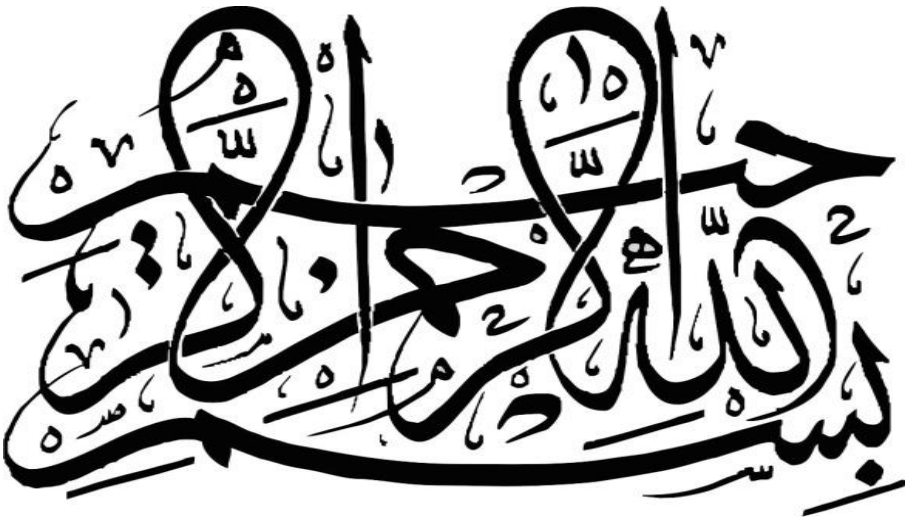

بسم الله الرحمن الرحيم

۹۱۱٤۲۲۹

دانشگاه شهید چمران اهواز

دانشکده مهندسی

پایان‌نامه کارشناسی ارشد مهندسی کامپیوتر

گرایش هوش مصنوعی

عنوان:

دسته‌بندی ترافیک شبکه با استفاده از الگوریتم‌های هوش محاسباتی

استاد راهنما:

آقای دکتر علیرضا عصاره

استاد مشاور:

خانم دکتر بیتا شادگار

نگارنده:

شیرین جمشیدیان

آبان ۱۳۹۱

باسمه‌تعالی

دانشگاه شهید چمران اهواز
دانشکده‌ی مهندسی

(نتیجه ارزشیابی پایان‌نامه دوره کارشناسی ارشد/دکتری)

پایان‌نامه خانم شیرین جمشیدیان دانشجوی رشته مهندسی کامپیوتر گرایش هوش مصنوعی

از دانشکده مهندسی به شماره دانشجویی ۸۸۱۴۲۰۴

با عنوان:
دسته‌بندی ترافیک شبکه با استفاده از الگوریتم‌های هوش محاسباتی

جهت اخذ مدرک کارشناسی ارشد در تاریخ ۲۹/ ۸ /۱۳۹۱ توسط هیأت داوران مورد ارزشیابی قرار گرفت و با درجه‌ی ... تصویب شد.

۱- اعضا هیأت داوران:	**رتبه علمی**	**امضا**
استاد راهنما: دکتر علیرضا عصاره	دانشیار	
استاد مشاور: دکتر بیتا شادگار	استادیار	
استاد داور: دکتر سید عنایت الله علوی	استادیار	
استاد داور: دکتر مرجان نادران طحان	استادیار	
نماینده تحصیلات تکمیلی دانشگاه: دکتر کریم انصاری اصل	استادیار	
۲- مدیر گروه: دکتر سید عنایت الله علوی	استادیار	
۳- معاون پژوهشی و تحصیلات تکمیلی دانشکده: دکتر علی حقیقی	استادیار	
۴- مدیر تحصیلات تکمیلی دانشگاه: دکتر مسعود قربانپور نجف‌آبادی	استاد	

تقدیم به

امید زندگی‌ام، پدرم

زیبایی زندگی‌ام، مادرم

و عشق زندگی‌ام، همسرم

به پاس لطف بی‌کران، امیدها و آرزوهایشان

تقدیر و تشکر

خداوند مهربان را شاکرم که لطف خویش را شامل حال من فرمود تا بتوانم در سایه‌ی یاری حضرتش مرحله‌ای دیگر از زندگی علمی خود را به اتمام برسانم. اینک که در پرتو الطاف بی‌کران خداوندی، نگارش این پایان‌نامه را به اتمام رسانده‌ام، برخود واجب می‌دانم از تلاش همه‌ی کسانی که به نحوی مرا در رسیدن به این مرحله یاری داده‌اند سپاس‌گزاری نمایم.

پیش از همه از پدر و مادر عزیزم که در تمامی مراحل زندگی، مشوق من بوده‌اند، از همسر مهربانم که با هم فکری‌های خود کمک شایانی به انجام این کار نمود و از برادران عزیزم که همیشه موفقیت و شادی مرا خواسته‌اند بسیار ممنونم.

از اساتید گرامی‌ام جناب آقای دکتر عصاره و سرکار خانم دکتر شادگار که با علم فراوان خویش راه انجام این پروژه را بر من هموار ساختند تشکر می‌کنم.

همچنین از دوست عزیزم سرکار خانم سمیرا دریکوند که زیستن در کنارش جزئی از خاطرات زیبای زندگی‌ام شد سپاس‌گزارم.

و در آخر از همه آنانی که از موفقیت من دل‌شاد می‌شوند و در به سرانجام رسیدن این پروژه یاری‌ام داده‌اند تشکر می‌کنم.

فهرست مطالب

فصل اول: مقدمه

فصل دوم: مروری بر ادبیات موضوع و پیشینه‌ی تحقیق

فصل سوم: ابزارها و روش‌ها

فصل چهارم: روش انجام کار و ارزیابی نتایج

فصل پنجم: نتیجه‌گیری و کارهای آینده

فهرست شکل‌ها

فهرست جداول

نام خانوادگی: جمشیدیان	**نام:** شیرین	**شماره دانشجویی:** ۸۸۱٤۲۰٤
عنوان پایان نامه: دسته‌بندی ترافیک شبکه با استفاده از الگوریتم‌های هوش محاسباتی		
استاد راهنما: دکتر علیرضا عصاره		
استاد مشاور: دکتر بیتا شادگار		
درجه تحصیلی: کارشناسی ارشد	**رشته:** مهندسی کامپیوتر	**گرایش:** هوش مصنوعی
دانشگاه: شهید چمران اهواز	**دانشکده:** مهندسی	**گروه:** مهندسی کامپیوتر
تاریخ فارغ التحصیلی: آبان ۱۳۹۱		**تعداد صفحه:** ۱۳۱

کلید واژه‌ها: دسته‌بندی ترافیک، یادگیری ماشین، هوش محاسباتی، نمونه‌برداری، انتخاب ویژگی، الگوریتم ژنتیک، دسته‌بندهای ترکیبی

چکیده: با توجه به استفاده‌ی روزافزون از شبکه‌ی اینترنت و لزوم وجود ابزارهایی جهت کنترل و مدیریت ترافیک شبکه‌ها و نیز پیدایش مداوم کاربردهایی که با استفاده از تکنیک‌های بدیع و پیچیده‌ی مبهم‌سازی قصد فریب این ابزارها را دارند، طراحی دقیق سیستم‌های شناسایی و دسته‌بندی کاربردهای مختلف شبکه از اهمیت زیادی برخوردار است. هدف این پایان‌نامه طراحی و پیاده‌سازی یک سیستم هوشمند و کارا جهت شناسایی و دسته‌بندی ترافیک شبکه، مبتنی بر تکنیک‌های یادگیری ماشین و هوش محاسباتی است. ساخت چنین سیستمی شامل سه مرحله است؛ مرحله‌ی اول مربوط به ساخت مجموعه داده و مرحله‌ی دوم شامل اقدامات پیش‌پردازشی از قبیل انتخاب ویژگی است. در این مرحله انتخاب ویژگی‌های مناسب به‌وسیله‌ی یک مدل سه لایه‌ی پیشنهادی انجام می‌شود. در مرحله‌ی سوم یعنی مرحله‌ی دسته‌بندی، روش‌های ترکیبی پایه شامل بگینگ، آدابوست و استکینگ و همچنین یک روش پیشنهادی روی داده‌ها اعمال شده‌اند. نتایج مقایسات، نشان از دقت بالای سیستم پیشنهادشده دارد. روش پیشنهادی، ترافیک UDP را با دقت ۹۱/۲۱ و ترافیک TCP را با دقت ۹۹/٦٤، شناسایی و دسته‌بندی می‌کند. همچنین این پژوهش از نخستین کارهایی است که روش استکینگ را در زمینه‌ی دسته‌بندی ترافیک شبکه به کار گرفته است.

فصل اول

مقدمه

در این فصل پس از ذکر برخی از مفاهیم پایه در زمینه‌ی دسته‌بندی ترافیک[1] شبکه و شرح اجمالی روش‌های شناسایی و دسته‌بندی جریان‌ها[2]، به معرفی مراحل اصلی ساخت یک سیستم هوشمند دسته‌بندی ترافیک پرداخته می‌شود. از آن‌جا که اولین گام در انجام این پایان‌نامه، ساخت مجموعه داده مبتنی بر جریان و استخراج ویژگی‌های جریان‌ها با استفاده از سرآیند[3] بسته‌هاست[4]، در ادامه‌ی این فصل، شرح مختصری از لایه‌های شبکه و سرآیند بسته‌های پروتکل‌های مختلف، ارائه و در انتها نیز ساختار کلی این پایان‌نامه توضیح داده شده است.

۱-۱ آشنایی

امروزه استفاده از شبکه‌ی اینترنت به‌طور وسیعی در سراسر دنیا مرسوم شده است. از این‌رو متصدیان[5] شبکه به‌منظور کنترل و مدیریت ترافیک شبکه‌هایشان از جمله:

- مدیریت بافرهای هوشمند برای فراهم کردن میزان کیفیت خدمات[6] هر کاربرد[7]

---

[1] Traffic classification
[2] Flows
[3] Header
[4] Packets
[5] Administrator
[6] Quality Of Service (QOS)
[7] Application

- شناسایی ترافیک‌های زمان‌حقیقی[۱] جهت تخصیص اولویت بالاتر به آن‌ها
- شناسایی حملات و موارد نقض سیاست امنیتی
- مدیریت پهنای باند
- کنترل دسترسی پویا (به‌منظور محدود یا مسدودسازی ترافیک کاربردهای خاص)

نیازمند استفاده از ابزارهای مدیریت شبکه هستند. لازمه‌ی موفقیت چنین ابزارهایی توانایی بالای آن‌ها در تشخیص[۲] و دسته‌بندی[۳] کارا و زمان‌حقیقی انواع جریان‌های شبکه است.

۱-۲ جریان و کاربردهای شبکه

جریان مجموعه‌ای از بسته‌های پی‌درپی است که آدرس IP[۴] مبدأ و مقصد، شماره‌ی درگاه[۵] مبدأ و مقصد و نوع پروتکل یکسانی دارند. جریان‌ها عمدتاً دوطرفه[۶] تعریف می‌شوند. اولین بسته‌ی دیده شده توسط سیستم دسته‌بندی، جهت مستقیم جریان را مشخص می‌کند.

دسته‌بندی جریان‌های جاری در شبکه بر اساس نوع کاربردهای تولیدکننده‌ی آن‌ها، از طریق تحلیل ترافیک شبکه را دسته‌بندی ترافیک شبکه گویند. از جمله‌ی این دسته‌ها می‌توان به کاربردهای نظیربه‌نظیر[۷] (P2P)، چندرسانه‌ای[۸]، وب‌گردی[۹] و فعالیت‌های ایمیلی اشاره کرد. رایج‌ترین دسته‌بندی کاربردها مطابق جدول ۱-۱ است [۱, ۲].

[1] Real time
[2] Identification
[3] Classification
[4] Internet Protocol (IP)
[5] Port
[6] Bi-directional
[7] Peer2Peer (P2P)
[8] Multimedia
[9] Web-browsing

جدول ۱-۱: لیست کاربردهای متعلق به هر گروه

گروه	کاربردها
Bulk	FTP
Database	Postgres, slqnet, oracle, ingres
Interactive	SSH, klogin, rlogin, telnet
Mail	IMAP, POP2, POP3, SMTP
Service	X11, DNS, LDAP, NTP
WWW	WWW
P2P	KaZaA, BitTorrent, GnuTella
Malicious	Internet work and virus attacks
Games	Half-life
Multimedia	Windows media player, Real player

۱-۳ روش‌های شناسایی و دسته‌بندی ترافیک شبکه

تحقیقات انجام شده در زمینه‌ی شناسایی و دسته‌بندی ترافیک شبکه به چند دسته‌ی کلی تقسیم می‌شوند.

- روش‌های مبتنی بر درگاه

یکی از رایج‌ترین روش‌های دسته‌بندی ترافیک، روش مبتنی بر درگاه[1] است که از طریق بررسی شماره‌های درگاه لایه‌ی انتقال که توسط سازمان IANA[2] [۳] به برنامه‌های کاربردی مختلف اختصاص یافته، صورت می‌گیرد. ترافیک کاربردهای سنتی به خاطر استفاده از شماره‌های درگاه مخصوص به خود که در IANA ثبت شده است، با این روش به راحتی شناخته می‌شوند. اما امروزه این روش به دلایل زیر کارایی بالایی ندارد [۱, ٤]:

- پیدایش کاربردهای جدیدی (مانند P2P) که به دلایل مختلف –نظیر جلوگیری از شناخته و محدود شدن توسط ناظران شبکه– از شماره‌های درگاه ثابت و ثبت شده استفاده نکرده و حتی برای فریب سیستم‌های نظارتی، از شماره‌ی درگاه اختصاص یافته به سایر کاربردها استفاده می‌کنند.

[1] Port-based
[2] International Assigned Number Authority (IANA)

- گسترش تکنیک‌های مبهم‌سازی نظیر رمزگذاری لایه‌ی IP که باعث عدم دسترسی به شماره‌ی درگاه واقعی می‌شود.

- روش‌های مبتنی بر محتوا

روش‌های دسته‌بندی مبتنی بر محتوا[1] برای حل برخی از معضلات روش مبتنی بر درگاه، الگوهای شناخته شده و ثابت هر کاربرد که امضای[2] آن کاربرد نامیده می‌شود را در محتوای بسته‌ها جست‌وجو می‌کنند؛ یعنی محتوای بسته‌ها جهت یافتن دنباله‌هایی خاص از بایت‌ها مورد بررسی قرار می‌گیرند [۱, ۵, ٦]. با وجود رفع برخی از چالش‌های تکنیک مبتنی بر درگاه، این روش همچنان در شناسایی کاربردهایی مانند P2P که از روش‌های مبهم‌سازی مانند رمزگذاری محتوای بسته‌ها استفاده می‌کنند، کارایی پایینی دارد. همچنین از سایر معایب آن می‌توان به موارد زیر اشاره کرد:

- تجسس در محتوای بسته‌ها که مغایر با اصل استقلال حریم کاربر است.
- ناتوانی در شناسایی کاربردهایی که از قالب و امضای آن‌ها آگاه نیست
- دشواری به روز نگه داشتن امضاها

- روش‌های مبتنی بر رفتار میزبان

یکی از راه‌های دسته‌بندی ترافیک، بدون استفاده از محتوای بسته‌ها دسته‌بندی مبتنی بر رفتار میزبان‌هاست که به سبب عدم نیاز به بررسی محتوای بسته‌ها در برابر رمزنگاری ترافیک، مقاوم است. در این روش با استفاده از روش‌های اکتشافی[3] الگوهایی برای رفتار ارتباطی، کارکردی و اجتماعی میزبان، هنگام استفاده از برنامه‌های کاربردی مختلف جست‌وجو می‌شود.

- روش‌های مبتنی بر استفاده از ویژگی‌های آماری جریان

[1] Payload-based
[2] Signature
[3] Heuristic

با توجه به محدودیت‌های روش‌های اشاره شده، استفاده از ویژگی‌های آماری[1] لایه‌ی انتقال[2] جریان‌ها مطرح شد. در این روش‌ها، مجموعه‌ای از ویژگی‌های آماری جریان به عنوان معیارهای دسته‌بندی انتخاب می‌شوند. یک ویژگی، یک خصیصه‌ی آماری توصیفی است که با استفاده از اطلاعات یک یا چند بسته از یک جریان خاص و با استفاده از سرآیند بسته‌ها محاسبه می‌شود. به عنوان مثال، مدت زمان استمرار جریان[3]، حجم جریان بر اساس تعداد بسته و بایت، طول بسته‌ها (کمینه، بیشینه، میانگین و انحراف معیار) و فاصله‌ی زمانی میان رسیدن بسته‌ها می‌توانند به عنوان ویژگی‌های آماری انتخاب شوند.

این رویکرد بر این حقیقت تکیه دارد که جریان کاربردهای مختلف در هنگام عبور از شبکه، الگوها و ویژگی‌های رفتاری مجزایی دارند که می‌توان با کمک این الگوها آن‌ها را دسته‌بندی کرد و برای این کار نیازی به بررسی محتوای بسته‌ها هم نیست.

همچنین به‌علت سروکار داشتن با مجموعه داده‌های بزرگ و پیچیده و فضای چندبعدی ویژگی‌های مربوط به جریان‌ها و بسته‌ها، استفاده از تکنیک‌های یادگیری ماشین در این حوزه رونق گرفت [۷].

۱-٤ انگیزه و هدف پایان‌نامه

همان‌طور که گفته شد به‌علت ترفندهای پیچیده‌ی به کار رفته در ساخت کاربردهای P2P مانند استفاده از شماره‌های درگاه پویا و رمزگذاری محتوای بسته‌ها، روش‌های مبتنی بر درگاه و مبتنی بر محتوا در شناسایی ترافیک این نوع کاربردها موفقیت زیادی ندارند [۵]. از طرف دیگر شناسایی این کاربردها به علت:

- مصرف زیاد پهنای باند، ایجاد یک شبکه‌ی پرازدحام[4] و کاهش کارایی شبکه

[1] Statistical Features
[2] Transport
[3] Duration
[4] Congested Network

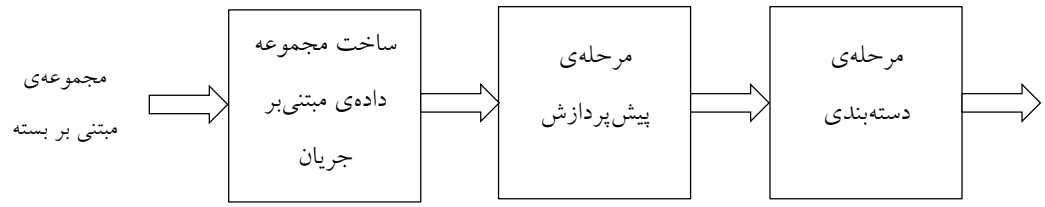
- در اختیار داشتن درصد بالایی از ترافیک فعلی اینترنت

بخش مهمی از مسئله‌ی دسته‌بندی ترافیک است. مطالعات نشان می‌دهد بیش از ٦٠٪ ترافیک فعلی اینترنت را کاربردهای P2P ایجاد می‌کنند [۸].

هدف از انجام این پژوهش، طراحی یک سیستم دسته‌بندی هوشمند، به‌منظور شناسایی و دسته‌بندی کاربردهای شبکه، مبتنی بر استفاده از الگوریتم‌های هوش محاسباتی و تکنیک‌های یادگیری ماشین است که دقت خوبی در شناسایی کاربردهای مختلف از جمله کاربردهای P2P داشته باشد. این سیستم به جای تکیه بر شماره‌ی درگاه یا تفسیر محتوای بسته‌ها به رفتار و مشخصه‌های آماری جریان‌ها متکی است.

۱-۵ مراحل اصلی ساخت سیستم هوشمند دسته‌بندی ترافیک

شکل ۱-۱ مراحل اصلی ساخت یک سیستم هوشمند دسته‌بندی با استفاده از ویژگی‌های آماری جریان‌ها را نشان می‌دهد.



شکل ۱-۱: مراحل کلی ساخت سیستم دسته‌بندی ترافیک شبکه

مراحل نشان داده شده در شکل ۱-۱ به شرح ذیل هستند:

- مرحله‌ی ساخت مجموعه داده مبتنی بر جریان

برای ایجاد یک مجموعه داده‌ی مبتنی بر جریان، ابتدا لازم است یک مجموعه‌ی مبتنی بر بسته در اختیار داشته باشیم. برای این کار می‌توان ترافیک شبکه را با اتصال تجهیزات لازم در مسیر شبکه به‌صورت خصوصی تولید کرد یا از مجموعه داده‌هایی که به‌صورت عمومی در دسترس هستند استفاده نمود [۹-۱۱]. سپس این بسته‌ها باید به‌صورت جریان‌های مجزا تفکیک شده و مقادیر ویژگی‌های موردنظر هر جریان، از طریق بسته‌های تشکیل‌دهنده‌ی آن استخراج شود. این کار می‌تواند از طریق ابزارهای آماده مانند NetMate [۱۲] یا دیگر ابزارهای تجاری تحلیل ترافیک، صورت پذیرد. بدی استفاده از این‌گونه ابزارها این است که آن‌ها آمار محدودی را راجع به جریان‌ها تولید می‌کنند. راه دیگر استخراج تعداد زیادی ویژگی درباره‌ی جریان‌ها، از طریق کدنویسی است.

- مرحله‌ی پیش‌پردازش

در این مرحله به‌منظور ایجاد بهبود نتایج مرحله‌ی بعد، اقداماتی روی مجموعه داده‌ی اولیه صورت می‌گیرد، اقداماتی از قبیل نمونه‌برداری مجدد[1] و انتخاب ویژگی‌های[2] مناسب. هدف این‌گونه اقدامات، عموماً کاهش بار محاسباتی سیستم و افزایش کیفیت داده‌های مورد بررسی است.

- مرحله‌ی دسته‌بندی

در مرحله‌ی پایانی، الگوریتم‌های یادگیری روی مجموعه‌ی حاصل از مرحله‌ی قبل، آموزش می‌بینند تا با دیدن نمونه‌های جدید، بتوانند در مورد نوع آن‌ها تصمیم بگیرند.

[1] Resampling
[2] Feaure Selection