



**دانشگاه اصفهان**

**دانشکده مهندسي کامپیوتر**

**گروه مهندسی فناوری اطلاعات**

**پایان‌نامه کارشناسی**

**رشته‌ مهندسی کامپيوتر گرايش فناوری اطلاعات**

**عنوان پروژه:**

تحلیل زمانی رفتار ترافیکی شبکه با استفاده از الگوریتم‌های تحلیل سری زمانی

**استاد راهنما:**

دکتر بهروز شاهقلی

**پژوهشگران:**

|  |  |
| --- | --- |
| رضا پازن | علی هداوند |

**شهریور 1400**



دانشگاه اصفهان

دانشکده مهندسي کامپیوتر

گروه مهندسی فناوری اطلاعات

پروژه کارشناسی رشته‌ي مهندسی کامپيوتر گرايش فناوری اطلاعات  
آقایان علی هداوند و رضا پازن

تحت عنوان

تحلیل زمانی رفتار ترافیکی شبکه با استفاده از الگوریتم‌های تحلیل سری زمانی

در تاريخ / / 13 توسط هيأت داوران زير بررسي و با نمره به تصويب نهايي رسيد.

1- استاد راهنماي پروژه:

دکتر امضا

2- استاد داور :

دکتر امضا

امضاي مدير گروه

تقديم به

پدر و مادر عزیزمان، که تلاش بی‌دریغ‌شان روشنی‌بخش آینده‌ی ما بود.

چکيده:

این فایل حاوی قالب گزارش پروژه کارشناسی می‌باشد. در این فایل ترتیب بخش‌های مختلف گزارش فرمت آن شامل: فونتها، حاشیه‌ها و... مشخص شده است، صفحات اولیه را با توجه به اطلاعات خودتان تکمیل کنید(نام و نام خانوادگی، موضوع پروژه و ...) دقت کنید که اندازه فونت‌ها و حاشیه‌ها تغییر نکند. در تمام متن پروژه، برای متن فارسی از فونت B Nazanin 13 و برای متن انگلیسی از فونت Times New Roman 11 استفاده کنید. اندازه فونت فارسی 13 و اندازه فونت انگلیسی 11 باشد(به غیر از صفحات اولیه و عناوین بخش‌ها ). متن چکیده از 200 کلمه کمتر نباشد و از یک صفحه بیشتر نشود. بعد از متن چکیده، ترتیب مطالب به این صورت باشد: فهرست مطالب، فهرست شکل‌ها، فهرست جدول‌ها، لیست مخفف‌ها، فصل اول(مقدمه)، فصل دوم(معرفی مفاهیم) و فصلهای بعدی(تعداد فصلها با انتخاب خودتان) در مورد کاری که شما انجام داده اید و نهایتا فصل آخر، نتیجه‌گیری و پیشنهادات می‌باشد. بعد از آن پیوستها(در صورت وجود) و نهایتا لیست مراجع آورده شود.

هر فصلی با زیر بخش "مقدمه" شروع می‌شود و با زیربخش "خلاصه فصل" خاتمه می‌یابد(البته فصل اول[مقدمه] و فصل آخر[نتیجه‌گیری و پیشنهادات] نیازی به زیربخش‌های "مقدمه" و "جمع‌بندی" ندارند)

در متنی که می‌نویسید هر جا که مطلبی از جایی آورده می‌شود باید با استفاده از [] شماره مرجع مشخص شود. شماره مرجع‌ها از 1 شروع و افزایش می‌یابد. در قسمت لیست مراجع، اطلاعات کامل مرجع آورده می‌شود.

شکل‌ها و جدول‌ها باید شماره گذاری شوند(مثلا شکل 3-1، یعنی اولین شکل از فصل 3) و برای آنها عنوان گذاشته شود و در متن نیز به آنها ارجاع داده شود.

پیشنهاد می‌شود که یک کپی از این فایل تهیه فرمایید و بخش‌های مختلف آن را پر نمایید.

واژگان کليدي: دانشگاه اصفهان، گرایش سخت افزار کامپیوتر، پروژه کارشناسی(حداقل 3 کلمه کلیدی پروژه خود را بنویسید)

فهرست مطالب

[1− فصل اول: مقدمه 1](#_Toc80630209)

[1−1 بیان مسئله 1](#_Toc80630210)

[2−1 ارزش پروژه 1](#_Toc80630211)

[3−1 هدف پروژه 1](#_Toc80630212)

[4−1 ساختار پایان‌نامه 2](#_Toc80630213)

[2− فصل دوم: مبانی و مفاهیم تحلیل سری زمانی 3](#_Toc80630214)

[1−2 مقدمه 3](#_Toc80630215)

[2−2 سری زمانی 3](#_Toc80630216)

[3−2 ویژگی‌های رفتاری سری زمانی 3](#_Toc80630217)

[1−3−2 فصلی بودن 3](#_Toc80630218)

[2−3−2 تناوب 3](#_Toc80630219)

[3−3−2 روند 3](#_Toc80630220)

[4−3−2 خطا 3](#_Toc80630221)

[5−3−2 باقی‌مانده 3](#_Toc80630222)

[6−3−2 ایستایی 3](#_Toc80630223)

[4−2 پیش‌گویی 3](#_Toc80630224)

[3− فصل سوم: شرح پروژه 4](#_Toc80630225)

[1−3 مقدمه 4](#_Toc80630226)

[2−3 جمع‌بندی 4](#_Toc80630227)

[4− فصل چهارم: نتایج 5](#_Toc80630228)

[1−4 مقدمه 5](#_Toc80630229)

[2−4 جمع‌بندی 5](#_Toc80630230)

[5− فصل پنجم: نتیجه‌گیری و پیشنهادات 6](#_Toc80630231)

[6− پیوست۱: لیست برنامه‌ها 7](#_Toc80630232)

[7− منابع 8](#_Toc80630233)

فهرست شکل‌ها

[شکل ‏1‑1: شکل اول 1](#_Toc80296369)

فهرست جدول‌ها

[جدول ‏1‑1: تست جدول اول 1](#_Toc80296438)

مخفف‌ها

|  |  |
| --- | --- |
| Computer Networks | CN |
| Time Series | TS |
| Auto Correlation Function | ACF |
| Partial Auto Correlation Function | PACF |
| Auto Regressive | AR |
| Moving Average | MA |
| Auto Regressive Moving Average | ARMA |
| Autoregressive Integrated Moving Average | ARIMA |

# فصل اول: مقدمه

## بیان مسئله

در طول تاریخ یکی از اساسی‌ترین نیازهای انسان برقراری ارتباط بوده است که با گذر زمان، ابزار و روش‌های آن نیز دستخوش تغییر شده‌اند. امروزه، فناوری شبکه‌های کامپیوتری[[1]](#footnote-1) (CN) مهم‌ترین ابزار برآورده کردن خواسته‌های تعاملی انسان است. از ابتدای پیداش این فناوری تا کنون، به دلیل گسترش جوامع و به دنبال آن گسترش استفاده از شبکه‌های کامپیوتری، میزان داده‌ی جاری در این شبکه‌ها افزایش یافته و چالش‌هایی نیز پیش روی کاربران و توسعه‌دهندگان این فناوری قرار داده است. مواردی مانند تامین محرمانگی، وجود یا عدم وجود خطا، صحت انتقال اطلاعات، تأخیر و ... از جمله این چالش‌ها هستند.

با توجه به گسترده شدن کاربرد شبکه‌های کامپیوتری و حضور این فناری در تمام عرصه‌های زندگی انسان، از سازمان‌های بزرگ تا کاربری خانگی، حفظ سلامت این شبکه‌ها امر بسیار مهمی تلقی می‌شود. یکی از راه‌های کنترل CNها، پایش[[2]](#footnote-2) آن‌ها است. پایش به معنی جمع‌آوری و تحلیل بسته‌های[[3]](#footnote-3) انتقال داده شده در شبکه است [۱]. کارشناسان و متخصصان، با استفاده از ابزارها و روش‌های مختلف پایش شبکه مانند پروتکل مدیریت ساده‌ی شبکه[[4]](#footnote-4) (SNMP) و پایش بلادرنگ[[5]](#footnote-5) قادر به مشاهده، ذخیره و بررسی داده‌های شبکه هستند که به آن‌ها کمک می‌کند رفتار شبکه‌ی موردنظر را ثبت و در صورت بروز ناهنجاری[[6]](#footnote-6) یا خطا در جریان داده‌ها، آن را گزارش کنند.

تحلیل داده‌های شبکه و تشخیص الگوی رفتاری آن کار آسانی نیست؛ زیرا جریان داده‌ای شبکه متغیری پیوسته در زمان است و فرایند تحلیل و نتیجه‌گیری در این مورد باید بازده قابل‌قبولی داشته باشد. در این پروژه با استفاده از مفاهیمی از علم آمار به نام سری‌های زمانی[[7]](#footnote-7) (TS)، روشی برای تحلیل داده‌های شبکه و دسته‌بندی[[8]](#footnote-8) رفتار آن در بازه‌های زمانی مختلف، جهت ثبت یک الگوی ثابت، ارائه شده است.

## ارزش پروژه

ابزارهایی که در حال حاضر برای پایش وضعیت شبکه وجود دارند صرفاً به جمع‌آوری داده‌های جاری و نمایش نتایج به کاربر به صورت مصور می‌پردازند. همچنین پروژه‌های مشابهی وجود دارد که خروجی آن‌ها با استفاده از ابزارهایی مانند XML تولید می‌شود. در این پروژه، داده‌های جمع‌آوری شده از دید مفاهیم سری زمانی مورد بررسی قرار گرفته و با استفاده از الگوریتم‌های موجود خروجی مورد نظر تولید می‌شود.

مفاهیم سری زمانی ویژگی‌های رفتاری داده‌های ورودی را مشخص می‌کنند؛ سپس با استفاده از ویژگی‌های به‌دست‌آمده، از میان الگوریتم‌های موجود، بهترین الگوریتم جهت تولید خروجی موردنظر انتخاب می‌شود. رویکرد بیان شده به کاهش خطاهای احتمالی پیش‌آمده در روند تحلیل داده‌ها کمک می‌کند و امکان پیش‌بینی رفتار آینده‌ی شبکه را نیز به کاربر می‌دهد.

## هدف پروژه

هدف از انجام این پروژه، آشنایی با مفاهیم موجود در الگوریتم‌های تحلیل سری زمانی و استفاده از آن‌ها در کنار ساختارهای شبکه‌محور کامیپوتری است. با توجه به نیاز جدی نظارت بر رفتار و جریان داده‌های شبکه‌های کامپیوتری و همچنین پیچیدگی زیاد و زمان‌بر بودن استفاده از ابزارهای موجود برای تحلیل داده‌ها، پیاده‌سازی ابزاری که فرایند تحلیل و نتیجه‌گیری و در نهایت تصمیم‌گیری را به صورت خودکار ارائه می‌دهد امری قابل توجه است.

به‌طورکلی این پروژه اهداف زیر را در نظر دارد.

* افزایش بهره‌وری از داده‌های موجود در شبکه جهت شناسایی ناهنجاری‌ها
* ایجاد زمینه برای پیاده‌سازی سیستم‌های تشخیص خطا در شبکه
* دسته‌بندی رفتار شبکه طبق تغییرات جریان داده‌ها طی گذر زمان
* بهره‌گیری از تحلیل‌های آماری در توصیف رویدادهای شبکه

## ساختار پایان‌نامه

این پایان‌نامه شامل هفت فصل است که بخش ششم پیوست‌ها و بخش هفتم منابع را ارائه می‌دهند. در فصل دوم مفاهیم موردنیاز جهت تحلیل و بررسی نتایج اولیه‌ی سری زمانی، روش‌ها و الگوریتم‌های ارائه شده در این زمینه و همچنین تحقیقات انجام‌شده‌ مورد بررسی قرار گرفته‌اند. در بخش سوم بیان تفصیلی روند پروژه به همراه معرفی ابزارهای استفاده شده ارائه شده است. در بخش چهارم نتایج به دست آمده

# فصل دوم: مبانی و مفاهیم تحلیل سری زمانی

## مقدمه

با توجه به گسترش علوم و فنون تحلیل داده‌ها، روش‌های زیادی در زمینه‌های تخصصی مختلف در این زمینه ارائه شده است. در این مورد، تحلیل TSها در دهه‌های اخیر بسیاری از محققان را به خود جذب کرده است. این پروژه با استفاده از مفاهیم موجود در علم تحلیل TSها، داده‌های جمع‌آوری شده از CNها را مورد بررسی قرار می‌دهد.

توضیح دلیل استفاده از TS................................

توضیح زیربخش‌ها و مطالب................................ (ابتدا مدل‌های جمع‌اوری داده سپس مدل‌های تحلیل سپس مدل‌های پیش‌بینی)

در این فصل به بررسی تعاریف و مفاهیم ابتدایی سری‌های زمانی، الگوریتم‌ها و مدل‌های موجود و نکات آن پرداخته شده است. همچنین تحقیقات انجام شده در این زمینه نیز مورد بررسی قرار گرفته و تعاریف بیان‌شده در آن‌ها توضیح داده شده است.

## معرفی سری‌های زمانی (Time Series)

به طور کلی مقادیری از داده که وابسته به زمان هستند و می‌توان طی گذر زمان آن‌ها را جمع‌آوری کرد، یک TS تشکیل می‌دهند. داده‌های TS می‌توانند متعلق به تغییرات دما، تغییرات جمعیت یک منطقه، تغییرات ارزش سهام در بازار بورس و یا داده‌های جمع‌آوری شده از یک CN باشند پس این داده‌ها، بیانگر تغییرات ایجاد شده در یک پدیده در طول زمان را منعکس می‌کنند.

به دلیل وابسته بودن داده‌های TS به زمان، می‌توان یک بردار مانند X در نظر گرفت و سری زمانی را به صورت زیر معرفی کرد:

در این عبارت t بیانگر زمان و X یک متغیر تصادفی است. همان‌طور که در عبارت بیان شده است، زمان صفر نیز قابل استفاده است. این زمان، می‌تواند لحظه‌ی شروع یک پدیده و یا لحظه‌ی شروع جمع‌آوری داده‌های یک پدیده‌ی در جریان باشد. وابسته بودن داده‌های TS به زمان و تغییراتی که منعکس می‌کنند اهمیت ترتیب را در آن‌ها نشان می‌دهد. اگر در هر یک از مراحل جمع‌آوری،‌ تحلیل و پیش‌بینی[[9]](#footnote-9) ترتیب داده‌ها دستخوش تغییر شود، نتایج به دست آمده قابلیت اعتماد ندارند.

### مدل‌های جمع‌آوری داده در سری زمانی

در جمع‌آوری داده‌های TS اگر فقط از یک ویژگی پدیده‌ی موردنظر استفاده شود، متغیر X در عبارت بیان شده یک‌بعدی بوده و مدل سری زمانی را یک متغیره[[10]](#footnote-10) می‌نامند. ولی اگر از چندین ویژگی برای جمع‌آوری داده استفاده شود، به مدل سری زمانی چند متغیره[[11]](#footnote-11) گویند. وابستگی داده‌های TS به زمان امری اساسی است. ولی اگر در کنار ویژگی متغیر بودن با زمان تغییرات مکان و مختصات داده‌ها نیز لحاظ شود، مباحث مورد بحث وارد علم آمار فضایی[[12]](#footnote-12) می‌شوند.

معمولا مرحله‌ی جمع‌آوری داده‌های TS بدون توقف و به صورت پیوسته انجام می‌شود که به آن زمان-پیوسته[[13]](#footnote-13) گویند؛ در غیر این صورت مدل جمع‌آوری داده را زمان-گسسته[[14]](#footnote-14) می‌نامند. از مثال‌های نام برده شده در بخش قبل، تغییرات جمعیت یک منطقه مثالی از مدل زمان-گسسته و تغییرات دما مثالی از مدل زمان-پیوسته هستند. در تحلیل‌های انجام شده معمولا از روش جمع‌آوری زمان-گسسته انجام می‌شود که مقاطع مشخصی از زمان برای آن‌ها در نظر گرفته می‌شود. به طور مثال جمع‌آوری داده‌ها به صورت بازه‌های ساعتی، روزانه، هفتگی، ماهانه و سالانه انجام می‌شود. در نهایت سری‌های زمان-پیوسته قابلیت تبدیل شدن به سری‌های زمان-گسسته را دارند.

در این پروژه مدل جمع‌آوری داده‌های شبکه به صورت یک‌بعدی و زمان-گسسته است. همچنین داده‌های شبکه فقط از نظر تغییرات زمانی قابل تحلیل و بررسی هستند و مباحث علم آمار فضایی در این پایان‌نامه بررسی نمی‌شوند.

## مولفه‌های رفتاری سری زمانی

داده‌های بررسی شده در سری زمانی، داده‌های پیوسته در زمان هستند که با گذر زمان رفتار متفاوتی از خود نشان می‌دهند. همیشه برای شروع تحلیل داده‌های سری زمانی، ابتدا باید ویژگی‌های مشخصی را از رفتار آن‌ها استخراج کرد تا بتوان با توجه به ویژگی‌های رفتاری داده‌ها بهترین مدل تحلیل را انتخاب و در نتیجه دقیق‌ترین پیش‌بینی را ارائه کرد.

توضیح دلیل استفاده از این ویژگی‌ها در این پروژه.....

### روند (Trend)

### الگوهای تناوب (Cyclic Patterns)

### فصلی‌بودن (Seasonality)

### خطا (Error)

### باقی‌مانده (Residual)

### ایستایی (Stationarity)

## پیش‌بینی (Forecasting)

### مدل‌های پیش‌گویی در سری‌های زمانی

# فصل سوم: شرح پروژه

## مقدمه

## جمع‌بندی

# فصل چهارم: نتایج

## مقدمه

## جمع‌بندی

# فصل پنجم: نتیجه‌گیری و پیشنهادات

# پیوست۱

# منابع

[1] Jung, Sangjoon, Chonggun Kim, and Younky Chung. "A prediction method of network traffic using time series models." In International Conference on Computational Science and Its Applications, pp. 234-243. Springer, Berlin, Heidelberg, 2006.

1. Computer Networks [↑](#footnote-ref-1)
2. Monitoring [↑](#footnote-ref-2)
3. Packets [↑](#footnote-ref-3)
4. Simple Network Management Protocol [↑](#footnote-ref-4)
5. Real-Time Monitoring [↑](#footnote-ref-5)
6. Anomaly [↑](#footnote-ref-6)
7. Time Series [↑](#footnote-ref-7)
8. Classification [↑](#footnote-ref-8)
9. Forecasting [↑](#footnote-ref-9)
10. Univariate [↑](#footnote-ref-10)
11. Multivariate [↑](#footnote-ref-11)
12. Spacial Statistics [↑](#footnote-ref-12)
13. Continuous Time [↑](#footnote-ref-13)
14. Discrete Time [↑](#footnote-ref-14)