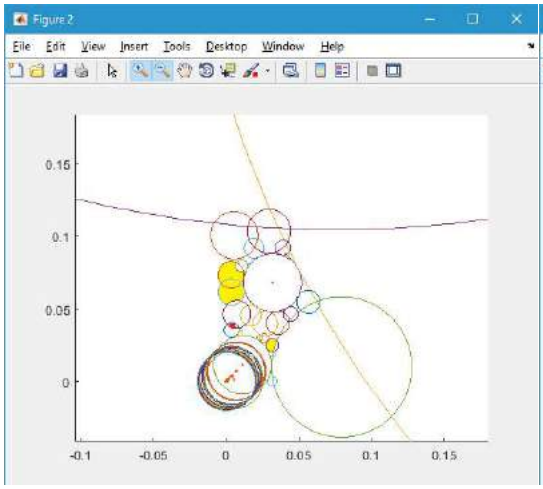
Mat - D - TA



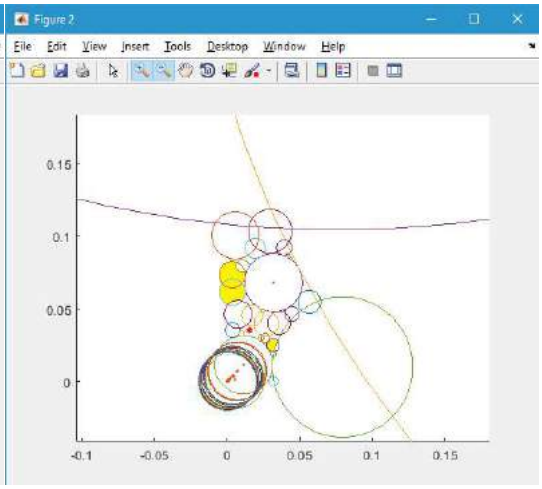
Mat - H - TA



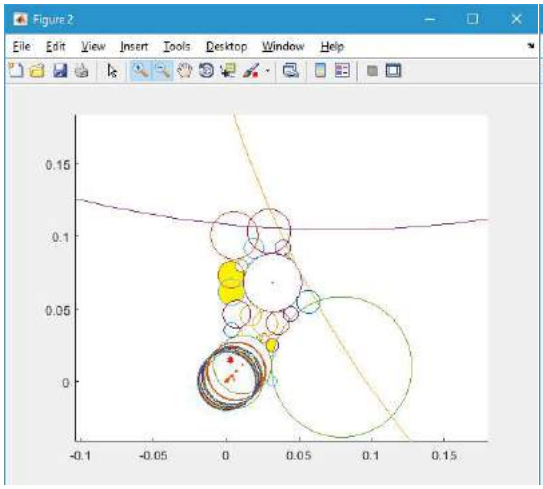
شکل ۲۴ - پلاتهای مربوط به هشت وضعیّتِ Sure در دادگانِ تستِ نخست
(بازنمایشِ دو بعدی با استفاده از دو ویژگی ۵ و ۶ دادگان UNSW – NB15)



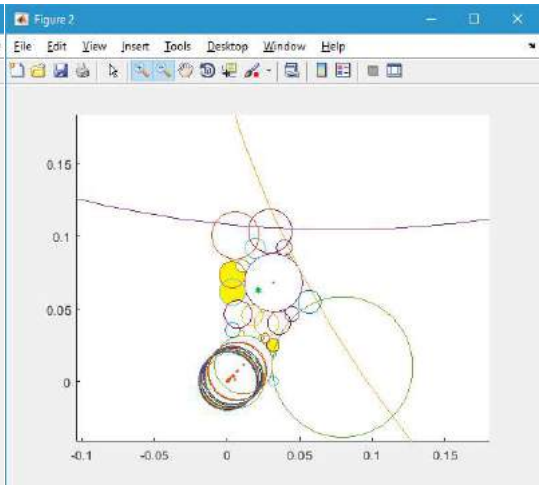
Mat - D - TN



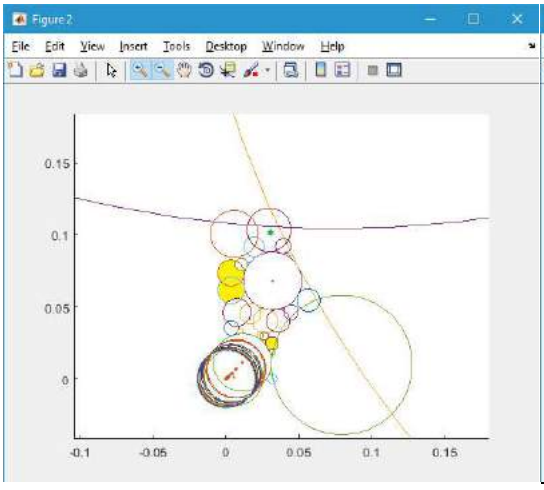
Mat - H - TN



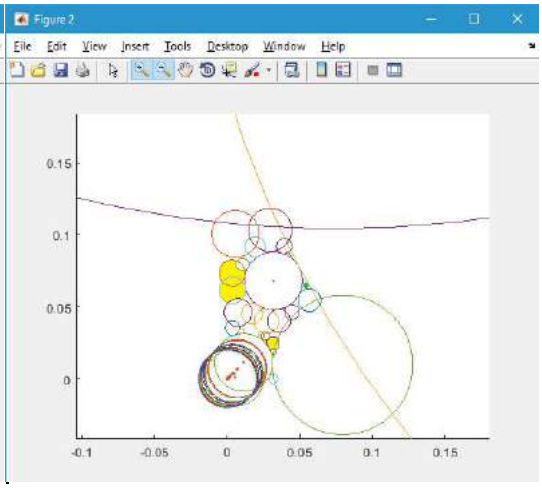
Mat - NSD



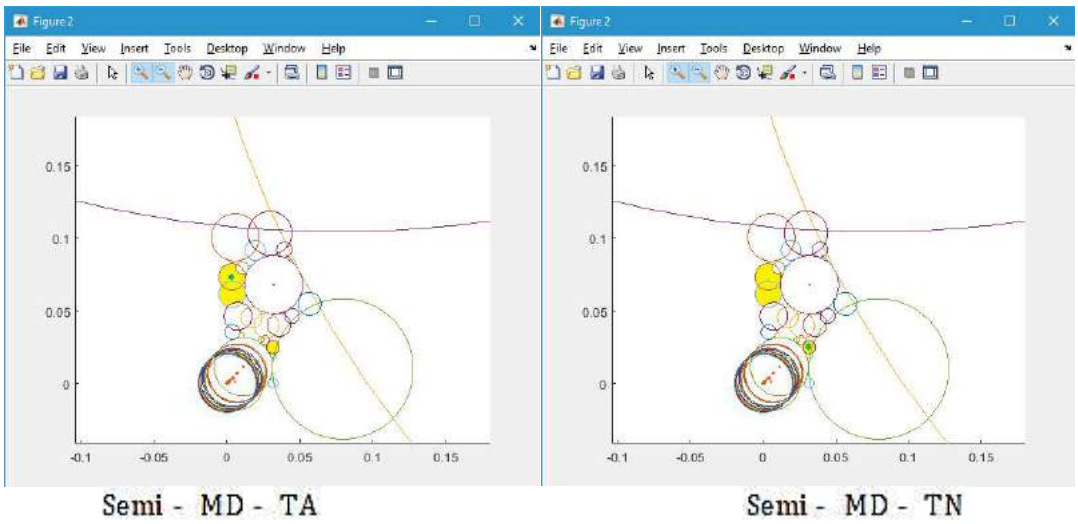
Semi - ASD



Semi - H - TA



Semi - D - TA



شکل ۲۵ - پلاتهای مربوط به هشت وضعیّتِ $not-Sure$ در دادگانِ تستِ نخست، بازنمایشِ دو بعدی با دو ویژگی ۵ و۶

با ترکیب این دو خط دفاعی ، مجموع حالات ممکن در روش هیبریدیِ پیشنهادی برابر با ۳۲ حالت ممکن خواهد بود. اگر خروجی دو وضعیت متوالی در خطوط دفاعی با هم برابر باشند وضعیت خروجی سیستم به حالت Sure رفته و بدین معناست که در سیکل جاری ، هر دو الگوریتم NS و DC نتیجه یکسانی را در خصوص برچسب احتمالی نمونه تست اعمال نموده اند. مثلاً اگر برچسب واقعی کلاس نمونه UA بوده و نتیجه اجرای دو خط دفاعی ، برچسب احتمالی نمونه را $N-UN$ ، $N-KN$,$A-UA$ یا $A-KA$ بزنند خروجی نهایی سیستم با اطمینان نزدیک به یقین می تواند برچسب نمونه را تشخیص صحیح بدهد.

دو حالت از این چهار حالت پیش آمده TP و دو حالت دیگر نیز FN خواهند بود. چالش اصلی ، کاهش رخداد دو وضعیتی است که منجر به خطای FN می گردند. در این مثال دو حالت $N-UN$ و $N-KN$ می باشند.

بنابراین سیستم پیشنهادی باید به گونه ای اجرا گردد که از بین چهار حالت ممکن برچسب زنی برای نمونه فرضی با برچسب واقعی X در نهایت پس از گذر از دو خط دفاعی به دو حالت True دست یابد. اگر $X=KN$ نرمال شناخته شده باشد پس از بین چهار حالت درستی ممکن که برابر با $N-KN, N-UN, A-KA, A-UA$ هستند باید تنها دو حالت $N-KN$,$N-UA$ رخ دهند تا در نهایت به حالت صحیح TN در هر دو خط دفاعی برسیم و با اطمینان برچسب نهایی را اعمال نماییم.

بنا به مشاهدات تجربی انجام شده در آزمایشات بخش های قبل، از انجایی که مکانیسمِ الگوریتمِ سلّولهای دِندریت (خط نخست دفاعی) به گونه ای است که همواره ثُبات در بالا بودنِ نرخ تشخیص آن بالاست بنابراین احتمال آنکه نمونه نفوذ (با برچسب واقعی UA یا KA) در خط نخست به طور صحیحی برچسب A بخورد

بالای ۸۳ درصد می باشد. حال در این وضعیت ، برچسب خروجی الگوریتم انتخاب منفی در خط دوّم ممکن است چهار برچسب KN, UN, KA, UA باشد که تنها دو برچسب KA یا UA حالت اطمینان (Sure) سیستم را منجر می شوند. از آنجایی که عملکرد خط دوّم نیز به گونه ای است که بنا به مشاهدات تجربی انجام شده همواره دو نرخ TP و TN آن بالای ۹۱ و ۹۴ درصد می باشند ، پس احتمال برچسب نهایی نمونه ی t برابر با ضرب دو احتمال $0.9191 \times 0.8325 = 0.7651$ برای TP و $0.9453 \times 0.5487 = 0.5186$ برای TP خواهد بود. به بیان بهتر ، بالای ۷۶ درصد احتمال دارد که هر دو خط دفاعی برچسب نمونه نفوذ را $A-KA$ و یا $A-UA$ شناسایی نمایند.

یک نکته ای که لازم است بدان اشاره شود آنست که اگر برچسب نمونه ی t که کلاس واقعی آن UA می باشد در خط دفاعی اوّل A و در خط دفاعی دوّم KA و یا UA شناسایی گردد در هر دو حالت بالاخره نفوذ (شناخته شده و یا ناشناخته) شناسایی شده و هر دو وضعیت True خواهد بود. برداشت خطا از وجود تفاوت بین KA و UA در اصل خطا نیست بلکه برچسب زنی اشتباه این دو توسط الگوریتم انتخاب منفی بیشتر به دلیل بروز نبودن پایگاه امضاء های حملات میباشد. برای مثال نمونه فرضی با برچسب واقعی KA را به اشتباه UA میزند.

امّا اگر برچسب آن توسط الگوریتم انتخاب منفی ، KN و یا UN شناسایی گردد به معنای رخداد خطای FN و وضعیت عدم اطمینان سیستم $not-Sure$ خواهد بود که احتمال رخداد این حالت با در نظر گرفتن کلّیه حالات 6.734925 درصد می باشد. این وضعیت مطابق با ردیفهای ۲۷-۳۰ جدول ۱۱ زمانی رخ خواهد داد که نمونه t با برچسب واقعی UA در خط نخست برچسب N به اشتباه خورده باشد و در خط دوّم نیز برچسب صحیح KA یا UA. درنتیجه احتمال رخداد این حالت نیز 15.394925 درصد خواهد بود. اگر در خط نخست برچسب A بخورد که صحیح است امّا در صورتیکه در خط دوّم به اشتباه KN یا UN خورده باشد این چهار حالت باز هم وضعیّت سیستم را به حالت عدم قطعیّت خواهند برد.

تحلیل ردیفهای ۱ تا ۸

برای نمونه ی t با برچسب واقعی نرمال- شناخته شده ، درمجموع هشت حالت ممکن الوقوع وجود دارد که چهار حالت آن با قطعیّت کامل (Sure) برچسب نهایی را با احتمال 51.868611 درصد درست (TN-KK یا TN-KU) و با احتمال 2.468611 درصد نادرست (FP-KK یا FP-KU) اعمال خواهند نمود. احتمال رخداد وضعیتهای $TN-1K$ و $FP-1K$ به ترتیب در حدود ۵۵ و ۴۵ درصد و در یک حدود می باشند. بدین معنی که احتمال آنکه نمونه t (با برچسب واقعی نرمال-شناخته شده) در خط دفاعی اوّل برچسب نرمال بخورد در مقایسه با حالتی که همین نمونه باشتباه برچسب آنومالی بخورد احتمال هر دو حالت تقریباً یکسان است. از طرفی در الگوریتم انتخاب منفی (خط دوّم) احتمال بالای ۹۱ درصد وجود دارد که به طور صحیحی

برچسب KN و یا UN را به نمونه بزند در عوض احتمال خطای مثبت کاذب آن پایین و در حدود کمتر از ۹ درصد می باشد. پس در مجموع با این تحلیل می توان نتیجه گرفت که احتمال رخداد ردیف های ۱ و ۲ بیشتر از رخداد ردیفهای ۷ و ۸ می باشد (جدول ۱۱)

تحلیل ردیفهای ۹ تا ۱۶

همانند تحلیل قبل می باشد و احتمال رخداد ردیفهای ۹ و ۱۰ بیشتر از احتمال رخداد ردیفهای ۱۵ و ۱۶ می باشند. زیرا ردیفهای ۹ و ۱۰ دارای دو برچسب متوالی خطوط دفاعی ، TN – TN بوده و در ردیفهای ۱۵ و ۱۶ ام نیز FP – FP می باشد.

تحلیل ردیفهای ۱۷ تا ۲۴

اگر برچسب نمونه t برابر با آنومالی-شناخته شده (یا KA) باشد ، خط دفاعی اوّل برچسب نمونه را باشتباه نرمال و یا آنومالی ممکن است بزند. مطابق مشاهدات تجربی حاصل از آزمایشات بخشهای قبل و میانگینهای بدست آمده از احتمالات برچسب زنی نمونه ها ، متوسط نرخ تشخیص TP-1K خط نخست به طور متوسط بالای ۸۳ درصد بوده و برای حالتی که برچسب نمونه باشتباه نرمال زده شود (FN-1K) در حدود کمتر از ۱۶ درصد می باشد. (جدول ۱۱) از طرفی در خط دفاعی دوّم نیز همین احتمالات صحّت و خطای برچسب زنی به ترتیب بالای ۹۵ و کمتر از ۵ درصد می باشد. بنابراین از دیدگاه تحلیل نظری می توان استنباط نمود که دو ردیف ۲۳ و ۲۴ بیشترین احتمال رخداد و ردیفهای ۱۷ و ۱۸ ام نیز کمترین میزان رخداد را خواهند داشت.

تحلیل ردیفهای ۲۵ تا ۳۲

مشابه تحلیل قبل می باشد و احتمال رخداد ردیفهای ۳۱و۳۲ بیشتر از احتمال رخداد ردیفهای ۲۵ و ۲۶ می باشند.

جدول ۱۳ – مجموع احتمالات محاسبه شده جدول ۱۱ برای سیکل نخست اجرای حیات مصنوعی (از دیدگاه تئوری)

Rows	Hybrid	Sum of Probability	Conf.of Hybrid	
			Total False Rates	Total True Rates
1 – 8	Sure	54.337222	FP-FP = 2.468611 (Total FP = 4.54 Percentage)	TN-TN = 51.868611 (Total TN = 95.46 Percentage)
	Not – Sure	45.662778	-	-
9-16	Sure	54.337222	FP-FP = 2.468611	TN-TN = 51.868611

	Not – Sure	45.662778	(Total FP = 4.54 Percentage)	(Total TN = 95.46 Percentage)
17-24	Sure	77.87015	-	-
	Not – Sure	22.12985	FN-FN = 1.355075 (Total FN = 1.74 Percentage)	TP-TP = 76.515075 (Total TP = 98.26 Percentage)
25-32	Sure	77.87015	-	-
	Not – Sure	22.12985	FN-FN = 1.355075 (Total FN = 1.74 Percentage)	TP-TP = 76.515075 (Total TP = 98.26 Percentage)

از دیدگاه تحلیل نظری ، مطابق اطلاعات جدول ملاحظه می گردد که سیستم تشخیص نفوذ پیشنهادی قادر است به طور متوسط حداقل در نیمی از موارد به طور قطع برچسب نمونه نرمال را با قطعیّت مشخص نماید که البته حدود ۴,۵۴ درصد از مواردی که به قطعیّت منجر شده و سیستم تشخیص در مورد برچسب نهایی آنها تصمیم گیری نهایی را اتّخاذ نموده اشتباه مثبت کاذب و ۹۵,۴۶ درصد تشخیص نرمال صحیح بوده اند.

در مورد ترافیک نفوذ از دیدگاه تئوری در سیکل نخست اجرای حیات مصنوعی ، سیستم در بیش از ۷۷ درصد از موارد به قطعیّت لازم در خصوص برچسب نهایی دست می یابد و حدود ۲۲ درصد نیز به عدم قطعیّت میرسد. نمونه های با "برچسب عدم قطعیّت" در سیکلهای بعدی مورد بررسی بیشتر قرار خواهند گرفت. از بین نمونه هایی که به قطعیت سیستم منجر شده اند ۱,۷۴ درصد از آنها به تشخیص اشتباه (FN) و ۹۸,۲۶ درصد به شناسایی صحیح منجر می شوند.

به عبارت دیگر در سیکل اوّل اجرای حیات مصنوعی در سیستم پیشنهادی - از بُعد تحلیل تئوری- از مجموع ۲۱۱۰ نمونه ترافیک شبکه در دادگان تست که ۶۶۰ مورد آنها نرمال (حدود ۳۱,۲۸ درصد) و مابقی ۱۴۵۰ نمونه آنومالی هستند (حدود ۶۸,۷۲ درصد) ۳۲۱ نمونه آنومالی و ۳۰۱ نمونه نرمال به حالت عدم قطعیّت می روند و در سیکل بعد مورد سنجش و ارزیابی قرار میگیرند.

حال در بین نمونه های باقیمانده که سیستم در مورد برچسب آنها به قطعیّت لازم رسیده (یعنی ۱۱۲۹ مورد ترافیک نفوذ و ۳۵۹ ترافیک نرمال) حدود ۱۹ نمونه علیرغم قطعی بودن برچسب آنها دارای خطای منفی کاذب هستند و حدود ۱۶ نمونه نرمال هستند که برچسب آنها توسط سیستم پیشنهادی به اشتباه آنومالی درج شده و دارای خطای مثبت کاذب می باشند. پس در مجموع از دیدگاه تئوری ، ۳۵ رکورد ترافیک شبکه (نرمال و غیر نرمال) از بین مجموع ۱۱۲۹+۳۵۹=۱۴۸۸ رکورد ترافیک شبکه که همگی ، قطعیّت سیستم تشخیص در مورد آنها ثابت شده و چیزی حدود ۲,۳۵ درصد را از مجموع نمونه های دارای قطعیّت ما را تشکیل می دهند دارای خطا خواهند بود که البتّه ناچیز است.

بنابراین می توان نتیجه گرفت که در مجموع نرخ تشخیص سیستم در سیکل نخست ۹۸,۲۶ درصد می باشد که در مقایسه با حالتی که الگوریتم انتخاب منفی یا الگوریتم سلولهای دندریت هر یک به طور مجزّا مورد استفاده قرار گیرند ($Dr_{NSA} = 94.53$, $Dr_{DCA} = 83.25$) نرخ بالایی به نظر میرسد. بنابراین با کاربرد " استراتژی تصمیم گیری قطعیّت / عدم قطعیّت" ملاحظه می گردد که از بُعد آنالیز نظری و با صرف نظر از

نمونه هایی که سیستم در مورد آنها به عدم قطعیّت لازم رسیده (در اینجا ۳۰۱+۳۲۱=۶۲۲ نمونه ترافیک شبکه) بهبود چشمگیری حاصل شده است. نتیجه این تحلیل را با مثالی ساده در ادامه متن اصلی پایان نامه توضیح داده ایم.

بخش پانزدهُم – تشریح روابط مربوط به درجه تشخیص

۱- اگر فاصله نمونه آنتی ژن t به فضای خودی - نرمال نزدیکتر باشد :

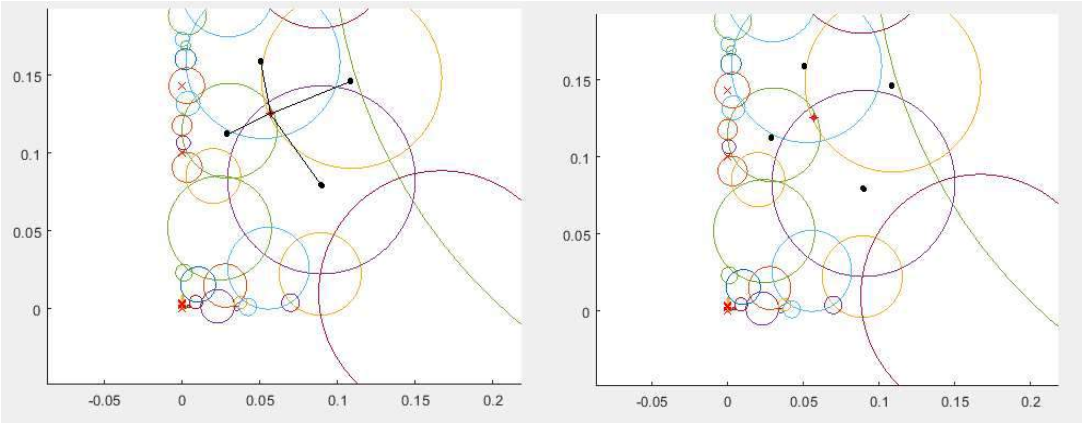
- اگر t واقع در حفره باشد که در این صورت روابط (۲) به منظور برچسب زنی به کار میرود. حال اگر وضعیت سیگنال خروجی آنتی ژن t در خط نخست را نیز در نظر بگیریم و سیگنال بالغ نسبت به سیگنال نیمه بالغِ آن قویتر باشد ، بدین ترتیب $\text{not} - \text{Sure}$ رخ داده است که امکان تصمیم گیری نهایی به سیکلهای بعدیِ سیستمِ تشخیص موکول گردیده و نمونه t در پایگاه موقّتی FeedBack بافر می شود تا در سیکل بعدی محتویات این بافِر تحویلِ لایه نخستِ سیستم گردد.

ولی اگر سیگنال نیمه بالغ قویتر باشد چون t در حفره واقع شده و ضمناً نزدیک به فضای خودی نرمال است بنابراین در این حالت با استفاده از روابط (۲) می توان برچسب نمونه را نزدیک به یقین " نرمال" دانسته و وضعیت Sure اتفاق خواهد افتاد. ضمناً حفره نیز در این حالت با آنتی ژن خودی نرمال با شعاع اولیّه r_s پوشش داده شده و ضمن کاهش پارامتر $\text{AVote} - -$ آن به اندازه یک واحد مقدار آن برابر -1 می گردد. سپس حفره های بوجود آمده و جدید اطراف آن نیز با ارسال سیگنالی به متد پیشنهادی در فاز یادگیری اوّلیه که از استراتژی مختصات قطبی برای پوشش استفاده می کند پوشش داده خواهند شد.

- اگر t را حداقل یک آنتی بادی بالغ حافظه پوشش دهد بدون در نظر گرفتن آنکه سیگنال بالغ یا نیمه بالغ قویتر است احتمالات برچسب نمونه با روابط (۱) محاسبه می گردند. در غیر اینصورت اگر آنتی بادی ، حافظه نباشد و تنها در صورتیکه سیگنال نمونه بالغ بوده باشد مانند بند اوّل رابطه الف) با آن برخورد شده و تمام آنتی بادیهای بالغ همپوشانِ [۱] آن نمونه همگی تبدیل به آنتی بادی حافظه شده و یک واحد به مقدار پارامتر $\text{AVote} = 1$ آنها افزوده می شود (شکل زیر).

در غیر اینصورت اگر سیگنال نیمه بالغ نمونه قویتر بود و توسط حداقل یک آنتی بادی غیر حافظه پوشش داده شده باشد مانند بند دوّم قسمت الف) با آن برخورد شده و روابط (۲) مورد در تعیین احتمال برچسب مورد استفاده قرار میگیرد.

[۱] اگر چند آنتی بادی در ناحیه اشتراکی بتوانند نمونه ای را شناسایی و پوشش دهند ، مشترکاً پارامتر AVote مربوط به تمامی آنتی بادیهای پوشش دهنده به دلیل همپوشانی نمونه مذکور ، یک واحد افزایش می یابد اگر نمونه نزدیک به فضای ASD باشد و یک واحد کاهش می یابد اگر نمونه به فضای NSD نزدیک تر باشد.

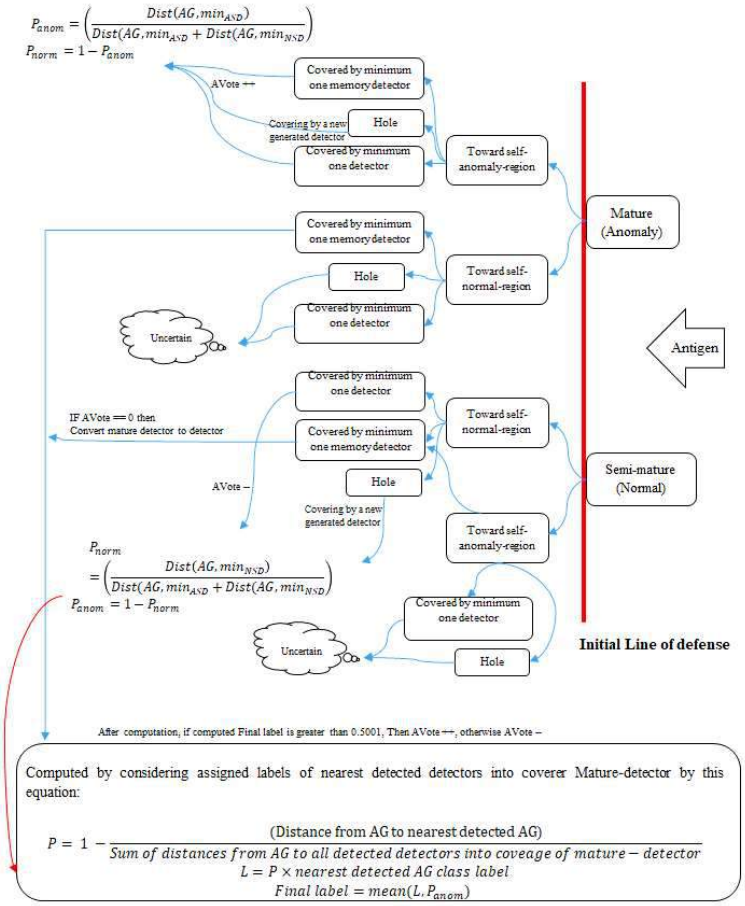


شکل ۲۶ – تحت پوشش واقع شدن یک نمونه تست (آنتی ژن) توسط چهار آنتی بادی بالغ غیر حافظه و تبدیل این آنتی بادیها به آنتی بادیهای بالغ حافظه ($\mathrm{AVote} = 1$)

۲- اگر فاصله نمونه آنتی ژن t به فضای خودی غیر نرمال (پایگاه امضاءهای حملات) نزدیکتر باشد :

- در صورتی که در خط دفاعی اوّل سیگنال بالغ t قویتر از سیگنال نیمه بالغ باشد ، چه آنتی بادی (های) پوشش دهنده حافظه باشند یا نباشند و حتّی در صورتی موقعیت مربوطه نیز حفره باشد در هر سه حالت روابط (۳) برای تعیین برچسب نهایی نمونه به کار می روند. ضمناً یک واحد به مقدار پارامتر $\mathrm{AVote}++$ هر کدام از آنتی بادیهای پوشش دهنده افزوده می شود. اگر $\mathrm{AVote} >= 1$ باشد آنتی بادی بالغ تبدیل به حافظه می شود و با افزایش این مقدار و فاصله گرفتن آن از صفر ، به تثبیت / تکثیر آن افزوده می شود. سپس سیگنال تعدیل و بروز رسانی به فاز یادگیری ارسال می شود تا حفره های احتمالی پوشش داده شوند.
- در صورتی که سیگنال نیمه بالغ t قویتر از سیگنال بالغ بود :

- وضعیت عدم قطعیّت $\mathrm{not} - \mathrm{Sure}$ پیش آمده و نمونه وارد سیکل بعدی سیستم میشود تا مجدّداً مورد بررسی قرار گیرد. ما در اینفوگرافیک زیر استراتژی پیشنهادی فوق را به طور دقیق با ذکر جزئیات کامل نشان داده ایم.



شکل ۲۷ - اینفوگرافیکِ مربوط به استراتژی پیشنهادی برای فاز تست خط دوّمِ دفاعی

پیوستِ ب - جداول ، اَشکال و نمودارها

بخش نخست – زیر مجموعه های یادگیری و تست استخراج شده از دو دادگانِ مَحک

در این قسمت زیر مجموعه دادگانِ مورد نیاز برای یادگیری و تستِ روشِ پیشنهادی را استخراج و نرمالیزه سازی نموده ایم.

۱-۱- مجموعه دادگان استفاده شده

ما به منظور تست و ارزیابیِ روشِ تشخیص نفوذ پیشنهادی از دو دادگانِ شبکه NSL – KDD و UNSW – NB15 استفاده نمودیم. دادگانِ دوّم جدیدترین حملات اخیر را گردآوری نموده که در نوع خود مجموعه جامعی به نظر میرسد. این دادگان برخی از مهمترین خصوصیات ترافیک شبکه را نیز در نظر گرفته که در دادگان قبلی موجود نبود که عبارتند از: طول مدّت کانکشن [1] ، تعداد بایتهای ارسالی و دریافتی بین مبداء و

[1] Connection time

مقصد[1] ، نرخِ ژیترِ (اختلال و بی ثباتی در ترافیک شبکه [2]) ، مهلت زمانِ آستانه عُمر بسته در مبداء و مقصد و غیره. از حیث دسته بندی زیر حملات ، مشخصات کامل تمامی حملات موجود در این دو دادگان در (۱-۲) به طور کامل توصیف شده اند. با دقّت در این پیوست مشاهده می گردد که این دو از نظر نوع ویژگیهای در نظر گرفته شده تا چه حد با هم تفاوت دارند. وجود این تفاوت این الزام را فراهم می کند که هر دو در فرایند ارزیابی در سری آزمایشات روش پیشنهادی استفاده شوند هر چند تاکید ما برای آزمایش نهایی ، استفاده از UNSW – NB15 می باشد.

۱-۱-۱- آماده سازی

UNSW – NB15 نُه زیر کلاسِ حمله دارد که با کلاسِ نرمال دَه مورد را شامل می شود. امّا این کلاسها را نمی توان به طور کلی در دسته بندی های رایج قرار داد. زیرا اصولاً شناسایی حملات R2L و U2R و اعمالِ تفکیک میان این دو کارِ مشکلی است. در بیشتر موارد یک الگوی ترافیک دریافت شده همیشه مرتبط با یک نوع حمله نبوده و ممکن است خود مقدمه ای باشد که در کنار سایر حملات آتی و ماقبل خود ، حمله ای به مراتب بزرگتر را ترتیب دهد. مانند حملات از نوع شناسایی ، که در واقع آنومالی ضِمنی بوده و خود در نهایت می توانند منجر به بروز آنومالی تجمّعی و شکل گیری حملاتی مانند انکار سرویس توزیع شده گردند. برای همین مثلاً ممکن است یک حمله از نوع R2L در نتیجه تاثیرِ مثبتِ سه نوع حمله shellcode, backdoor, reconnaissance ترتیب داده شود. پس در این مورد خاص ، این سه نوع زیر حمله برچسبِ R2L می خورند. در نتیجه به دلیل عدم امکانِ تبدیل و نگاشتِ کاملِ تمامی این نُه زیر حمله در دادگان نفوذ مذکور به یکی از چهار گروه حمله شناخته شده [3] ، این کار ، کار مشکلی است و پیشنهاد می شود در فاز پیش بینیِ مدل (تست و ارزیابی سیستم تشخیص) صرفاً از "دسته بندی باینری" استفاده شود.

امّا در مواردی مانند استخراج و گردآوری دو زیر مجموعه یادگیری و تست از دو دادگان مختلف ، معادلسازی حملات در این دو دادگان ضروری به نظر میرسد. ما مبنای ارزیابی مقایسه ای را بر پایه این چهار کلاس رایج قرار داده و پس از مشورت با برخی از مهمترین صاحبنظران و توسعه دهندگانِ دادگانِ UNSW – NB15 در سامانه آنلاین ریسرچ گیت [4] ، به پیشنهاد آنها معادلسازی زیر را در نظر گرفتیم.

[1] Network Packet Send & Recieve
[2] Jitter Rate (sjit / djit)
[3] DoS,Probe,U2R,R2L
[4] https://www.researchgate.net/post/About_Class_labels_of_the_KDDCU99

طبق این معادلسازی ، حملات با برچسب reconnaissance در دادگان جدید ، معادل حمله probe بوده و به جز حملات از نوع dos مابقی حملات با برچسب های Fuzzers, Shellcode, Exploits, Analysis, Generic, Worms که در زیر مجموعه هر دو نوع حمله Remote to local و User to root قرار میگیرند همگی برچسب جدیدِ Exploits می خورند زیرا هدفِ تمامی آنها " حمله شناسایی " یا "حمله انکار سرویس (DoS/DDoS) " نبوده و در نهایت منجر به نفوذ به سیستم قربانی می شوند.

۱-۱-۲- زیر مجموعه دادگان یادگیری و تست

در زیر مجموعه دادگانِ تست، علیرغم کمتر بودن حجم آن نسبت به زیرمجموعه دادگان یادگیری ، تعدادی حمله جدید تعبیه کردیم که در دادگان یادگیری موجود نبودند. این مورد جهت انجام کلیه آزمایشات " تشخیص و پیش بینی حملات ناشناخته " امری ضروری می باشد.

دادگان NSL – KDD بیش از ۴ و نیم میلیون و UNSW – NB15 نیز در مجموع به اندازه ۲۵۷۶۷۳ رکورد ترافیک شبکه را در خود جای داده اند. ما اندازه زیر مجموعه دادگان یادگیری و تست را برای هر دو دادگان به ترتیب برابر ۲۰۰۰۰ و ۲۱۱۰ رکورد ترافیک شبکه در نظر گرفتیم.

به منظور توزیع یکنواخت و یِکدَست بودن الگوهای ترافیک ، نحوه استخراج ترافیک از دادگانِ کُل برای تشکیل دو زیر مجموعه به صورت کاملاً تصادفی و یکنواخت با استفاده از نرم افزار متلب انجام شد. ما در بین ۲۱۱۰ رکورد ترافیکِ تست ، ۱۱۵ حمله جدید را تعبیه نمودیم. این میزان در حدود ۵,۴۵ درصد از کُل حجمِ ترافیک تست ما را تشکیل می دهند. مشخصات حملات تعبیه شده در جدول ۱۴ به تفکیک بیان شده اند.

جدول ۱۴ – جزئیات کلاسهای حملات موجود در دادگان نفوذ kddcup_data_10_percent_corrected و corrected

Category	Sub Category	Size	Percentage (%)
Normal		**157871**	19.69100
Probe	Ipsweep	1553	1.02763
	Nmap	315	
	Portsweep	1394	
	Satan	3222	
	Mscan	1053	
	Saint	736	
Dos	Back	3301	77.17669
	Land	30	
	Neptune	165202	

	Pod	351	
	Smurf	444881	
	Teardrop	991	
	Apache2	794	
	Mail bomb	5000	
	process table	759	
	Udpstorm	2	
U2R	Buffer overflow	52	0.034780
	Load module	11	
	Rootkit	23	
	Ps	16	
	httptunnel	158	
	Sqlattack	2	
	Xterm	13	
	Perl	5	
R2L	ftp write	11	2.150798
	Guess password	4420	
	Imap	13	
	Multihop	25	
	Warez client	1020	
	Warez master	1622	
	Snmp guess	2406	
	Snmp get attack	7741	
	Send mail	17	
	Named	17	
	Worm	2	
	Xlock	9	
	Xsnoop	4	
	Phf	6	
	Spy	2	
Total		**805050**	**1**

به دلیل عدم وجود برخی از حملات در kddcup_data_10_percent_corrected ما آنرا با پایگاه اصلاح شده corrected ترکیب نمودیم. مجموع اندازه این دو پایگاه برابر با ۸۰۵۰۵۰ رکورد ترافیک می باشد. (جدول ۱۴) نمودار زیر نحوه توزیع داده ها را در دو زیر مجموعه دادگان یادگیری و تست بدست آمده بهتر نشان می دهد.

نمودار ۸ - نمودارهای توزیع حملات در کلاسهای مختلف در دو دادگان نفوذ

۱-۱-۳- نُرمالیزه سازی

برای انجام آزمایش، ما ابتدا دادگان را نرمال سازی کردیم تا بتوان به راحتی با ویژگیهای آن کار کرد. زیرا بیشتر متدها نمی توانند با ویژگیهای از نوع غیر عددی [1] کار کنند. این فاز تحت عنوان فاز پیش پردازش

[1] Nominal

داده[1] می باشد. فاز پیش پردازش هم در تست و هم در مرحله یادگیری استفاده می شوند. پس از نرمالیزه سازی ، مقادیر بیشینه و کمینه هر ویژگی تعیین می شوند. ویژگیهای دو و سه و چهار باضافه ویژگی برچسب کلاس که در تست استفاده نمی شوند ، هر سه میبایست در تست و هم در فاز یادگیری نرمالیزه سازی شوند. نرمالیزه سازی روشهای مختلفی دارد و اغلب از روش $\min - \max$ استفاده میشود. ما نیز از این روش استفاده نمودیم. این فرایند به دلیل آنکه کلیّه ی داده ها را تَراز نموده و پس از یکسان سازی در رنجی در بازه [0,1] تبدیل می کند از این نظر روش مفیدی به نظر میرسد ضمن اینکه دسته بند به منظور انجامِ دسته بندی باینری نیازمند آنست که تمام ویژگیهای دادگانِ مورد آزمایش هم در تست و هم در یادگیری، مقادیری پیوسته بوده و از قبل نرمالیزه سازی شده باشند. از طرفی به دلیل آنکه متدهای ایمنی مصنوعی به طور کلّی از معیار فاصله به عنوان استراتژی و معیاری برای شناسایی عوامل نفوذ استفاده می کنند در نتیجه قاعدتاً نیازمند آن هستند که مقادیر تمام ویژگیها ضمن حقیقی بودن ، همگی در یک بازه مشخصی باشند.

نرمالیزه سازی یک شرط لازم برای کار با اکثر متدهای دسته بندی و بهینه سازی می باشد زیرا همانگونه که اشاره شد این متدها نمی توانند با مقادیر سَمبولیک کار کنند در نتیجه محاسبه آنها برای توابع مختلف مشکل است. مراحل اجرای الگوریتم نرمالسازی استفاده شده به ترتیب به صورت زیر می باشد.

- ابتدا ستون های مربوط به ویژگیهای سمبولیک (غیر عددی) تعیین می شوند (سه ویژگی دو تا چهار).
- سپس احتمالات مقادیر این ویژگیها محاسبه می شوند. مثلاً برای ویژگی نوع پروتکل (Protocol Type) تعداد هر یک از مقادیر ممکنِ این ویژگی (فرضاً تعداد tcp) به تفکیک جمع شده و بر کلّ تعداد عاملها در دادگان تقسیم می شوند تا احتمال tcp بدست آید. بدین صورت :

```
% calculate the probabilities of the protocol type
for i = 1: Nprotocol
M(i) = sum(strcmp(protocol_type(i), proto_col));
pdf_p(i) = M(i)/row_count;
end
```

این سورس مربوط به ویژگی پروتکل می باشد . برای بدست آوردنِ احتمالاتِ دو ویژگی دیگر (یعنی سرویس و فلگ) نیز سورس بالا استفاده میشود.

- سپس جایگزین کردن تمام احتمالات بدست آمده با مقادیرِ معادل غیر عددیِ این سه ویژگی در دادگان.
- ذخیره نمودن دادگان بدست آمده (با فرمتِ csv)
- خواندن فایل csv دادگان و نرمالیزه سازی $\min - \max$.

[1] PreProcessing Phase

بدین ترتیب با انجام عملیات نرمالیزه سازی تمامی مقادیرِ دادگان به مقداری در بازه $[0,1]$ تَراز شده و نرمال خواهند شد.

۱-۲- جزئیات دادگان تست و یادگیری استخراج شده

جدول ۱۵ - دو زیر مجموعه یادگیری و تست استخراج شده از دادگان kddcup_data_10_percent_corrected

Category	Class Label	Training Set Size	Test Set Size
Normal	*normal*	3867	354
Probe	Ipsweep	31	15
	Nmap	5	2
	Portsweep	36	17
	Satan	90	10
	Mscan	20	24
	Saint	13	16
Dos	Back	106	15
	Land	1	6
	Neptune	4102	357
	Pod	9	6
	Smurf	11089	978
	Teardrop	20	14
	Apache2	14	15
	Mail bomb	147	18
	process table	17	16
	Udpstorm	0	2
Exploits (U2R,R2L)	Buffer overflow	0	52
	Load module	0	7
	Rootkit	1	12
	Ps	0	10
	httptunnel	4	25
	Sqlattack	0	2
	Xterm	0	10
	Perl	0	0
	ftp write	1	5
	Guess password	127	11
	Imap	1	8
	Multihop	0	10
	Warez client	28	8
	Warez master	34	19
	Snmp guess	51	9
	Snmp get attack	185	30
	Send mail	1	5
	Named	0	4
	Worm	0	2
	Xlock	0	5
	Xsnoop	0	4
	Phf	0	5
	Spy	0	2
	New Attacks	0	115
	Total	20000	2110

جدول ۱۶ - زیر مجموعه یادگیری استخراج شده از دادگان UNSW_NB15_training_set با اندازه ۸۲۳۳۲ ترافیک شبکه

Category	Training Set			
	Class Label	Protocol		Size
Normal		*Tcp*	*6844*	9076
		Udp	*1973*	
		Arp	*241*	
		Igmp	*10*	
		Ospf	*8*	
Probe (reconnaissance)		*Tcp*	*457*	848
		Udp	*308*	
		Unas	*38*	
		Ospf	*14*	
		Sctp	*2*	
		Any, ax.25, cftp, compaq-peer, dgp, eigrp, encap,etherip, fire, gre, ib, idpr, ifmp, ip, ipcv, ipip, ipv6, ipv6-frag, ipv6-opts, isis, iso-tp4, mfe-nsp, micp, ptp, rdp, snp, st2,tp++, trunk-2	*1*	
Dos		*Unas*	*257*	942
		Tcp	*257*	
		Ospf	*52*	
		Sctp	*25*	
		Udp	*38*	
		Sep	*11*	
		Any	*7*	
		Rsvp,su-nd,swipe,	*6*	
		Ttp, trunk-2, stp, scps, pipe, leaf-2, i-nlsp,	*5*	
		Aris, ax.25, iatp, idrp, igp, ipv6-no, isis, iso-ip, mtp, nvp, pim, prm, sccopmce, wb-expak, xtp,	*4*	
		Xns-idp, uti, sm, sdrp, sat-expak, rvd, rdp, pnni, pgm, narp, micp, mhrp, merit-inp, leaf-1, kryptolan, iso-ipv4, ipv6-route, ipv6, iplt, ipcv, il, encap, eigrp, cpnx, cbt, aes-sp3-d, 3pc	*3*	
		a/n, Compaq-peer, crtp, crudp, ddp, ddx, etherip, gmtp, gre, ib, idpr, idpr-cmtp, ip, ipip, ipnip, irtp, larp, nsfnet-igp, pvp, qnx, sat-mon, secure-vmtp, snp, sprite-rpc, sps, st2, tlsp, vines, visa, vmtp, zero,	*2*	
		Argus, bbn-rcc, bna, br-sat-mon, cftp, chaos, dgp, egp, emcon, fc, fire, ggp, hmp, ifmp, ipcomp, ippc, ipv6-frag, mfe-nsp, mux, netblt, pri-enc, tcf, tp++, trunk-1, vrrp, wb-mon, xnet,	*1*	
Exploits (remaining)	*Fuzzers*	*Tcp*	*909*	1484
		udp	*274*	
		unas	*128*	
		ospf	*16*	
		Sctp,sm	*4*	
		Any, crudp, dcn, ib, micp, narp, pnni, rsvp, wb-mon	*3*	
		aes-sp3-d, argus, bna, crtp, ddp, egp, eigrp, fire, ggp, gre, idpr, idpr-cmtp, il, ip, iplt, ipv6, ipv6-frag, ipv6-opts, ipx-n-ip, irtp, kryptolan, mhrp, mobile, netblt, pim, pri-enc, pup, pvp, skip, smp, sprite-rpc, swipe, visa, vrrp, wb-expak, xnet, xns-idp, xtp	*2*	
		3pc, a/n, aris, bbn-rcc, br-sat-mon, cbt, cftp, chaos, compaq-peer, ddx, dgp, encap, fc, hmp, iatp, idrp, ifmp, igp, ipv6-no, isis, iso-ip, mfe-nsp, mtp, nsfnet-igp, ptp, qnx, rdp, rvd, sat-expak, sat-mon, sccopmce, sdrp,	*1*	

		sep, snp, sps, st2, stp, tlsp,tp++, trunk-1, trunk-2, ttp, uti, vines, vmtp, wsn		
	Shellcode	*Tcp*	*44*	81
		udp	*37*	
	Backdoors	*Unas*	*57*	154
		Tcp	*15*	
		Sdrp	*5*	
		Ospf	*4*	
		Sun-nd	*4*	
		Swipe, sccopmce, pup, prm, nvp, mobile, iso-ip, irtp, ipv6, ipip, ipcv, ip, iatp, hmp,	*2*	
		a/n, aes-sp3-d, argus, cftp, chaos, cpnx, crudp, ddp, ddx, eigrp, ggp, idpr, il, iplt,ipnip, ipv6-opts, ipv6-route, ipx-n-ip, iso-tp4, l2tp, merit-inp, micp, mtp, narp, netblt, nsfnet-igp, pgm, pipe, pvp, rdp, rsvp, sm, srite-rpc, srp, tcf, trunk-2, udp, wb-expak, wsn, xnet, xns-idp	*1*	
	Analysis	*unas*	*49*	166
		Tcp	*21*	
		Prm, any	*3*	
		Argus,ax.25, bbn-rcc, cbt,crudp, idrp, il, ip, ipv6-frag, ipx-n-ip, irtp, iso-tp4, pnni, pri-enc,qnx, scps, srp, tlsp, vrrp,xns-idp	*2*	
		Visa,vines, ttp, trunk-2, sun-nd, st2, snp,smp,sm, skip, secure-vmtp,sdrp, sccopmce, sat-expak,rvd,rsvp,pup,pim,nvp,narp,mtp,mfe-nsp,mert-inp,leaf-2,leaf-1,l2tp,kryptolan,iso-ip,ipv6-route,ipv6-opts,ipv6-no,ipv6, ippc,ipnip,iplt,ipip,igp,ifmp,ib,iatp,eigrp,egp,ddx,dcn,crtp,cphb,Compaq-peer,chaos,cftp, aris,	*1*	
	Exploits	*Tcp*	*1908*	2727
		Unas	*301*	
		Udp	*61*	
		Ospf	*58*	
		Sctp	*38*	
		ipv6, secure-vmtp	*7*	
		Any, gre, sun-nd, visa	*6*	
		bbn-rcc, cbt, cftp, idpr, ipcv, ipip, ipx-n-ip, rsvp, sep, sm,sprite-rpc	*5*	
		Cphb, cpnx, crtp, eigrp, ib, igp, ipv6-opts, irtp, isis, l2tp, nsfnet-igp, nvp, pipe, pri-enc, prm, rdp, rvd, smp, stp, trunk-2, wb-mon	*4*	
		3pc, aris, ddx, dgp, encap, ggp, gmtp, iatp, ifmp, iplt, ipv6-no, larp, micp, mtp, mux, pgm, pim, ptp, qnx, sdrp, skip, snp, sps, srp, trunk-1, vines, wb-expak, xnet	*3*	
		Argus, ax.25, bna, br-sat-mon, chaos, compaq-peer, dcn, ddp Egp, etherip, fire, i-nlsp, idpr-cmtp, idrp, il, ip, ipcomp, ipnip Ippc, ipv6-frag, ipv6-route, iso-ip, iso-tp4, leaf-1, merit-inp, mfe-nsp, mhrp, mobile, narp, pnni, pvp, sat-expak, sccopmce Scps, st2, swipe, tcf, tlsp, tp++, vrrp, xns-idp, xtp, zero	*2*	
		a/n, aes-sp3-d, crudp, fc, hmp, kryptolan, leaf-2, netblt, pup sat-mon,ttp, uti,vmtp,wsn	*1*	
	Generics	*udp*	*4385*	4510
		Tcp	*113*	

		Unas	*4*	
		Ospf, sctp, sun-nd	*2*	
		Mobile, prm	*1*	
	Worms	*Tcp*	*11*	12
		Udp	*1*	
Total	20000			

جدول ۱۷ - زیر مجموعه تست استخراج شده از دادگان UNSW_NB15_testing_set با اندازه ۱۷۵۳۴۱ ترافیک شبکه

Category	**Test Set**				
	Class Label	**Protocol**		**Size**	**Normalized Sub Class**
Normal		*Tcp*	*430*	660	0.31280
		Udp	*197*		
		Arp	*32*		
		Icmp	*1*		
Probe (reconnaissance)		*Tcp*	*57*	133	0.063033
		Udp	*43*		
		Unas	*7*		
		3pc,cpnx	*6*		
		St2	*3*		
		Ospf, bna, ddp, fc, pri-enc, sctp, srp, swipe, tcf, trunk-2, xns-idp	*1*		
Dos		*Unas*	*55*	175	0.082938
		Tcp	*37*		
		mobile	*25*		
		Ospf	*6*		
		Udp, sctp	*4*		
		Any	*3*		
		a/n, eigrp, sm, bbn-rcc, chaos	*2*		
		Aris,cftp, compaq-peer, encap, fc, iatp, idrp, ippc, ipv6,ipv6-opts, iso-ip, mfe-nsp, micp, nsfnet-igp, ptp, sccopmce, sep, sprite-rpc, sun-nd, visa, vrrp, xtp, zero, argus, cbt, emcon, igp, mux, pup,trunk-2, xnet	*1*		
Exploits (remaining)	*Fuzzers*	*Tcp*	*126*	211	0.1
		Udp	*57*		
		Unas,nvp,mux	*5*		
		Ospf, 3pc, br-sat-mon, cphb, iplt, mtp, pgm, sctp, sdrp, sun-nd, vrrp, ipnip, xns-idp	*1*		
	Shellcode	*Tcp*	*8*	16	0.007582
		Udp	*8*		
	Backdoors	*Unas*	*9*	17	0.008056
		Tcp, pgm	*2*		
		Bna, i-nlsp, pvp, swipe	*1*		
	Analysis	*Tcp*	*7*	21	0.099526
		Unas	*5*		

		Ospf, a/n, gre, ipcomp, ipv6-no, sctp, xtp Igp, ip	*1*		
	Exploits	*Tcp*	*259*	412	0.19526
		Unas	*57*		
		Ospf	*17*		
		Udp, sctp	*8*		
		aes-sp3-d, dgp, etherip, fire, ipv6, mobile, mtp, pipe, sm, wb-expak, leaf-2	*2*		
		a/n, any, aris, ax.25, bna, compaq-peer, cphb, crudp, ddp, eigrp, encap, i-nlsp, iatp, ib, idrp, ipcomp, iplt, ipv6-no, l2tp, mhrp, pim, ptp, sat-mon, secure-vmtp, stp.sun-nd, tcf,tlsp,tp++, uti, wb-mon, xtp, argus, bbn-rcc, chaos, irtp, leaf-1, netblt, prm, rdp, xnet	*1*		
	Generics	*Udp*	*444*	465	0.22038
		Tcp,ptp,unas, wb-expak, wb-mon, wsn, xtp, zero, cbt, chaos, dcn, emcon, ggp, igp, ip, ipnip, irtp, iso-tp4, leaf-1, leaf-2, mux	*1*		
	Worms	-	*0*	0	-
Total	2110				

در جدول فوق قسمت هایی که با رنگ زرد مشخص شده اند ، حملات ناشناخته هستند که در زیر مجموعه تست تعبیه شده اند.

جدول ۱۸ – معادلسازی و نگاشتِ زیر حملاتِ موجود در دو دادگانِ $NSL-KDD$، $UNSW_{NB15}$ به یکدیگر

Attack Category	
$ISCX\ NSL-KDD \mid UNB$	$UNSW-NB15$
Normal	
Dos	
Probe	***Reconnaissance***
U2R & R2L (Exploits)	***Fuzzers***
	Shell code *(FreeBSD,Linux,OpenBSD,SCO Unix, Mac OS X, Net BSD,BSD, IRIX, AIX, Windows,Decoders,Multiple OS,Solaris, HP-UX,NetBSD)*
	Analysis *(HTML, Port Scanner, Spam)*
	Backdoors
	Exploits

	(Evasions, SCCP, SSL, VNC, Backup Appliance, Browser,ClientSide Microsoft Office, Interbase, Miscellaneous Batch, Socks, TCP, Apache, IMAP, Microsoft IIS, SOCKS, Client Side Microsoft Paint, IDS, SSH, ICMP, DCERPC, FTP, RAIDUS, SSL, WINS, POP3, Unix r Service, Cisco IOS, Client Side Microsoft Media Player, Dame ware, IMAP, LPD, MSSQL, Office Document, RTSP, SCADA, VNC, Webserver, All, LDAP, NNTP, IGMP, Oracle, RDesktop, Telnet, LPD, Apache, PHP, SMB, SunRPC, Web Application, DNS, Evasions, SMTP, Browser FTP, PPTP, SCCP, SIP, TFTP)
	Generics *(All, SIP, HTTP, SMTP, IXIA, TFTP, Super flow)*
	Worms

بخش دوّم - جزئیّات دادگان تست استفاده شده برای انجام آزمایشات

ما بدین منظور از دادگان تست UNSW – NB15 استفاده نموده و در مجموع پنج زیر مجموعه ی تصادفی یکنواخت از این دادگان بزرگ در اندازه های ۲۱۱۰ رکورد و یکبار ۵۰۰۰ رکورد ترافیک شبکه استخراج نمودیم. نتایج استخراج بهمراه جزئیّات حملات در هر زیر مجموعه دادگان به صورت جدول زیر می باشد.

جدول ۱۹ – جزئیات دادگان تست استفاده شده در آزمایش حیات مصنوعی سیستم تشخیص نفوذ پیشنهادی

Subset	*Size of Records*	*Traffic Details*									
		Normal	*Analysis*	*Backdoor*	*DoS*	*Exploits*	*Fuzzers*	*Reconnaissanc*	*Shellcode*	*Worms*	*Generic*
1	*2110*	*660*	*21*	*17*	*175*	*412*	*211*	*133*	*16*	*0*	*465*
2		*656*	*19*	*32*	*149*	*385*	*210*	*154*	*12*	*4*	*489*
3		*678*	*14*	*28*	*126*	*395*	*225*	*133*	*19*	*1*	*491*
4		*673*	*16*	*21*	*183*	*427*	*202*	*124*	*10*	*2*	*452*
Total Test Set Size 8440		*2667*	*70*	*98*	*633*	*1619*	*848*	*544*	*57*	*7*	*1897*
Train Set 20000		*9076*	*166*	*154*	*942*	*2727*	*1484*	*848*	*81*	*12*	*4510*

بخش سوّم - نویسه های اختصاری

جدول ۲۰ - شرح نویسه های اختصاری موجود در جدول ۱۱

Abbreviation	***Description***
KN	*Known Normal*
UN	*Unknown Normal*
KA	*Known Abnormal*
UA	*Unknown Abnormal*

TN-1K	*TN Rate associated with First Defense Line when the actual class label is KN, and DCA's assigned label be Normal*
FP-1K	*FP Rate associated with First Defense Line when the actual class label is KN and DCA's assigned label be Abnormal*
TN-1U	*TN Rate associated with First Defense Line when the actual class label is UN and DCA's assigned label be Normal*
FP-1U	*FP Rate associated with First Defense Line when the actual class label is UN and DCA's assigned label be Abnormal*
FN-1K	*FN Rate associated with First Defense Line when the actual class label is KA and DCA's assigned label be Normal*
TP-1K	*TP Rate associated with First Defense Line when the actual class label is KA and DCA's assigned label be Abnormal*
FN-1U	*FN Rate associated with First Defense Line when the actual class label is UA and DCA's assigned label be Normal*
TP-1U	*TP Rate associated with First Defense Line when the actual class label is UA and DCA's assigned label be Abnormal*
TN-KK	*TN Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be KN* *OR* *TN Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be KA*
TN-KU	*TN Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be UN* *OR* *TN Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be UA*
TN-UK	*TN Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be KN* *OR* *TN Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be KA*
TN-UU	*TN Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be UN* *OR* *TN Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be UA*
FP-KK	*FP Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be KN* *OR* *FP Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be KA*
FP-KU	*FP Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be UN* *OR* *FP Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be UA*
FP-UK	*FP Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be KN* *OR* *FP Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be KA*
FP-UU	*FP Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be UN* *OR* *FP Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be UA*
FN-KK	*FN Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be KN* *OR*

	FN Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be KA
FN-UK	*FN Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be KN* *OR* *FN Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be KA*
FN-KU	*FN Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be UN* *OR* *FN Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be UA*
FN-UU	*FN Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be UN* *OR* *FN Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be UA*
TP-KK	*TP Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be KN* *OR* *TP Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be KA*
TP-KU	*TP Rate associated with Second Defense Line when the actual class label is KN and NSA's assigned label be UN* *OR* *TP Rate associated with Second Defense Line when the actual class label is KA and NSA's assigned label be UA*
TP-UK	*TP Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be KN* *OR* *TP Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be KA*
TP-UU	*TP Rate associated with Second Defense Line when the actual class label is UN and NSA's assigned label be UN* *OR* *TP Rate associated with Second Defense Line when the actual class label is UA and NSA's assigned label be UA*
Semi-NSD	*Antigen be Semi-Mature on First Defense Line , and it's position-Vector be inside of NSD Space*
Semi-D-TN	*Antigen be Semi-Mature on First Defense Line ,it is detected by an Antibody and it's position-Vector be toward to the Normal Space (NSD)*
Semi-H-TN	*Antigen be Semi-Mature on First Defense Line ,it is not detected by an Antibody and it's position-Vector toward to the Normal Space (NSD)*
Semi-ASD	*Antigen be Semi-Mature on First Defense Line , but it's position-Vector be inside of ASD Space*
Semi-D-TA	*Antigen be Semi-Mature on First Defense Line ,it is detected by an Antibody ,but it's position-Vector be toward to the Normal Space (ASD)*
Semi-H-TA	*Antigen be Semi-Mature on First Defense Line ,it is not detected by an Antibody ,but it's position-Vector toward to the Normal Space (ASD)*
Semi-MD-TA	*Antigen be Semi-Mature on First Defense Line ,it is detected by a Memory-Antibody ,but it's position-Vector be toward to the Normal Space (ASD)*
Semi-MD-TN	*Antigen be Semi-Mature on First Defense Line ,it is detected by a Memory-Antibody and it's position-Vector be toward to the Normal Space (NSD)*
Mat-NSD	*Antigen be Mature on First Defense Line , but it's position-Vector be inside of NSD Space*
Mat -D-TN	*Antigen be Mature on First Defense Line ,and it is detected by an Antibody, but it's position-Vector be toward to the Normal Space (NSD)*
Mat -H-TN	*Antigen be Mature on First Defense Line ,and it is not detected by an Antibody ,but it's position-Vector toward to the Normal Space (NSD)*
Mat –ASD	*Antigen be Mature on First Defense Line , and it's position-Vector be inside of ASD Space*

Mat -D-TA	*Antigen be Mature on First Defense Line ,and it is detected by an Antibody ,and it's position-Vector be toward to the Normal Space (ASD)*
Mat -H-TA	*Antigen be Mature on First Defense Line ,and it is not detected by an Antibody ,and it's position-Vector toward to the Normal Space (ASD)*
Mat -MD-TA	*Antigen be Mature on First Defense Line ,and it is detected by a Memory-Antibody ,and it's position-Vector be toward to the Normal Space (ASD)*
Mat -MD-TN	*Antigen be Mature on First Defense Line ,and it is detected by a Memory-Antibody , but it's position-Vector be toward to the Normal Space (NSD)*

بخش چهارُم

جدول ۲۱ – بهترین پیکربندی در هشت رویکرد الگوریتم سلولهای دندریت پیشنهادی در دادگان تست نخست (میانگین ۲۰ بار آزمایش)

Test ID	Proposed DCA Approach				Iteration	Runtime	Recall	FPR	TNR	FNR	ACC	ERR	Proposed Function for Fitness Evaluation. (3-11)
	Min Clone Size for PresentedAg	Max Clone Size for PresentedAg	Radius of Presenting DCs	Migration Threshold									
1	5	10	by mean	by median	21	0.178856	90.1	51.2	48.8	9.9	77.2	22.8	59.8336
2	3	10	by mean	by median	21	0.176326	92.8	52.7	47.3	7.2	78.6	21.4	62.8232
3	3	15	by mean	by median	21	0.223302	91.7	52.9	47.1	8.3	77.7	22.3	60.0722
4	5	15	by max	by median	21	0.186085	93.6	57.6	42.4	6.4	77.6	22.4	55.648
5	5	15	by max	by mean	21	0.225727	97	63.2	36.8	3	78.2	21.8	52.6452
6	5	15	by median	by mean	21	0.20068	95.5	62.9	37.1	4.5	77.3	22.7	50.1726
7	5	15	by median	by median	21	0.181126	90.8	52.4	47.6	9.2	77.3	22.7	59.1394
8	5	15	by mean	by mean	21	0.202256	95.1	55.8	44.2	4.9	79.2	20.8	62.0432
Best proposed Approach for DCA's Input Parameters													
9	5	15	AVG	By median	21	0.255311	95.9	59.2	40.8	4.1	78.7	21.3	57.5528

جدول ۲۲ – بهترین پیکربندی هشت رویکرد الگوریتم سلولهای دندریت پیشنهادی با اعمال انتخاب الگوریتم ویژگی ده-پا در دادگان تست نخست (میانگین ۲۰ بار آزمایش)

Test ID	Proposed DCA Approach				Feature Subset	Iteration	Runtime	Recall	FPR	TNR	FNR	ACC	ERR	Proposed Function for Fitness Evaluation. (3-11)
	Min Clone Size for PresentedAg	Max Clone Size for PresentedAg	Radius of Presenting DCs	Migration Threshold										
1	5	15	by mean	by median	14	8	0.249765	82	57.9	42.1	18	69.5	30.5	33.194
[2,3,4,7,10,12,13,15,28,29,32,33,34,40]														
2	5	20	by mean	by median	16	8	0.185186	77.3	49.8	50.2	22.7	68.8	31.2	37.528
[2,3,4,5,6,7,10,12,13,15,28,29,32,33,34,40]														
3	3	10	by mean	by median	16	8	0.179267	78.6	45.8	54.2	21.4	71	29	46.286
4	3	15	by max	by median	16	8	0.175893	79.6	53.8	46.2	20.4	69.1	30.9	35.3466
5	5	50	by max	by median	16	8	0.178993	83.7	56.7	43.3	16.3	71	29	38.05
6	5	50	by median	by mean	16	13	0.206841	91.2	58.8	41.2	8.8	75.5	24.5	48.679
7	5	50	by mean	by median	16	8	0.207566	80.1	54.2	45.8	19.9	69.3	30.7	35.5904
8	5	50	by median	by median	16	8	0.173867	79.1	45.6	54.4	20.9	71.4	28.6	47.552
9	5	15	By mean	By median	42	19	0.181506	94.8	48.6	51.4	5.2	81.2	18.8	74.8408

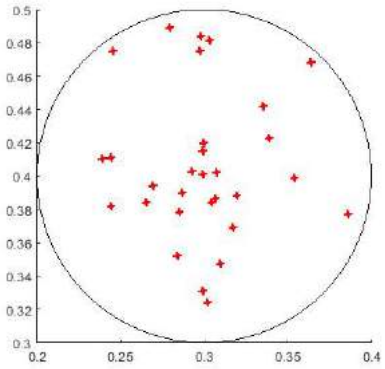
- **شِبِه کُد مربوط به کاربرد استراتژی متخصات قطبی برای دو بعد**

```
circle_raduis = 0.1;
circle_center = [0.3 0.4]; % [X Y]
rectangle('Position',[circle_center(1) − circle_raduis,circle_center(2) − circle_raduis,2
                ∗ circle_raduis,2 ∗ circle_raduis],'Curvature',[1,1]);
axis ([0 1 0 1]);
axis ('image');
hold on
nodes_x = 0;
nodes_y = 0;
for nn = 1:30
  a = 0;
  b = circle_raduis;
  distance = a + (b − a) ∗ rand(1);

  a = 0;
```

```
  b = 2 * pi;
  teta = a + (b - a) * rand(1);

  nodes_x(nn) = distance * cos(teta);
  nodes_y(nn) = distance * sin(teta);
    plot(circle_center(1) + nodes_x, circle_center(2)
                + nodes_y, 'rs', 'LineWidth', 5, 'MarkerSize', 1.5);
end
nodes_x = nodes_x + circle_center(1);
nodes_y = nodes_y + circle_center(1);
```

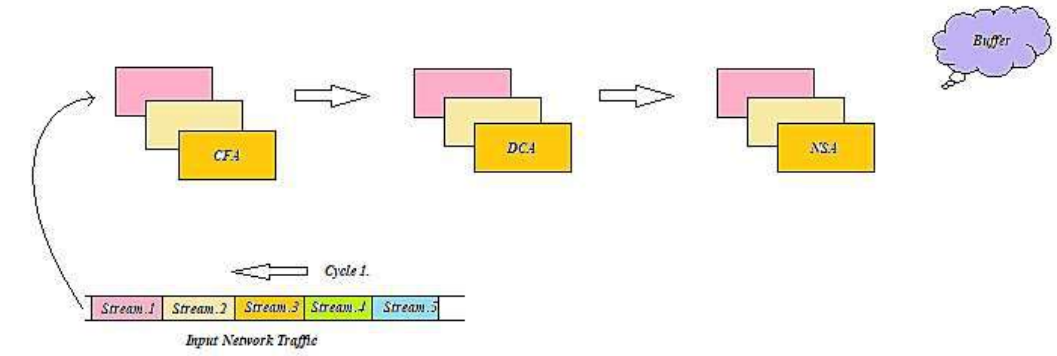


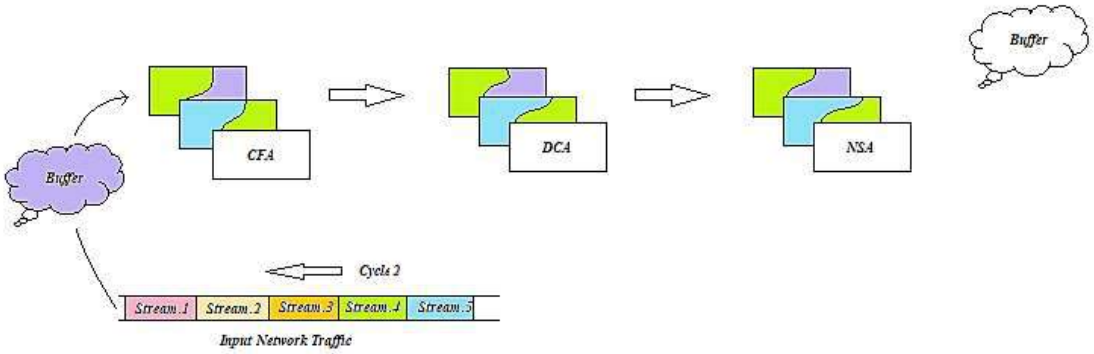
شکل ۲۸ - سورس کد متلب مربوط به کاربرد مختصات قطبی برای تولید 30 نقاط تصادفی درون دایره (0.3,0.4) به شعاع 0.1 بهمراه پلات

جدول ۲۳ – ارزیابی مقایسه ای روش هیبریدی پیشنهادی در دادگان تست نخست

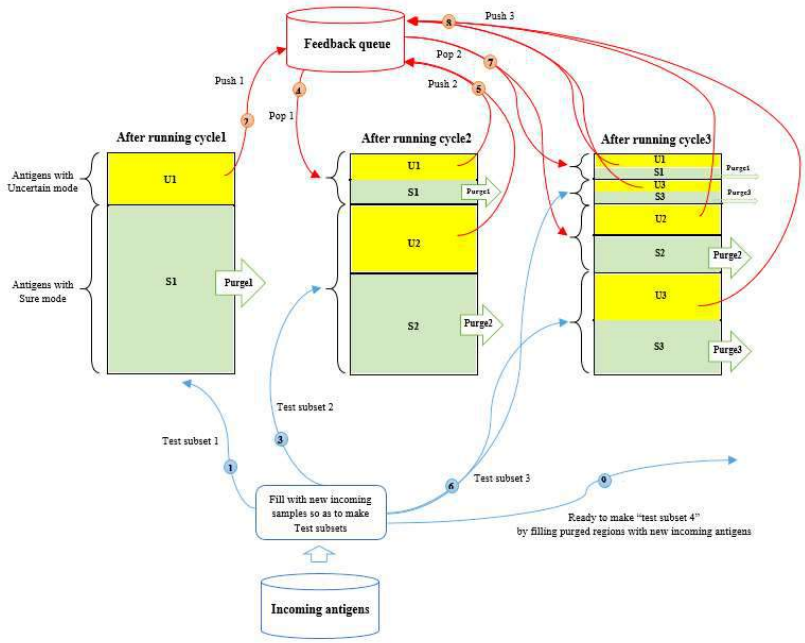
Method ID	Method	DR	FPR	TNR	FNR	ACC	ERR	FM	PR	Sens	Spec	cc
1	CoarseKNN	77.8	5.3	94.7	22.2	83.1	16.9	84.98	93.62	77.80	94.7	0.7356
2	CoarseKNN_PCA	77.4	6.1	93.9	22.6	82.6	17.4	84.36	92.69	77.40	93.9	0.7229
3	CoarseTree	65.7	7.9	92.1	34.3	73.9	26.1	75.69	89.26	65.63	92.1	0.5993
4	CoarseTree_PCA	75.9	12.6	87.4	24.1	79.5	20.5	80.53	85.76	75.90	87.4	0.6372
5	CubicKNN	77.6	15.3	84.7	22.4	79.8	20.2	80.46	83.53	77.60	84.7	0.6246
6	CubicKNN_PCA	81	7.3	92.7	19	84.6	15.4	86.03	91.73	81	92.7	0.7421
7	CubicSVM	72.1	22.6	77.4	27.9	73.8	26.2	74.06	76.13	72.10	77.4	0.4957
8	CubicSVM_PCA	79	17.9	82.1	21	80	20	80.24	81.52	79	82.1	0.6113
9	EnBaggedTrees	86.5	3.6	96.4	13.5	89.6	10.4	91.00	96.00	86.50	96.4	0.8331
10	EnBaggedTrees_PCA	81.3	12.7	87.3	18.7	83.2	16.8	83.81	86.48	81.30	87.3	0.6872
11	EnBoosted	46.3	2.9	97.1	53.7	62.2	37.8	62.06	94.10	46.30	97.1	0.5039
12	EnBoosted_PCA	90.9	12.1	87.9	9.1	90	10	89.56	88.25	90.90	87.9	0.7884
13	FineGausianSVM	65.7	0.2	99.8	34.3	76.4	23.6	79.20	99.69	65.70	99.8	0.6968
14	FineGausianSVM_PCA	79.8	4.4	95.6	20.2	84.7	15.3	86.64	94.77	79.80	95.6	0.7636
15	FineKNN	83.2	19.1	80.9	16.8	82.5	17.5	82.25	81.33	83.20	80.9	0.6412
16	FineKNN_PCA	82.6	8.6	91.4	17.4	85.3	14.7	86.40	90.57	82.60	91.4	0.7429
17	FineTree	24.4	2	98	75.6	47.4	52.6	38.61	92.42	24.40	98	0.3309
18	FineTree_PCA	81	13.8	86.2	19	82.7	17.3	83.16	85.44	81	86.2	0.6729

19	LinearDiscriminant	62.4	5.2	94.8	37.6	72.6	27.4	74.46	92.30	62.40	94.8	0.6046
20	LinearDiscriminant_PCA	92.2	17.1	82.9	7.8	89.3	10.7	88.10	84.35	92.20	82.9	0.7543
21	LinearSVM	70.5	1.5	98.5	29.5	79.2	20.8	81.98	97.92	70.50	98.5	0.7188
22	LinearSVM_PCA	79	7.4	92.6	21	83.3	16.7	84.77	91.43	79	92.6	0.7227
23	LogisticReg	4.9	0.8	99.2	95.1	34.4	65.6	9.27	85.97	4.9	99.2	0.1232
24	LogisticReg_PCA	90.5	16.4	83.6	9.5	88.3	11.7	87.48	84.66	90.50	83.6	0.7428
25	MediumGausianSVM	81.7	5.6	94.4	18.3	85.6	14.4	87.24	93.58	81.70	94.4	0.7672
26	MediumGausianSVM_PCA	77.2	3.9	96.1	22.8	83.1	16.9	85.26	95.20	77.20	96.1	0.7465
27	MediumKNN	78.3	7	93	21.7	82.9	17.1	84.51	91.79	78.30	93	0.7208
28	MediumKNN_PCA	80.6	6.8	93.2	19.4	84.5	15.5	86.02	92.22	80.60	93.2	0.7439
29	MediumTree	35	3.3	96.7	65	54.3	45.7	50.61	91.38	35	96.7	0.4028
30	MediumTree_PCA	78.6	12.3	87.7	21.4	81.4	18.6	82.35	86.47	78.60	87.7	0.6658
31	QuadraticSVM	70.8	5	95	29.2	78.3	21.7	80.55	93.40	70.80	95	0.6782
32	QuadraticSVM_PCA	71.7	18.9	81.1	28.3	74.6	25.4	75.24	79.14	71.70	81.1	0.5303
33	Weighted KNN	80.3	7.9	92.1	19.7	84	16	85.33	91.04	80.30	92.1	0.7291
34	Weighted KNN_PCA	81.9	8.9	91.1	18.1	84.8	15.2	85.85	90.20	81.90	91.1	0.7331
35	**Proposed Hybrid iNIDS (After 3 Cycles of Artificial Life)**	**98.7**	**17.2**	**82.8**	**1.3**	**96.2**	**3.8**	**91.43**	**85.16**	**98.7**	**82.8**	**0.8255**
	Number of remained not-Sure samples	**79**					**$\alpha = \beta$**					
36	**Proposed Hybrid iNIDS (After 5 Cycles of Artificial Life)**	**99.1**	**15.3**	**84.7**	**0.9**	**97.3**	**2.7**	**92.45**	**86.63**	**99.1**	**84.7**	**0.8468**
	Number of remained not-Sure samples	**34**					**$\alpha = \beta$**					
37	**Proposed Hybrid iNIDS (After 5 Cycles of Artificial Life)**	**99.4**	**18**	**82**	**0.6**	**97.4**	**2.6**	**91.45**	**84.67**	**99.4**	**82**	**0.8266**
	Number of remained not-Sure samples	**34**					**$\alpha = 0.3, \beta = 0.7$**					
38	**Proposed DCA**	**95.9**	**59.2**	**40.8**	**4.1**	**78.7**	**21.3**	**75.19**	**61.83**	**95.9**	**40.8**	**0.4398**
	Radius of Presenting DCs	AVG	**Migration Threshold**		By median	**Iteration Steps**	21	**Signal Generation Method**	IG	**Min-max of Clone Presenting Antigen**	5-15	
39	**Real Valued NSA**	**98.6**	**16.2**	**83.8**	**1.4**	**93.9**	**6.1**	**91.81**	**85.89**	**98.6**	**83.8**	**0.8332**





شکل ۲۹ – اینفوگرافیکِ مربوط به سیکلهای اجرای حیات مصنوعی در روش تشخیص نفوذ هیبریدی پیشنهادی

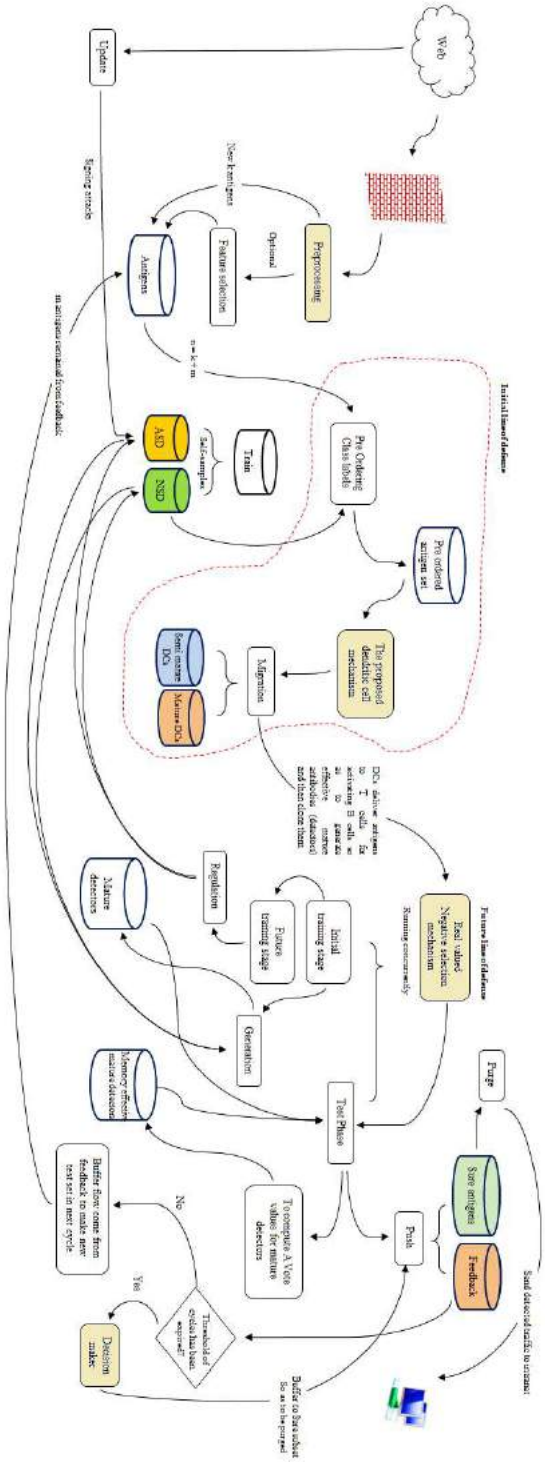


شکل ۳۰- اینفوگرافیکِ مربوط به سازوکار ماژول تصمیم گیری نهایی

Original Class Label	First Defense Line DCA by Exp	Second Defense Line (RNSA)	Conf.1	Conf.2	Evaluation of Results	State
Normal	Normal	Normal	TN1 △	TN2 △	△△	Sure
Normal	Normal	Anomaly	TN1 △	FP2 ▽	△	-
Normal	Anomaly	Normal	FP1 ▽	TN1 △	△	-
Normal	Anomaly	Anomaly	FP1 ▽	FP2 ▽	▽▽	Sure
Anomaly	Normal	Normal	FN1 ▽	FN2 ▽	▽▽	Sure
Anomaly	Normal	Anomaly	FN1 ▽	TP1 △	△	-
Anomaly	Anomaly	Normal	TP1 △	FN2 ▽	△	-
Anomaly	Anomaly	Anomaly	TP1 △	TP2 △	△△	Sure

Original Class Label	First Defense Line DCA by IG	Second Defense Line (RNSA)	Conf.1	Conf.2	Evaluation of Results	State
Normal	Normal	Normal	TN1 ▽	TN2 △	△	Sure
Normal	Normal	Anomaly	TN1 ▽	FP2 ▽	▽▽	-
Normal	Anomaly	Normal	FP1 △	TN1 △	△△	-
Normal	Anomaly	Anomaly	FP1 △	FP2 ▽	△	Sure
Anomaly	Normal	Normal	FN1 ▽	FN2 ▽	▽▽	Sure
Anomaly	Normal	Anomaly	FN1 ▽	TP1 △	△	-
Anomaly	Anomaly	Normal	TP1 △	FN2 ▽	△	-
Anomaly	Anomaly	Anomaly	TP1 △	TP2 △	△△	Sure

شکل ۳۱ – جدولِ تحلیل نظری وضعیّتهای ممکن ترافیک شبکه در خطوط دفاعی



شکل ۳۲ – معماری سیستم تشخیص نفوذ هیبریدی پیشنهادی

243 -244

پیوست ج

Meth	Title	Goal	Author(s)	Pros And Cons	Future Works
RNSA	An efficient negative selection algorithm with further training for anomaly detection	To reduce time complexity of test phase by regulating the self-region	**FtNSA** by Maogu G, and et al. Ref.[6] (2012)	**Adv:** 1) proposing the future stage to regulate self-normal-region by generating self-detectors (SD) and improve the self-region coverage and holes 2) reduce test time **Dis:** 1) KDDCup99 is old 2) lack of regulation for self-abnormal-region	Focusing on the optimization of the generated detectors distribution and the reduction of detector overlap
	An immune optimization based real-valued negative selection algorithm	To cover maximum non-self-space and its regulation through generating minimum detectors	**IO-RNSA** by Xin X, and et al. Ref [8] (2014)	**Adv:** 1) Utilizing the Polar coordinates idea for detector mutation 2) Comparative analysis of several RNS algorithms under UCI datasets **Dis:** 1) Complexity for implementation 2) computing the exact coverage of non-self is challenge due to duplication among detectors 3) only UCI datasets was used for evaluation while intrusion detection datasets should also be taken	Proposing more efficient negative selection algorithm for dynamic self-set
	A distributed intrusion detection system using multi-agent AIS approach	Designing a hybrid anomaly IDS by using the AIS paradigms in KVM hypervisor to achieve to a collaborative immunity and self-adaptation	**MAIS-IDS** by Afzali seresht N, and et al. Ref [9] (2014)	**Adv:** 1) Each detector agent is responsible for identifying non-self-antigens from self-ones in a decentralized mode 1) working on essential characteristics of AIS methods are collaborating, cloning, mutation and memory mechanisms in a host based distributed architecture using KVM environment **Dis:** 1) Only 19 features have been used only by recommendation of a reference while it was better to adopt a feature selection algorithm for doing it. 2) Security challenge especially for inform connections between hosts in real world conditions.	Implementing MAIS-IDS in real environment
	An Efficient Proactive Artificial Immune System based Anomaly Detection and Prevention System	To achieve dissimilarity among detectors by Self-tuning which makes detector adapt dynamically with respect to the detector's closeness towards the self-sample. *Response Module* takes measure to prevent attacks. It achieves this task through collaborative agents	**EPAADPS** by P Saurabh and B Verma. Ref [10] (2016)	**Adv:** 1) To achieve self-tuning characteristic to provide equations for disproportion among detectors which aids to a better coverage of space 2) The mechanism of Voting employed in VAM module helps the system in lowering down the FP rate **Dis:** 1) KDDCup99 is old 2) Don't showing up how to analyze unseen anomalies by charts separately	
Danger Theory (DCA)	A survey of the dendritic cell algorithm	To make general review of the powerful characteristics of the DCA and discuss about open research areas on providing improved approaches for the DCA	By Zeinab Chelly and Zeid Elouedi. Ref [15] (2015)	**Adv:** 1) Almost all of approaches for DCA improvements are evaluated 2) Good to utilize as a baseline for proposed hybrid method	
	Network Intrusion detection Using Danger Theory and Genetic Algorithm	to calculate *p safe* and *p danger* values for DCA	By J.L.Santanelli and F.B de Lima Neto. Ref [45] (2017)	**Adv:** Adopting GA to analyze which one of selected attribute subsets is optimal to calculate *p safe* and *p danger* signal values **Dis:** KDDCup99 is old	To apply the concepts of Distributed Systems to enhance the performance of the Genetic Algorithm, allowing it possibly to reach better results so as to obviate high time complexity challenge

	Theoretical and Empirical extensions of the dendritic cell algorithm		PHD thesis by FENG GU University of Nottingham. Ref [47] (2012)	**Adv:** Provide a basic foundations for investigating the signal calculation problem and preprocessing phase of DCA	Immuno-Engineering approach for developing immune-inspired algorithms for anomaly detection by combining with machine learning techniques. The automated data pre-processing methods are tailored for the DCA, however they still show the possibility of using techniques in machine learning and statistics to widen the use and improve the performance of immune inspired algorithms
	CloudIDS: Cloud Intrusion Detection Model Inspired by Dendritic Cell mechanism	To make a decentralized mechanism for DC so as to model its activities completely similar to DCs activities	By Azuan Ahmed and et al. Ref [37] (2017)	**Adv:** 1) Could environment can be very beneficial for developing DCA basic functionalities that are challenge in client implementations due to needs to high memory capacity 2) A more sensitivity and TPR and PPV in long-term than other non-cloud based implementations **Dis:** KDDCup99 is very old	The successful of Dendritic Cell in protecting human body will also bring a success in protecting Cloud environment if the same mechanisms are being implemented in the real world applications
Survey	Intrusion Detection System using Artificial Immune System	To make an IDS with two lines of defense by mimicking the innate and adaptive immunity response of HIS so as to detect malicious files from network traffic	By Inadyuti Dutt and et al. Ref [31] (2016)	The T-cell and B-cell defensive mechanisms are used to detect the vulnerability of the files. The results exhibit that the proposed methodology works efficiently for detecting intrusion after inducing malicious attacks on the host-based system	
	Immunological Approach for Intrusion Detection	Comparative analysis of the DCA with Binary NSA	By M Zekri and L Souici-Mesati. Ref [21] (2014)	**Adv:** DCA as a winner of evaluations needs to be improved in future	
Mixed	Toward an artificial immune server against cyber attacks	To provide a framework to make an artificial immune server by mimicking two innate and adaptive immunity responses to detect cyber-attacks and to obviate unknown vulnerabilities	By Takeshi Okamato and Mitsubishi Tarao. Ref [3] (2016)	**Adv:** Drawing a good road map for producing an immune inspired hybrid prototypes	A cyber IS must be Big Data ready to handle the high volumes and speed at which data traffic is generated. This requires enough processing power to handle all the traffic.
	Cyber Immunity A Bio-Inspired Cyber Defense System	Comparing the BIS with AIS from aspects of their potentials so as to make an immune-inspired cyber-defense system	By Peter Wlodarczak. Ref [48] (2017)		

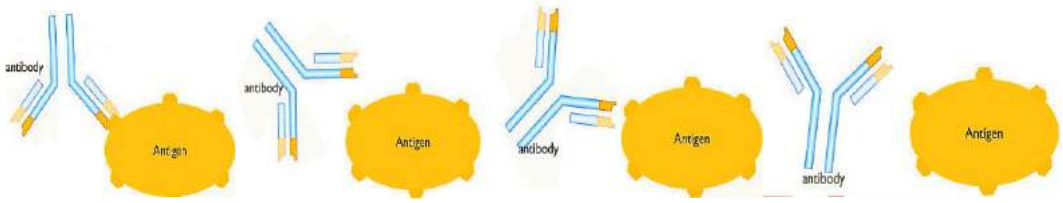
Table. Pros and Cons about some proposed AIS based intrusion detection approaches in recent years

واژه نامه ی تخّصصیِ ایمنی شناسی

به منظور گردآوریِ این واژه نامه از مرجع [۵۷] اقتباس شده است.

آنتی بادی : پادتن ، نوعی پروتئین است که در دستگاه ایمنی بدن به منظور مقابله با انتی ژنها تولید می شود. بدین صورت که پس از نفوذ آنتی ژن خاص این پروتئینها تولید شده و در خون به گردش در می آیند یا در محل باقی می مانند تا به آنتی ژنِ خاصِ مورد نظرِ خود بچسبند و خطر آلودگی آن را از بین ببرند. بنابراین هر آنتی بادی، یک نوع آنتی ژنِ خاص را هدف قرار می دهد.

هر آنتی بادی، یک نوع آنتی ژن خاص را هدف قرار می دهد. این به این دلیل است که شکل آنتی بادیها به صورت V انگلیسی است که دوشاخک دارند بطوریکه سر این شاخک ها به صورت فرو رفته به شکل قالبهایی هستند که در -اشکال مختلفی می توانند داشته باشند. هر نوع آنتی بادی بوسیله این شاخک ها بر روی نوع خاصی از آنتی ژن که پُرزهای برآمده بر روی سطح آن قالب این شاخک باشد متصل می شود. با این عمل، آنتی بادیها دراثر ترکیب با آنتی ژن آن را غیر فعال نموده و یا از انتشار آن در بدن جلوگیری می کنند. شکل زیر مراحل پیوند یک نوع آنتی بادی خاص را با آنتی ژن خاص نشان می دهد.



شکل ۳۳ – فرایند بایند شدنِ آنتی بادی با سطح آنتی ژن - از چپ به راست

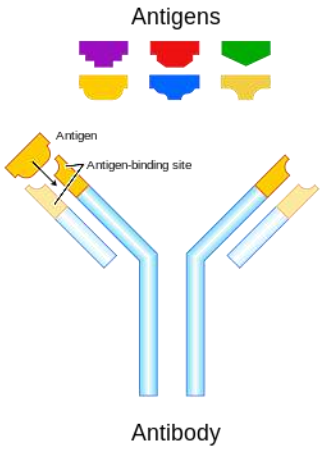
بنابراین وقتی چندین آنتی بادی از هر طرف با دو شاخه یشان بر روی سطح آنتی ژن قرار گرفته و قالب آن می شوند ، عملاً فعالیت آنتی ژن مختل می شود و نمی تواند آزادانه برای خودش در بدن بچرخد.

آنتی ژن : درسطح بیولوژیکی و مولکولی به باکتریها و ویروسها ، آنتی ژن می گویند. آنتی ژنها دارای ترکیباتی از پروتئینها و مواد شیمیایی هستند که پس از نفوذ به بافت های بدن موجب برانگیختن و واکنش سیستم ایمنی می شوند ، بطوریکه بدن انسان آنها را بیگانه تشخیص داده و با ترشح آنتی بادیها با آنها مقابله می کند.

البته ذکر این نکته نیز حائز اهمیت است که همه آنتی ژنها مضرر نیستند بلکه تعدادی از آنها به عنوان گونه های خودی در داخل بدن فعالیت می کنند و در کار نظارت در شناسایی نفوذ به سیستم دفاعی بدن کمک می کنند. اما دسته دیگر که آنتی ژنهای غیر خودی نیز شناخته می شوند بوسیله پاتوژنها از خارج از بدن تولید و

به اهداف خود در بدن هدایت می شوند. بنابراین من بعد هر جا در این گزارش مقصود از آنتی ژن همان غیر خودی است.

در واقع آنتی ژن های غیر خودی در حکم عوامل نفوذی در بافت های بدن هستند در مواقع لزوم به سلولهای خودی حمله می کنند. به آنتی ژنها ، پادتن یا پادزا نیز می گویند.



بر روی سطح خارجی سولهای آنتی ژن معمولاً پرزهایی برآمده وجود دارد که پپتید نامیده می شوند. هر نوع آنتی ژن پپتید خاصی دارد بطوریکه تنها آنتی بادی های خاصی که دارای شاخکهایی با فرمت مشخصی باشند در این پُرزها فرو رفته و قالب آنها باشد می توانند به آنتی ژن بچسبند. به عبارت دیگر برای هر نوع آنتی ژن خاص، آنتی بادی مخصوص با آن در بدن تولید می شود. منشأ تولید آنتی ژنها ، می تواند عوامل خارجی میکروبی باشد یا حتی سلولهای ناشناس در داخل بدن (مثل سلولهای سرطانی) یا خود سلولهای طبیعی بدن (بیماریهای خود ایمن) که قبلاً آلوده شده اند.

اُلگوی آنتی ژنیک : عوامل میکروبی و نفوذی به بدن که اصطلاحاً آنها را عوامل غیر خودی گویند ، اغلب دارای الگویی بر سطح مولکولی شان هستند که این الگوها در صورتی که با الگوی سطح مولکولی آنتی بادیهای حافظه بایند گردند ، پتانسیل مقاومت سیستم ایمنی بدن در جهت شناسایی و کشف نفوذ بالاتر می رود.

انتخاب منفی : آنتی بادیها به منظور بلوغ و موثر بودن در فرایند تشخیص ، به محض تولید در مغز استخوان با الگوهای خودی موجود چک می شوند و در صورتی برای ورود به چرخه شناسایی سیستم ایمنی انتخاب می گردند که با الگوهای خودی واکنشی نشان نداده باشند.

انتخاب تکثیری : تشخیص دهنده ها/آنتی بادیها بسته به تعداد الگوهای غیر خودی که در طول عمر خود شناسایی/بایند نموده اند تکثیر شده و در حافظه سیستم باقی می مانند.

تیموس : غدّه تیموس در بدن وظیفه آموزش دادن سیستم ایمنی را بر عهده دارد و به سلولها و پروتئینهای ایمنی بدن می آموزد تا چگونه با نفوذ و نفوذگر (آنتی ژنها) مقابله کنند.

تشخیص دهنده های حافظه[1]: در اصل همان آنتی بادیهای بالغ شده هستند که در طول عمر خود با الگوهای مختلف آنتی ژنیک بایند شده اند. این آنتی بایهای موثر مورد نیاز سیستم بوده و سیستم ایمنی آنها را حفظ می کند تا در هنگام نفوذ عواملی غیر خودی با الگویی مشابه ، بتوانند اقدم موثری را در جهت شناسایی آنها به انجام رسانند.

[1] Memory B Cells which have memory mature antibodies on its surface

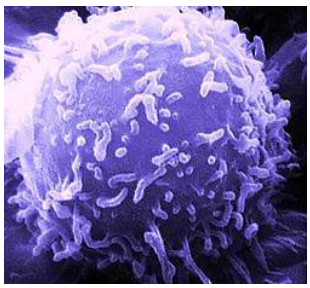
تطبیقِ الگو (بایند [1]) : اگر سطح مولکولی یک آنتی بادی با سطح مولکولی یک آنتی ژن واکنش نشان داده و بچسبد، به این عمل بایند شدن یا تطبیق الگو گویند. با تطبیق الگوی موفق ، آنتی ژن شناسایی میشود.

سیستم ایمنی زیستی (BIS) : در سطح مولکولی ، به سیستم ایمنی بدن موجودات زنده در طبیعت گویند. هدف این سیستم مقابله با پاتوژنها و عوامل آنها (آنتی ژنها) و تشخیص هر گونه رفتار غیر خودی در بدن جاندار می باشد که این نوع تشخیص یکی از مهمترین و اساسی ترین وظایف این سیستم است.

سلولهای دندریت : یاخته خواران/ بیگانه خواران یا APC سلولهایی هستند که در تمامی بافت های سلولی حضور دارند و پیوسته به سرکشی سلولها مشغولند. به محض وقوع نفوذ یا دریافت سیگنالهای خطر و مرگ سلولی از بافت ها، اقدام به پردازش سیگنالها و نمونه برداری از آنتی ژنها می نمایند. این سلولها به عنوان نخستین خط دفاعی بدن ، نقش موثری در کشف آسیب پذیری های سیستم و ایجاد زمینه برای شناسایی عوامل نفوذی و بررسی بیشتر الگوی آنها در سیستم ایمنی بدن ایفا می کنند.

عامل خودی : به عوامل بدن که الگوی خودی داشته و بدن خود آنها را تولید می نماید ، عامل با الگوی خودی گفته می شود. مثل آنتی بادیها ، نمونه هایی از آنتی ژنهای خودی و لنفوسیتها.

عامل غیر خودی : هر عامل پاتوژنیک که خارج از بدن تولید شده و بدن در تولید آنها نقشی نداشته ، عامل با الگوی غیر خودی قلمداد می گردد. آنتی ژنها معمولاً از این نوع هستند. آنتی ژنها در واقع مولکولهایی هستند که مخرب بوده و میکروب و باکتری هستند.



لنفوسیت [2] : لنفوسیت نوعی از گلبولهای سفید خون هستند که در سیستم ایمنی نقش اساسی دارند. سلولهای لنفوسیت به دو نوع B و T تقسیم می شوند. هر دو نوع در مغز استخوان جاندارتولیدمی شوند. نوع B در تولید پادتن (آنتی بادی) و ایمنی هومورال در بدن نقش دارد. این سلولها پس از تولید دوره ای را در محل تولید (مغز استخوان) می گذرانند که به "دوره تمایز" معروف است و در این مرحله تفاوت بین سلولهای خودی و غیر خودی (دشمن) و شناخت آنها را می آموزند! اما نوع T پس از تولید در مغز استخوان ، وارد غده تیموس شده و آنجا دوره تمایز خود را می گذراند. خود این لنفوسیت ها هر کدام انواعی دارند که در ادامه بدان اشاره می شود. هر لنفوسیت چه B و چه T فقط یک نوع گیرنده ی آنتی ژنی در سطح خود دارند ولی تعداد زیادی گیرنده ی آنتی ژنی از یک نوع دارند. لنفوسیتها، عمل فاگوسیتوز (یا بیگانه خواری، واکنش ذاتی) مانند یاخته خواران انجام نمی دهند و صرفاً در دفاع اختصاصی (تطبیق پذیر) نقش دارند.

[1] Pattern matching (Binding)
[2] lymphocyte

واژه نامه تخصّصیِ انگلیسی به فارسی و بالعکس

واژه انگلیسی	معادل	واژه فارسی	شرح
Innate immunity response		**مصونیّتِ ایمنی ذاتی**	سلولهای دندریت به سرعت در برابر هر گونه عامل خارجی و بیگانه واکنش نشان می دهند و نمونه های غیر خودی را قرنطینه می کنند. به این واکنش سریع که غالباً دارای خطای مثبت کاذب است پاسخ ایمنی ذاتی گویند.
Acquired immunity response		**مصونیّتِ ایمنی اکتسابی یا تطبیق پذیر**	لنفوسیتهای B و T دارای سازوکاری هستند که به آن پاسخ ایمنی اکتسابی یا تطبیق پذیر گویند.
Adaptive immunity response			
FtRNSA[6]	Future training real-valued negative selection algorithm	**الگوریتم انتخاب منفیِ حقیقی مبنا با یادگیری ثانویه**	
IO-RNSA[8]	Immune optimization based real-valued negative selection algorithm	**الگوریتم انتخاب منفیِ حقیقی مبنا مبتنی بر بهینه سازیِ مصون**	
DCA	Dendritic Cell Algorithm	**الگوریتم سلّول دِندریت**	
Artificial Life Cycle	AL	**سیکل حیات مصنوعی**	به هر بار اجرای سیستم تشخیص نفوذ پیشنهادی را یک سیکلِ حیات می گوییم.
HIS	Human Immune System	**سیستم ایمنیِ بدن انسان**	
AIS	Artificial Immune System	**سیستم ایمنیِ مصنوعی**	
DT	Danger Theory	**تئوریِ خطر**	یکی از متدهای ایمنی مصنوعی می باشد که بسطِ آن الگوریتم سلول دندریت می باشد. از دیدگاه ایمنی شناسی، بر سازوکار سلولهای دندریت نظامی حاکم است که با تئوری خطر بیان می شود.
Polar Coordinates		**مختصّات قطبی**	جهت ترسیم و تجسّم اشکال در ابعاد بالاتر نیازمند روابط مختصات قطبی هستیم.
Sure mode		**وضعیت قطعیت**	وضعیتی که در آن برجسب های اولیه و ثانویه آنتی ژن مربوطه در هر دو خط دفاعی یکسان باشند (یا هر دو آنومالی و یا هر دو نرمال)
Uncertain mode		**وضعیت عدم قطعیت**	وضعیتی که در آن برچسب اولیه و ثانویه هر آنتی ژن در خطوط دفاعی یکسان نباشند (برچسب اولیه نرمال و برچسب ثانویه آنومالی و یالعکس)
Hole		**حُفره**	به فضای پوشش داده نشده در فضای ابعاد مسئله گویند. این فضا می بایست توسط تشخیص دهنده های بالغ پوشش داده شود

Negative Selection mechanism		**سازوکار انتخاب منفی**	لنفوسیتهای B بالغ دارای این سازوکار هستند و با تولید و تکثیر آنتی بادیها با عوامل غیر خودی در بدن مقابله می کنند.
Clonal selection mechanism		**سازوکار انتخاب تکثیری**	
DC mechanism	**Dendritic cell mechanism**	**سازوکار سلولهای دندریت در بدن**	این سلولها دارای مصونیت ذاتی هستند و به سرعت در برابر عوامل میکروب یا آنتی ژنیک مقابله می کنند و گزارش آن را به لنفوسیتها می دهند تا فعال شوند.
Detector	**Antibody**	**آنتی بادی**	تشخیص دهنده
Mature detector	**Mature Ab**	**تشخیص دهنده بالغ**	تشخیص دهنده ها به محض تولید در مغز استخوان ، با عوامل خودی چک می شوند و در صورت عدم تطبیق با خودی ها ،بالغ شده و وارد بدن می شوند.
Mat effective detector	**Mat effective Ab**	**تشخیص دهنده های بالغ موثر**	اگر تشخیص دهنده با حداقل یک عامل غیر خودی تطبیق موفق انجام داد و آنرا شناسایی نمود ، در این صورت بدان آنتی بادی بالغ موثر گفته می شود.
Self-cell		**سلول خودی**	هر آنچه که در بافت های بدن وجود دارد به غیر از عواملی که از خارج وارد بدن می شوند خودی محسوب می شوند.
Non Self cell		**سلول غیر خودی**	انواع مختلف دارند و وارد بدن می شوند. آنها اغلب زیاد آور هستند هرچند آنتی ژنهای غیر خودی مفید هم داریم.
Memory detector	**Memory Ab**	**تشخیص دهنده حافظه**	تشخیص دهنده d هر چه بیشتر توانسته باشد با عوامل غیر خودی بایند گردد طول عمر آن بالا رفته و تبدیل به تشخیص دهنده حافظه می شود. بدن این دسته از تشخیص دهنده های بالغ را در حافظه خود نگه می دارد تا در صورت لزوم در آینده به عوامل مشابه غیر خودی پاسخ سریع دهد و زمان شناسایی به طور چشمگیری کاهش یابد.
Stability		**پایداری**	معیاری است وابسته به cc ضریب همبستگی که هرچه به ۱ نزدیکتر باشد بهتر است
Anomaly		**ناهنجاری**	

Signal		**سیگنال**	ابزار نمونه برداری از آنتی ژنها توسط سلولهای دندریت در طول یک پاسخ ایمنی ذاتی [1]
Sampling / Presenting		**نمونه برداری**	
Presenter DC		**سلول دندریت نمونه بردار**	
Intrusion Detection & Prevention System	**IDPS**	**سیستم تشخیص و جلوگیری از نفوذ**	
Information Security Management System	**ISMS**	**سیستم مدیریت امنیّتِ اطلاعات**	
Biological Immune System	**BIS**	**سیستم دفاعی ایمنی زیستی**	
National Security Agency	**NSA**	**آژانس امنیّت ملّی امریکا**	
Black Hole Algorithm	**BHA**	**الگوریتم خوشه بندی سیاه چاله**	
Real-valued negative selection algorithm	**RNSA**	**الگوریتم انتخاب منفیِ حقیقی مبنا**	
	Real Valued NS		
V-detector		**نسخه استاندارد و اوّلیه الگوریتم انتخاب منفی حقیقی مبنا**	
Nature Inspired Computations	**NIC**	**محاسبات الهام گرفته شده از طبیعت**	
Network based intrusion detection system	**NIDS**	**سیستم تشخیص نفوذ مبتنی بر شبکه**	
Host based intrusion detection system	**HIDS**	**سیستم تشخیص نفوذ مبتنی بر میزبان**	

[1] عوامل غیر خودی و آسیب رسان پس از ورود به بافت های بدن ، سلولها را مورد تخریب قرار می دهند و بدین ترتیب از سمت سلولهای آسیب دیده یا حتی سالم در هر لحظه سیگنالهایی مبنی بر اعلام وضعیت سلول منتشر می شود. حال در صورتی که سلول دندریتی در نزدیکی سلول قرار داشته باشد به سرعت سیگنال را دریافت نموده و باصطلاح نمونه برداری می نماید.