# طراحی یک سیستم ژنتیک فازی برای تشخیص نفوذ در شبکه های کامپیوتری

نرگس نوری[۱]، شیوا احمدی [۲]

[۱] دانشجوی کارشناسی ارشد مهندسی نرم افزار، دانشگاه آزاد اسلامی واحد الکترونیکی، تهران، ایران

[۲] دانشجوی کارشناسی ارشد مهندسی نرم افزار، دانشگاه آزاد اسلامی واحد الکترونیکی، تهران، ایران

## چکیده

**همزمان با رشد روزافزون سرویس های تحت شبکه، تهدیدهای امنیتی زیادی نیز در حال تولید هستند. همچنین از لحاظ تکنیکی ، سیستم فاقد هر گونه اسیب پذیری نمی توان ساخت، بنابراین سیستم تشخیص نفوذ (IDS) مورد توجه قرار گرفته است. هدف اصلی آن، دسته بندی فعالیتها به دو دسته اصلی است: (۱) فعالیتهای عادی (۲) فعالیت های نفوذی. بدلیل آنکه مرز بین این فعالیت ها را نمی توان به خوبی تعریف کرد، احتمال تولید هشدارهای نادرست بالا است. با منطق فازی این احتمال کاهش می یابد. این مقاله یک دسته بند ژنتیک فازی را برای تشخیص الگو در شبکه های کامپیوتری پیشنهاد می کند. قوانین تولید شده با منطق فازی با الگوریتم ژنتیک آموزش داده می شوند، در نهایت با این قوانین، الگوهای ورودی دسته بندی می شوند. کار اصلی این مقاله تولید قوانینی است که (۱) طول کوتاهی دارند (۲) ویژگی های الگو را بطور خودکار انتخاب می کنند (۳) نرخ تشخیص را بهبود می بخشند. نتایج آزمایشات انجام شده روی مجموعه داده KDDCUP99، بهبود نرخ تشخیص را نسبت به دیگر الگوریتم ها نشان می دهد.**

## کلمات کلیدی

**سیستم تشخیص نفوذ، منطق فازی، الگوریتم ژنتیک ، یادگیری ماشین**

## ۱- مقدمه

هر مجموعه فعالیتی که برای نقض[۱] یکپارچگی[۲] ، محرمانگی[۳] یا دسترس پذیری[۴] یک منبع تلاش می کند، نفوذ[۵] نامیده می شود [۱]. بنابراین برای محافظت سیستم از نفوذگران، یک سیستم تشخیص نفوذ[۶] نیاز است. IDS، با بکارگیری قوانین معین، دسترسی کاربر به سیستم کامپیوتری را نظارت و محدود می کند. این قوانین مبتنی بر دانش تخصصی مدیران حرفه ای که سناریو های حمله را می سازند، می باشد. این سیستم نفوذ های کاربر را شناسایی می کند و اقدام لازم برای جلوگیری از حمله روی پایگاه داده را اتخاذ می کند یا پیشنهاد می دهد [۲].

روشهای تشخیص نفوذ به دو دسته اصلی تقسیم می شوند: تشخیص سوء استفاده (الگو)[۷] و تشخیص ناهنجاری[۸]. تشخیص ناهنجاری مبتنی بر رفتارهای نرمال کاربران است و هر گونه انحراف از این رفتارها به عنوان نفوذ تشخیص داده می شود. روشهای آماری [۳]، سیستم های خبره [۴]، و شبکه های عصبی [۵] روشهایی در این زمینه هستند. در این روش، حملات جدید قابل شناسایی است اما نرخ

---

۱ compromise
۲ Integrity
۳ Confidentiality
۴ Availability
۵ Intrusion
۶ Intrusion Detection System (IDS)
۷ Misuse Detection
۸ Anomaly Detection

تولید هشدار اشتباه بالا است. روش تشخیص الگو، حملاتی که قبلا امضاء یا الگوی آنها برای سیستم تعریف شده است را می تواند شناسایی کند. این روش توانایی شناسایی حملات جدید (شناخته نشده) را ندارد اما نرخ هشدار اشتباه پایین دارد. روشهای سیستم خبره [۶]، الگوریتم ژنتیک[۷] و روشهای تطبیق الگو برای این روش بکار رفته اند.

یکی از رایج ترین روشها، ترکیب منطق فازی با الگوریتم ژنتیک است که منجر به سیستم های ژنتیک فازی[۹] می شود. GFS یک سیستم فازی است که با یک فرایند یادگیری مبتنی بر محاسبات تکاملی، مانند الگوریتم ژنتیک و دیگر الگوریتم های تکاملی، اموزش داده می شود.

برای حل مسئله تشخیص نفوذ، مطالعات مختلفی برای بدست اوردن سیستم دسته بند انجام شده است. در [۲] با طراحی و تحلیل انواع سیستم های ژنتیک فازی برای تشخیص الگو، قابلیت سیستم های ژنتیک فازی در رابطه با تشخیص نفوذ بررسی می شود. در [۸] قوانین فازی با استفاده از الگوریتم بهینه سازی اجتماع ذرات اموزش داده می شوند. در [۹] یک روش با استفاده از قوانین انجمنی به همراه منطق فازی ارائه شده است. در [۱۰] یک دسته بند ژنتیک فازی با تنها یک قانون برای دسته بندی الگو تولید می شود و هر دو تشخیص ناهنجاری و الگو را حمایت می کند.

همانطور که در این مرور کوتاه ملاحظه کردید، توجه کمی روی GFS ها برای IDS شده است. هدف اصلی این مقاله ارائه یک دسته بند ژنتیک فازی و مقایسه آن با سایر الگوریتم ها است. ساختار این مقاله به صورت زیر است. در بخش ۲ پیش پردازش داده بیان می گردد، بخش ۳، منطق فازی و بخش ۴ الگوریتم ژنتیک و تابع ارزیاب و انواع عملگرهای آنرا بیان می نماید. بخش ۵ دسته بند پیشنهادی را توضیح می دهد. در بخش ۶ نتایج آزمایشگاهی و در بخش ۷ نتیجه گیری اورده شده است.

## ۲- پیش پردازش داده

برای ارزیابی IDS پیشنهادی از مجموعه داده KDDCUP99[۱۰] استفاده شده است[۱۱] که شامل رکوردهای TCP/IP با ۴۱ خصوصیت برای هر رکورد می باشد. این خصوصیات از نوع عددی و الفبایی هستند. پیش از استفاده از مجموعه داده، باید آنها به مقادیر بین صفر ویک تبدیل شوند. هر مقدار عددی توسط (۱) نرمال می شود

$$y = \frac{x - Min}{Max - Min} \qquad (۱)$$

$x$ : مقدار عددی خصوصیت است.

$Min$ : حداقل مقداری که این خصوصیت در همه رکوردها دارد.

$Max$ : حداکثر مقداری که این خصوصیت در همه رکوردها دارد.

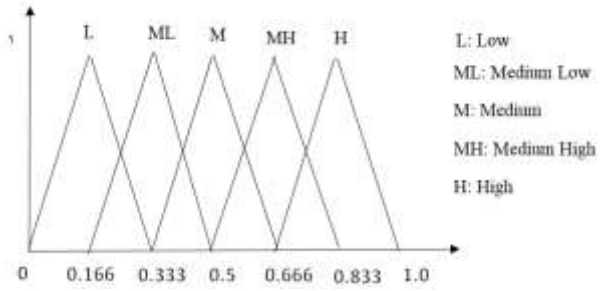
$y$ : مقدار عددی که $0 \le y \le 1$

## ۳- منطق فازی

تئوری مجموعه‌های فازی و منطق فازی را اولین بار پرفسور لطفی‌زاده در سال ۱۹۶۵ معرفی نمود. بنیاد منطق فازی بر شالوده نظریه مجموعه‌های فازی استوار است. این نظریه تعمیمی از نظریه کلاسیک مجموعه‌ها در علم ریاضیات است. در تئوری کلاسیک مجموعه‌ها، یک عنصر، یا عضو مجموعه است یا نیست. در حقیقت عضویت عناصر از یک الگوی صفر و یک و باینری تبعیت می‌کند. اما تئوری مجموعه‌های فازی این مفهوم را بسط می‌دهد و عضویت درجه‌بندی شده را مطرح می‌کند. به این ترتیب که یک عنصر می‌تواند تا درجاتی - و نه کاملاً - عضو یک مجموعه باشد.

درجه عضویت یک شی به مجموعه فازی با تابع عضویت تعیین می گردد. دامنه این تابع محدوده مقادیر شی و خروجی ان در بازه [۰...۱] است.

فضای فازی برای مقادیر عددی نرمال شده مجموعه داده در شکل (۱) نمایش داده می شود. خصوصیات الفبایی با قراردادن مقدار تصادفی صفر یا یک نرمال می شوند.




**شکل (۱) : توابع عضویت و فضای فازی برای خصوصیات عددی مجموعه داده**

## ۴- الگوریتم ژنتیک

الگوریتم ژنتیک نوع خاصی از الگوریتمهای تکامل است که از تکنیکهای زیست‌شناسی مانند وراثت و جهش استفاده می‌کند. این الگوریتم برای اولین بار توسط جان هلند معرفی شد.در واقع الگوریتم‌های ژنتیک از اصول انتخاب طبیعی داروین برای یافتن فرمول بهینه جهت پیش‌بینی یا تطبیق الگو استفاده می‌کنند.

در GA، یک جمعیت شامل کروموزوم ها تولید می شود. هر کروموزوم بیانگر یک راه حل ممکن برای مسئله است. این کروموزوم ها با استفاده از عملگرهای ژنتیک (انتخاب، برش،جهش) تکامل می یابند. شکل (۲) نمای کلی از سیستم پیشنهادی را نشان می دهد. این سیستم برای هر یک از انواع الگو به صورت جداگانه اجرا می شود.

---

[۹] Genetic Fuzzy Systems

[۱۰] Knowledge Discovery in Database

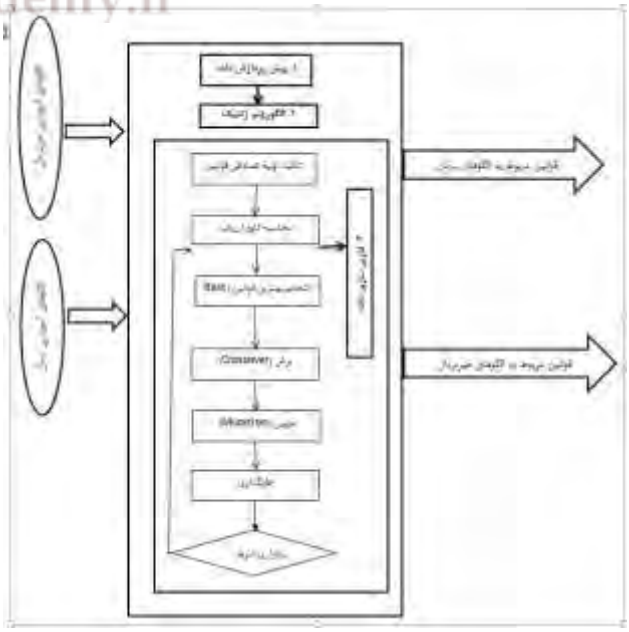
### ۴-۱- کدینگ کروموزوم

هر قانون به صورت یک کروموزوم نمایش داده می شود. گرامر زیر برای ساخت قسمت شرط قانون استفاده می شود.

1) <condition> ::=
< atomic_cond > < operator > < condition > |
< atomic_cond > < operator > < atomic_cond>

(2) <atomic_cond> ::==
<variable> <rel op> <set>





**شکل (۲): نمای کلی از سیستم در مرحله آموزش**

(3) <operator>::= $\vee \mid \wedge$
(4) <variable> ::== $x_1 \mid \ldots \mid x_n$
(5) <rel op> ::= $\in \mid \notin$
(6) <set> ::= L|ML|M|M|MH|H

یک کروموزوم مجموعه ای از n ژن است. هر ژن از یک شرط اتمی و عملگر فازی تشکیل شده است. ژن اخر فقط شامل شرط اتمی است. شکل (۳) یک کروموزم را نشان می دهد. مقدار n که بیانگر ژن ها یا در واقع تعداد خصوصیات است، به صورت تصادفی انتخاب می گردد. در واقع $2 < n \leq 41$.

| Gene1 | | | | | ... | Gene n | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Atomic condition 1 | | | | Operator 1 | ... | Atomic condition n | | | | Operator n |
| Attno 1 | Var 1 | Ro 1 | Set 1 | Op 1 | ... | Attno n | Var n | Ro n | Set n | Op n |

**شکل (۳) : نمایش کروموزوم**

هر یک از ژنها شامل مقادیر زیر هستند:

- بخش شرط اتمیک:

Attno: شماره خصوصیت به صورت تصادفی از 1...41 انتخاب می شود.

Var: مقدار خصوصیت

Ro: عملگر تعلق به صورت تصادفی از ($\notin$ و $\in$) انتخاب می شود.

Set: مجموعه فازی که به صورت تصادفی از ( low,Medium Low,Medium, Medium High, High برای خصوصیات عددی و cr1 و cr2 برای خصوصیات حروفی) انتخاب می شود.

- بخش عملگر :

Op : عملگر منطقی که به صورت تصادفی از ($\vee, \wedge$) انتخاب می شود. الویت عملگرها از چپ به راست است.

### ۴-۲ جمعیت اولیه

جمعیت اولیه به صورت تصادفی تولید می شود که شامل مجموعه ای از کروموزوم ها است که هر کروموزم تعداد متفاوتی ژن دارد.

### ۴-۳ تابع ارزیاب

تابع ارزیاب، درجه هر کروموزوم در جمعیت فعلی را معین می کند. معیارهای زیر در مسئله دسته بندی نفوذ در محاسبه تابع ارزیاب استفاده می شوند.

**جدول (۱):معیارهای مطرح در دسته بندی نفوذ**

| نوع خروجی | نوع پیش بینی شده | نوع واقعی |
|---|---|---|
| مثبت صحیح (TP) | نرمال | نرمال |
| مثبت نادرست (FP) | حمله | نرمال |
| منفی صحیح (TN) | حمله | حمله |
| منفی نادرست (FN) | نرمال | حمله |

مقدار تابع ارزیاب طبق (۲) محاسبه می شود [۱۰].

(۲)

$$TP = \sum_{i=1}^{p} Fuzzy(normal\_data_i)$$

$$TN = \sum_{i=1}^{q} [1 - Fuzzy(abnormal_{data_i})]$$

$$FP = \sum_{i=1}^{q} Fuzzy(abnormal\_data_i)$$

$$FN = \sum_{i=1}^{p} [1 - Fuzzy(normal\_data_i)$$

$$sensitivity = \frac{TP}{TP + FN} \quad , \quad specificity = \frac{TN}{TN + FP}$$

$$length = 1 - \frac{chrom\_length}{100},$$

$$fitness = w_1 * sensitivity + w_2 * specificity + w_3 * length$$

- FP، TN، TP و FN به ترتیب مثبت صحیح، منفی صحیح، مثبت نادرست و منفی نادرست می باشند.
- Fuzzy ، مقدار فازی قسمت شرط قانون است.
- P و q به ترتیب تعداد نمونه های نرمال و حمله در مجموعه داده آموزشی است.
- $w_1$ ، $w_2$ و $w_3$: وزن های اختصاص داده شده به هر قانون هستند که می تواند به صورت تصادفی مقداردهی شوند یا مقادیر ثابت داشته باشند. در این مقاله از مقادیر ثابت $w_1 = 0.45$ ،$w_2 = 0.45$ و $w_3 = 0.1$ استفاده می کنیم.
- $normal\_data_i$ : زیرمجموعه الگوهای اموزشی نرمال است.
- $abnormal\_data_i$ : زیرمجموعه الگوهای اموزشی حمله است.

برای محاسبه تابع ارزیاب برای قوانین تعیین کننده حمله، کافی است در فرمول بالا جای normal را با abnormal عوض می کنیم.

### ۴-۴ انتخاب بهترین ها

به علت اینکه برای تولید نسل بعدی ممکن است کروموزوم های بهینه نسل فعلی از بین روند، ابتدا کروموژوم های بهینه را در نسل جدید کپی می کنیم.

### ۴-۵ انتخاب

کروموزوم ها ی پدر برای تولید فرزند بر اساس انتخاب tournament انتخاب می شوند. در این نوع انتخاب چند مجموعه از کروموزوم انتخاب می گردد و سپس بهترین کروموزوم های هر مجموعه بر اساس تابع ارزیاب انتخاب می گردند.

### ۴-۶ برش

قابلیت ترکیب دو کروموزوم پدر و تغییر برخی از ژن های آنها، برش نامیده میشود.بر اساس احتمال pc برش اعمال می گردد.برای اعمال برش، نقطه برش تعیین می گردد که برابر شماره ژن است.تعداد ژن ها در دو کروموزوم ، بعد از برش بین دو و حداقل طول دو کروموزوم پدر هستند. در اینجا ما از برش تک نقطه ای استفاده کردیم.

### ۴-۷ جهش

جهش تغییر در مقادیر ژن های یک کروموزوم است. پس از برش، با احتمال pm جهش روی کروموزوم های فرزند اعمال می گردد.یک ژن از کروموزوم به صورت تصادفی انتخاب می گردد و مقادیر آن ژن به صورت تصادفی مقداردهی می شود.

### ۴-۸ جایگذاری

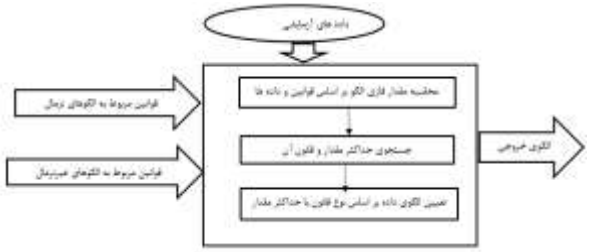
پس از تولید فرزندان، انها بهمراه کروکوزوم های بهینه از قبل انتخاب شده در نسل بعدی قرار می گیرند.

### ۴-۹ شرط پایان

شرط پایان در این کار برای تعداد ثابت تکرار الگوریتم می باشد.

## ۵- سیستم دسته بند ژنتیک فازی

پس از اموزش سیستم با داده های اموزشی نرمال و حمله، مجموعه قوانین مربوط به هریک از از این نوع الگوها حاصل می شود. حال در مرحله آزمایش، این قوانین را بر مجموعه داده آزمایشی اعمال می کنیم. هر قانونی که مقدار فازی بیشتری داشته باشد، به عنوان دسته الگو شناخته می شود. همانطور که در شکل (۴) نمایش داده می شود.




**شکل (۴) : نمای دسته بند**

## ۶- نتایج آزمایشگاهی

آزمایشات روی ده درصد از مجموعه داده KDDCUP99 [۱۱] و روی لپ تاپ با پردازنده corei7 و حافظه GB٤ انجام شده است. نتایج در جدول (۲) آورده شده است.

| | نرخ تشخیص | نرخ هشدار اشتباه |
|---|---|---|
| الگوی نرمال | ۹۵,۰۱ | ۰,۰۴ |
| الگوی حمله | ۹۴,۹۶ | ۰,۰۳۶۷ |

**جدول (۲): نتایج آزمایشگاهی**

## ۷- نتیجه گیری

در این مقاله ما یک سیستم دسته بند ژنتیک فازی را توسعه داده ایم که قادر است قوانینی با طول و تعداد خصوصیات متغیر تولید کند. سیستم پیشنهادی ، می تواند نوع الگوی ورودی را به صورت حمله یا نرمال دسته بندی کند. نرخ تشخیص بدست آمده برای هر دو نوع الگوها (۹۵ % برای الگوی نرمال و۹۴% برای الگوهای حمله) در مقایسه با دیگر الگوریتم های ژنتیک فازی، نتیجه خوبی است.

## مراجع

[1] Heady, R., Luger, G., Maccabe, A., & Servilla, M.,"The architecture of network level intrusion detection system", Technical Report. Department of Computer Science, University of New Mexico (1990).

[2] Abadeh, M. S., H. Mohamadi and J. Habibi, "Design and analysis of genetic fuzzy systems for intrusion detection in computer networks", Expert Syst. Appl. **38**(6), (2011) 7067-7075.

[3] Javitz, H.S., Valdes, A., Lunt, T.F., Tamaru, A., Tyson,M.,Lowrance, J., " Next generation intrusion detection expert system (NIDES). Technical Report A016-Rationales, SRI (1993).

[4] Vaccaro, H.S., Liepins, G.E., "Detection of anomalous computer session activity", In:Proceedings of the IEEE Symposium on Security and Privacy (1989).

[5] Debar, H., Becker, M., Siboni, D., "A neural network component for an intrusion detection system. In: Proceedings of the IEEE Symposium of Research in Computer Security and Privacy, pp. 240–250 (1992)
[6] Lunt, T.F., Jagannathan, R.: A prototype real-time intrusion-detection expert system.In: Proceedings of the IEEE Symposium on Security and Privacy, pp. 59–66 (1988)

[7] Ludovic, M.: GASSATA, A genetic algorithmas an alternative tool for security audit trails analysis. In: First International Workshop on the Recent Advances in Intrusion Detection (1998)

[8] Feng Hsuan-Ming, 'Particle Swarm Optimization Learning Fuzzy Systems Design', in Information Technology and Applications, 2005. ICITA 2005. Third International Conference on (2005), pp. 363-66 vol.1.

[9] Arman Tajbakhsh, Mohammad Rahmati, and Abdolreza Mirzaei, Intrusion Detection Using Fuzzy Association Rules', Applied Soft Computing, 9 (2009), 462-69.

[10] S. M. Al Naqshbandi, and V. W. Samawi, 'One-Rule Genetic-Fuzzy Classifier', in Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on (2012), pp. 204-08.

[11] KDD-Cup data set: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.