

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

١٤٢٣



دانشگاه شهید بهشتی
دانشکده مهندسی برق و کامپیوتر

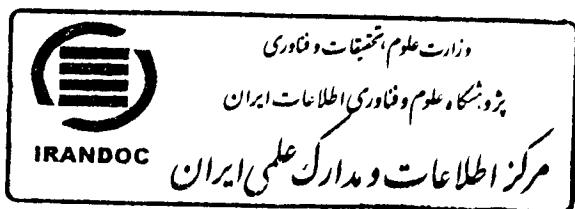
عنوان پایان نامه
شناسایی نفوذگر بر مبنای رفتار غیر متعارف با استفاده از روشهای داده کاوی و روشهای
ترکیبی یادگیری ماشین

پایان نامه کارشناسی ارشد مهندسی کامپیوتر
گرایش هوش مصنوعی

استاد راهنما:
جناب آقای دکتر مقصود عباسپور

توسط:
فرزانه گرامی راز

تابستان ۱۳۸۹



۱۴۹۳۲۳

۱۳۸۹/۱۰/۱۹

پایان نامه کارشناسی ارشد مهندسی کامپیوتر - گرایش هوش مصنوعی
تحت عنوان:

شناسایی نفوذگر بر مبنای رفتار غیر متعارف با استفاده از روشهای داده کاوی و روشهای
ترکیبی یادگیری ماشین

پایان نامه دانشجو، فرزانه گرامی راز، توسط کمیته تخصصی داوران مورد بررسی و تصویب

در تاریخ ۱۳۸۹/۶/۲۴
نهایی قرار گرفت.

امضاء
امضاء
فرزانه راز
امضاء
امضاء

جناب آقای دکتر مقصود عباسپور
جناب آقای دکتر فرشاد صفایی
جناب آقای دکتر احمد خونساری
جناب آقای دکتر اسلام ناظمی

۱- استاد راهنما اول:
۲- استاد داور (داخلی)
۵- استاد داور (خارجی)
۶- نماینده تحصیلات تکمیلی

تشکر و قدردانی

ستایش و سپاس خداوند را، آنکه پیوند دهنده‌ی ستایش است به نعمتها و پیوند دهنده‌ی نعمتهاست به سپاس. اکنون که به یاری ایزد یکتا توفیق تکمیل این رساله تحقیقاتی را یافته‌ام، بر خود لازم می‌دانم از زحمات و راهنماییهای جناب آقای دکتر مقصود عباسپور-استاد راهنمای پروژه- که در تمام مدت انجام پروژه صبورانه و دلسوزانه در کنارم بودند کمال تقدیر و تشکر را به عمل آورم.

کلیه حقوق مادی مترتب بر نتایج مطالعات،
ابتکارات و نوآوریهای ناشی از تحقیق موضوع
این پایان نامه متعلق به دانشگاه شهید بهشتی
می باشد.

نام و نام خانوادگی: فرزانه گرامی راز
عنوان پایان نامه: شناسایی نفوذگر بر مبنای رفتار غیر متعارف با استفاده از روشهای داده کاوی و
روشهای ترکیبی یادگیری ماشین
استاد/اساتید راهنما: جناب آقای دکتر مقصود عباسپور

اینجانب فرزانه گرامی راز تهیه کننده پایان نامه کارشناسی ارشد/دکتری حاضر خود را ملزم به حفظ
امانت داری و قدردانی از زحمات سایر محققین و نویسندگان بنا بر قانون Copyright می دانم. بدین
وسیله اعلام می نمایم که مسئولیت کلیه مطالب درج شده با اینجانب می باشد و در صورت استفاده از
اشکال؛ جداول، و مطالب سایر منابع، بلافاصله مرجع آن ذکر شده و سایر مطالب از کار تحقیقاتی اینجانب
استخراج گشته است و امانتداری را به صورت کامل رعایت نموده ام. در صورتی که خلاف این مطلب ثابت
شود، مسئولیت کلیه عواقب قانونی با شخص اینجانب می باشد.

نام و نام خانوادگی دانشجو: فرزانه گرامی راز

امضاء و تاریخ:

تقدیم به آنان که می‌بینند، می‌شنوند و در راه اهداف و اندیشه‌های
نیکشان تا پای نتیجه پای‌بندند.

تقدیم به پدر و مادر عزیزم.

فهرست مطالب:

۱	فصل اول: معرفی پروژه.....
۲	۱-۱ اهمیت موضوع:.....
۳	۲-۱ معرفی پروژه:.....
۷	فصل دوم: پایگاه داده استفاده شده NCL.....
۸	۱-۲ پایگاه داده KDD-99.....
۱۱	۲-۲ بررسی مجموعه داده KDD-99.....
۱۱	۳-۲ مشکلات ذاتی مجموعه داده KDD-99.....
۱۲	۴-۲ مجموعه داده‌های جدید:.....
۱۵	فصل سوم: مروری بر معماری‌های ارائه شده.....
۱۶	۱-۳ مقدمه.....
۱۶	۲-۳ تاریخچه سیستمهای شناسایی نفوذ.....
۱۸	۲-۳ تقسیم بندی روشهای شناسایی نفوذ.....
۲۱	۳-۳ آشنایی با تعدادی از سیستمهای تشخیص ناهنجاری.....
۲۲	۱-۳-۳ مدل‌های ایستا.....
۲۲	۱-۳-۳-۱ روشهای آماری.....
۲۵	۲-۳-۳-۱ روش‌های یادگیری ماشین.....
۲۸	۳-۳-۳-۱ روشهای داده‌کاوی در تشخیص ناهنجاری.....
۳۲	۴-۳ مروری بر سیستمهای شناساگر ناهنجاری وفقی.....
۳۳	۱-۴-۳ به‌روز رسانی مدل.....
۳۳	۱-۴-۳-۱ سیستم شناساگر نفوذ با قابلیت به‌روز رسانی اتوماتیک.....
۳۹	۲-۴-۳-۱ چارچوبی برای شناساگر نفوذ وفقی بر اساس بردارهای پشتیبانی توصیف داده.....
۴۲	۳-۴-۳-۱ شناساگر نفوذی وفقی بر اساس خوشه‌بندی و بر مبنای روش هسته.....

- ۴۵ ۴-۱-۴-۲ استفاده از تشدید تطبیقی فازی در شناسایی ناهنجاری وفقی شبکه
- ۴۶ ۴-۱-۴-۳ چارچوبی برای تشخیص ناهنجاری وفقی با استفاده از داده‌کاوی فازی
- ۴۷ ۴-۱-۴-۳ یادگیری سیستم شناساگر نفوذ با استفاده از الگوریتم بیزین وفقی
- ۴۹ ۴-۲-۲ جایگزینی زیرمدل جدید به جای مدل قدیم
- ۴۹ ۴-۲-۱ شناساگر بلادرنگ نفوذی با استفاده از روش داده‌کاوی
- ۵۱ ۴-۲-۲ چارچوبی برای تشخیص نفوذ با استفاده از شبکه بیزین
- ۵۳ ۴-۳ مدل‌های ترکیبی ارائه شده
- ۵۳ ۴-۳-۱ مدل ترکیبی افزایشی با استفاده‌ی گروهی از کلاس‌بندهای ضعیف
- ۵۶ ۴-۳-۲ سیستم شناساگر نفوذ وفقی با استفاده از تکامل اتصالات سیستمی
- ۵۸ ۴-۳-۳ سامانه تطبیقی یادگیری امضاء با استفاده از الگوریتم ژنتیک در تشخیص نفوذ
- ۵۹ ۳-۵ چالشهای پیش ر:
- 63 فصل چهارم: مروری بر معماریهای ارائه شده
- ۶۴ ۴-۱ آشنایی با معماری ارائه شده:
- ۶۶ ۴-۲ پیش‌نیازهای پروژه
- ۶۶ ۴-۲-۱ منطق فازی
- ۶۷ ۴-۲-۱-۱ مجموعه‌های فازی
- ۶۷ ۴-۲-۱-۲ قوانین فازی
- ۶۸ ۴-۲-۱-۲ کلاس بند فازی
- ۶۹ ۴-۲-۲ آشنایی با FCM
- ۷۲ ۴-۲-۳ الگوریتم ژنتیک
- ۷۲ ۴-۳ توصیف کلی معماری ارائه شده:
- ۷۴ ۴-۳-۱ معماری ارائه شده به جزئیات

۷۴ ۱-۱-۳-۴ مولد مدل تشخیص
۷۸ ۲-۱-۳-۴ موتور تجزیه و تحلیل
۷۹ ۳-۱-۳-۴ بافرها
۷۹ ۴-۱-۳-۴ میزان کننده فازی مدل
۸۱ ۲-۳-۴ نتیجه گیری: صورت کلی نحوه به روز رسانی
82 فصل پنجم: نتایج شبیه سازی
۸۳ ۱-۵ مقدمه:
۸۳ ۲-۵ مقایسه روش ارائه شده با سایر روشهای یادگیری ماشین:
۸۳ ۱-۲-۵ مقایسه با سایر روشهای یادگیری ماشین
۸۶ ۳-۵ بررسی عوامل مؤثر در کارایی سیستم ارائه شده
۸۷ ۱-۳-۵ میزان کننده فازی مدل (کنترل کننده فازی)
۸۷ ۱-۱-۳-۵ شیب توابع عضویت خروجی سیستم
۸۸ ۲-۱-۳-۵ افزودن قوانین تشویقی
۸۹ ۳-۱-۳-۵ تأثیر تأخیر در به روز رسانی سیستم بر دقت نهایی
۹۱ ۲-۳-۵ بررسی نحوه توزیع رکوردها بر اساس سختی در هر یک از دسته ها
۹۳ ۳-۳-۵ نتیجه گیری
۹۴ فصل ششم: نتیجه گیری و کارهای آیند
۹۵ ۱-۶ خلاصه پروژه ارائه شده
۹۷ ۲-۶ کارهای آینده
۱۰۰ مراجع:
۱۰۶ خلاصه به انگلیسی:

فهرست جداول

- جدول ۱-۲ ویژگیهای پایه برای هر ارتباط به صورت انفرادی ۸
- جدول ۲-۲ ویژگیهای محتوایی برای اتصالات بر اساس دانش قلمرو ۸
- جدول ۳-۲ ویژگیهای ترافیکی با در نظر گرفتن پنجره ی دو ثانیه ای ۹
- جدول ۴-۲ پراکندگی داده‌های هر دسته در مجموعه داده آموزش ۱۱
- جدول ۵-۲ پراکندگی داده‌های هر دسته در مجموعه داده تست ۱۱
- جدول ۱-۳ دقت تشخیص سیستم ارائه شده توسط یو و همکارانش ۳۷
- جدول ۲-۳ دقت تشخیص سیستم ارائه شده توسط یانگ ۳۹
- جدول ۳-۳ مقایسه روش ارائه شده توسط لی و همکارانش با سایر روشها ۴۰
- جدول ۴-۳ نرخ تشخیص سیستم ارائه شده با استفاده از F-ART ۴۱
- جدول ۵-۳ نرخ تشخیص سیستم ارائه شده توسط ژیانگ و همکارانش ۴۲
- جدول ۶-۳ مقایسه روش ارائه شده توسط فرید و همکارانش با سایر روشها ۴۴
- جدول ۷-۳ درجه تشیص سیستم ارائه شده توسط جلیلی برای تشخیص نفوذ ۴۸
- جدول ۸-۳ درجه تشیص سیستم ارائه شده توسط جلیلی برای تشخیص انواع نفوذ ۴۸
- جدول ۹-۳ درجه تشیص سیستم ارائه شده توسط جلیلی برای تشخیص انواع نفوذ ۵۷
- جدول ۱-۵ مقایسه دقت سیستم [۷۱] با معماری ارائه شده ۸۵
- جدول ۲-۵ مقایسه دقت سیستم [۷۲] با مقدار ثابت $p=0.5$ با معماری ارائه شده ۸۶
- جدول ۳-۵ مقایسه دقت سیستم [۷۲] با مقدار متغیر p با معماری ارائه شده ۸۶
- جدول ۴-۵ تأثیر تغییر شیب توابع عضویت خروجی سیستم بر دقت ۸۸
- جدول ۵-۵ تأثیر وجود تشویق در میزان کارایی سیستم ۸۹
- جدول ۶-۵ نمایش همزمان تأثیر تغییرشیب و وجود تشویق در کارایی کل ۸۹

- جدول ۷-۵ تأثیر تأخیر در شبیه‌های مختلف بر دقت نهایی در سیستم بدون تشویق ۹۰
- جدول ۸-۵ تأثیر تأخیر در شبیه‌های مختلف بر پارامترهای کارایی (همراه تشویق) ۹۰
- جدول ۹-۵ تأثیر تأخیر در شبیه‌های مختلف بر پارامترهای کارایی (بدون تشویق) ۹۱
- جدول ۱۰-۵ پراکندگی نمونه‌ها در دسته‌های مختلف و معیارهای مرتبط ۹۱
- جدول ۱۱-۵ پراکندگی نمونه‌ها در دسته‌های مختلف و معیارهای مرتبط ۹۲

فهرست شکلها:

- شکل ۱-۳ سیستم عمومی آشکار ساز نفوذ..... ۱۷
- شکل ۲-۳ شمای کلی معماری ارائه شده توسط یو ۳۴
- شکل ۳-۳ نمونه ای از عملکرد سیستم ۳۹
- شکل ۴-۳ نمایی از نحوه حرکت پنجره روی داده‌ها ۴۱
- شکل ۵-۳ نمونه‌ای از نمودار I ۴۲
- شکل ۶-۳ معماری ارائه شده توسط لی و استلفو ۵۱
- شکل ۷-۳ معماری ارائه شده توسط رسولی فرد ۵۳
- شکل ۸-۳ فازهای یادگیری مدل و به روزرسانی آن در مدل رسولی فرد و همکارانش ۵۵
- شکل ۹-۳ بررسی تأثیر افزایش تعداد کلاس‌بندها بر دقت سیستم در معماری ارائه شده توسط رسولی فرد و همکارانش ۵۹
- شکل ۱۰-۳ معماری ارائه شده توسط شفی و همکارانش ۵۹
- شکل ۱-۴ کنترل کننده فازی، به زبان ساده ۶۹
- شکل ۲-۴ نمونه‌ای از غیرفازی سازی ۷۱
- شکل ۳-۴ خوشه‌بندی کلاسیک ۷۱
- شکل ۴-۴ خوشه‌بندی فازی ۷۲
- شکل ۵-۴ معماری ارائه شده ۷۴
- شکل ۶-۴ توابع عضویت متغیرهای ورودی سیستم وفق‌دهنده فازی ۷۴
- شکل ۷-۴ توابع عضویت خروجی سیستم وفق‌دهنده فازی ۷۵
- شکل ۱-۵ مقایسه روش ارائه شده و سایر روشهای متداول کلاس‌بندی ۸۴
- شکل ۲-۵ توابع عضویت فازی برای $x \leq a$ و $x \geq a$ ۸۵

شکل ۳-۵ توابع عضویت خروجی سیستم میزان کننده فازی ۸۷

نمایه نرمال	Normal Profile
سیستمهای شناسایی نفوذ	Intrusion Detection System
نفوذ/حمله	Intrusion
الگوی حملات شناخته شده	Known Attack Patterns
اداره استخبارات فدرال ایالات متحده امریکا	FBI
ناهنجاری/رفتار غیر متعارف	Anomay
شناسایی نفوذگر بر مبنای تشخیص ناهنجاری	Anomay-Based Intrusion Detection
وفقی	adaptive
محرمانگی	Confidentiality
در دسترس بودن	Availability
جامعیت	Integrity
مبتنی بر الگو	Signature-Based
شخصی کردن	Customize
رخنه‌های نفوذی	Vulnerable
مدل سازی بر اساس قوانین فازی	Fuzzy Rule-Based Modeling
FCM	Fuzzy C-Mean Clustering
درصد اعتماد به پیش بینی	Prediction Confidence Ratio
میزان کردن	Tune
میزان کننده مدل	Model Tuner
استلفو	Stolfo
لی	Lee
میزبان یکسان	Same Host
قطعات	Fragments
فوری	Urget
داغ	Hot
مصالحه	Compromised
پوسته ریشه	Root Shell
ریشه	Root
اعلان پوسته	Shell Prompt
تولایی	Tavallae
تشخیص درست	SuccessfulPrediction
آندرسن	Anderson
دنینگ	Denning
اکسلسون	Axelsson
جمع آوری اطلاعات بازرسی	Audit data Collection
نگهدارنده اطلاعات بازرسی	Audit data storage
سوء استفاده	Misuse Detection
مثبت کاذب	False Positive
منفی کاذب	False Negative
منفی درست	True Negative
مثبت درست	True Positie
براساس شبکه	Network Based

پیگیری	Ttrace
مبتنی بر تک دستگاه	Host Based
ایستا	Static
بازخورد	FeedBack
معیار میزان شدت فعالیت	Activity Intensity Measure
معیار توزیع رکوردهای بازرسی	Audit record distribution measure
وصفی	Ordinal
غیر برخط/آفلاین	Offline
بلادرنگ	Real-Time
موسسه تحقیقاتی استنفورد	Stanford Research Institute(SRI)
سیستم موتور آماری تشخیص ناهنجاری در بسته-ها	Statistical Packet Anomaly Detection Engine(SPADE)
امتیاز ناهنجاری	Anomaly Score
پوشش دروازه	Port Scan
فورست	Forrest
جدول مراجعه ای	Lookup Table
هافمیر	Hofmeyr
اسکین	Eskin
والدز	Valdes
شبکه بیزین ساده	Naïve Bayesian Network
	Principle Component Analysis(PCA)
شیو	Shyu
ماهالانوبیس	Mahalanobis
یه	Ye
حداکثر کردن امید	Expectation Maximization(EM)
یونگ	Yeung
دستورهای یوسته‌ای	Shell Command
تولید قوانین استقرایی	Rule Generation Techniques Inductive
بدون سرپرست	Unsupervised
دیکرسون	Dickerson
FIRE	Fuzzy Intrusion Recognition Engine
راماداس	Ramadas
ANDSOM	Anomalous Network-Traffic Detection with Self Organizing Maps
SOM	Self-Organizing Map
SVM	Support Vector Machine
عنصر خارجی	Outlier Detection
MINDS	Minnesota Intrusion Detection System
یادگیری قوانین انجمنی	Association rule discovery
اقلام	Items
باربارا	Barbara'
ADAM	Audit Data Analysis and Mining

سیستم شناساگر نفوذ با قابلیت به روز رسانی اتوماتیک	An Adaptive Automatically Tuning Intrusion Detection System
یو	Yu
اسلیپر	SLIPPER
ممانعت از سرویس	DOS
چهارچوبی برای شناساگر نفوذ وقتی بر اساس بردارهای پشتیبانی توصیف داده	A Framework for Adaptive Anomaly Detection Based on Support Vector Data Description
یانگ	Yang
ابر کره	Hypersphere
شناساگر نفوذی وقتی بر اساس خوشه بندی و بر مبنای روش هسته	An Adaptive Intrusion Detection Algorithm Based on Clustering and Kernel-Method
عددی	Numeric
نمادین	Symbolic
نرخ یادگیری	Learning Rate
بردار مفهوم	Concept Vector
استفاده از تشدید تطبیقی فازی در شناسایی ناهنجاری وقتی شبکه	Investigation of Fuzzy Adaptive Resonance Theory in Network Anomaly Intrusion Detection
چارچوبی برای تشخیص ناهنجاری وقتی با استفاده از داده کاوی فازی	Detection System A Framework for an Adaptive Anomaly with Fuzzy Data Mining
ژیانگ	Xiang
یادگیری سیستم شناساگر نفوذ با استفاده از الگوریتم بیزین وقتی	Learning Intrusion Detection Based on Adaptive Bayesian Algorithm
شناساگر بلادرنگ نفوذی با استفاده از روش داده-کاوی	Real time data mining-based intrusion detection
چارچوبی برای تشخیص نفوذ با استفاده از شبکه بیزین	A Framework for an Adaptive Intrusion Detection System using Bayesian Network
جمیلی	Jemili
ذقدود	Zaghdoud
بن احمد	Ben Ahmed
مدل ترکیبی افزایشی با استفاده از گروهی از کلاس-بندهای ضعیف	Incremental Hybrid Intrusion Detection Using Ensemble of Weak Classifiers
یادگیرهای ضعیف	Weak Learner
سیستم شناساگر نفوذ وقتی با استفاده از تکامل اتصالات سیستمی	Adaptive anomaly detection with evolving connectionist systems
لی هوآ	Liao
بر خط	Online
بردارهای وزن داری	Weighted Vector
سامانه تطبیقی یادگیری امضاء با استفاده از الگوریتم ژنتیک در تشخیص نفوذ	An adaptive genetic-based signature learning system for intrusion detection
شفی	Shafi
ویژگی عملیاتی دریافت کننده	Receiver Operating Characteristic
اکسلسون	Axelsson
مولد مدل تشخیص	Detection Model Generator
خوشه بندی کاهشی	Subtractive Clustering

سبک	Style
میشیگان	Michigan
پیتزبرگ	Pittsburgh
کوچکترین ماکزیمم	Smallest Of Maximom
مرتبہ زمانی	Order

چکیده:

به نظر می‌رسد استفاده از روشهای سابق از جمله استفاده از دیوارهای آتش و سیستم‌های تشخیص نفوذ مبتنی بر امضا به تنهایی قادر به تأمین امنیت شبکه نیستند، در نتیجه صرف هزینه‌های بیشتر برای تأمین امنیت غیر قابل انکار است. یکی از این هزینه‌ها می‌تواند تشخیص نفوذ بر اساس تشخیص رفتار غیرمتعارف (ناهنجاری) باشد. در این روشها ابتدا نمایه نرمال^۱ سیستم مدل شده و از این مدل جهت شناسایی هر گونه رفتار مشکوک استفاده می‌گردد. روشهای مبتنی بر تشخیص ناهنجاری بر خلاف روش‌های سنتی می‌توانند حملات جدید را نیز تشخیص دهند اما تعداد زیاد هشدارهای کاذب مهمترین ضعف آن است، علاوه بر این، ساخت نمایه نرمال و به‌روز نگه داشتن آن از دیگر چالشهای این روش است.

در معماری ارائه شده از مدل سازی مبتنی بر قوانین فازی جهت مدل کردن رفتار سیستم استفاده شده است. مدل سازی مبتنی بر قوانین فازی تلاش می‌کند خروجی سیستم بر اساس ورودیهای آن را با استفاده از قوانین فازی مدل کند. چگونگی به‌دست آوردن قوانین و مجموعه‌های فازی متغیرهای ورودی از مشکلات اصلی این نوع روش مدل سازی است. الگوریتم ژنتیک برای پیدا کردن قوانین مدل و FCM^۲ برای پیدا کردن مجموعه‌های فازی متغیرهای ورودی استفاده شده است. برای هر یک از قوانین فازی آموزش دیده، درجه اعتماد به پیش بینی در نظر گرفته شده که در حقیقت میزان اطمینان سیستم به آن قانون محسوب می‌شود. این پارامتر نسبت نمونه‌های درست تشخیص داده شده به کل نمونه‌های آموزشی است.

با توجه به تغییر در رفتار نرمال شبکه و وقوع حملات جدید، استفاده از مدل‌های غیرانطباقی مناسب نیست و تلاش برای به‌روزرسانی مدل غیر قابل انکار است. در روش ارائه شده، درجه اعتماد هر یک از قوانین با توجه به عملکردشان در طول فاز تست و با استفاده از یک کنترل کننده فازی به‌روز رسانده می‌شوند. با استفاده از مدل غیرانطباقی، دقت حدود ۵۵ درصد عاید ما شد، در حالی که استفاده از روش انطباقی این دقت را تا ۷۵ درصد افزایش داد.

نگاه ساده به مسئله به‌روز رسانی مدل تشخیص، به‌روز رسانی مدل با استفاده از پارامترهای کم اما مؤثر، تفاوت محسوس در دقت در مقایسه با روش‌های غیرانطباقی و به‌روز رسانی تقریباً آنلاین از مهم‌ترین مزایای سیستم انطباقی پیشنهادی هستند.

کلمات کلیدی: تشخیص تطبیقی نفوذ بر مبنای تشخیص ناهنجاری، مدل سازی مبتنی بر قوانین فازی، کنترل کننده

فازی

^۱ Normal Profile

^۲ Fuzzy C-Mean Clustering

فصل اول:

معرفی پروژه