عنوان پروژه:

بررسی رابطه شدت ترافیک شبکه Packets/s وBytes/s با وقوع حمله DDoS در دستگاه‌های IoT خانه هوشمند

ابزارهای مورداستفاده:

Python (pandas ،scipy ،statsmodels ،matplotlib)

منبع داده:

در این پروژه از داده‌های شبیه‌سازی‌شده (Synthetic) استفاده شده است که باهدف تمرین مفاهیم آماری و ایجاد سناریوی واقع‌گرایانه‌ی ترافیک شبکه IoT در حالت عادی و تحت حمله DDoS تولید شده‌اند.

نام استاد:

دکتر چهکندی

پژوهشگر:

ملیکا باقری

دی ۱۴۰۴

معرفی داده‌ها و متغیرها

حجم داده

- تعداد رکوردها :۵۰۰۰
- برچسب‌ها:
  - Normal: **3342**
  - DDoS: **1658**

متغیرهای طبقه‌ای

1. label ∈ {Normal, DDoS}
2. device_type ∈ {camera, thermostat, light, speaker}

متغیرهای پیوسته

- flow_pkts_s: نرخ بسته‌ها(Packets/s)
- flow_byts_s: نرخ بایت‌ها(Bytes/s)
- flow_duration_s: مدت‌زمان جریان (ثانیه)
- avg_pkt_len: میانگین طول بسته (بایت)

آمار توصیفی

میانگین‌ها به تفکیک برچسب

flow_pkts_s (Packets/s)

- Normal: میانگین ≈ **۷۷.۱۰**، انحراف معیار ≈ **۳۶.۰۷**
- DDoS: میانگین ≈ **۵۴۹.۸۶**، انحراف معیار ≈ **۲۵۷.۷۲**

flow_byts_s (Bytes/s)

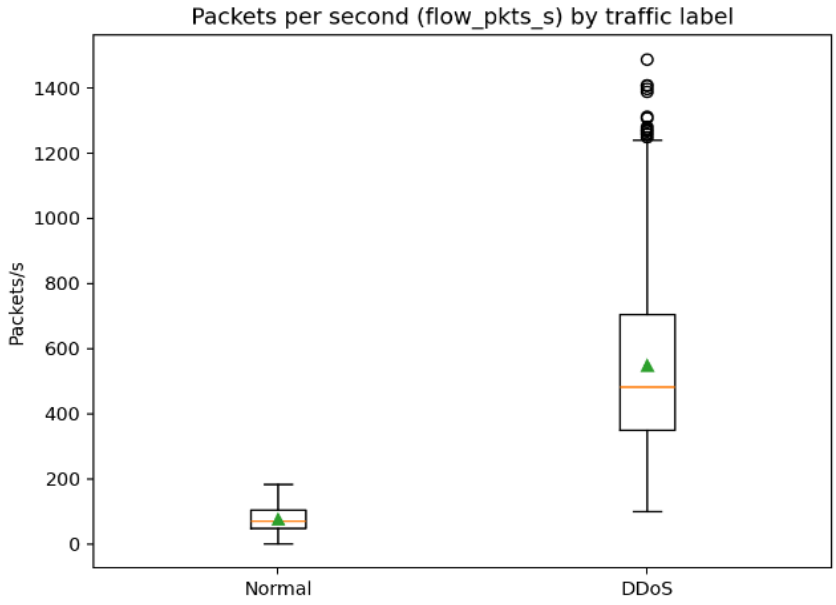
- Normal: میانگین ≈ **۴۵,۱۲۷.۲۶**، انحراف معیار ≈ **۳۲,۵۰۸.۹۸**
- DDoS: میانگین ≈ **۲۳۱,۲۲۵.۵۲**، انحراف معیار ≈ **۱۷۹,۵۸۴.۸۱**

تفسیر :در حالت DDoS هم نرخ بسته‌ها و هم نرخ بایت‌ها به طور چشمگیری بیشتر از حالت عادی است؛ بنابراین انتظار داریم آزمون‌های آماری اختلاف بسیار معنی‌دار نشان دهند.

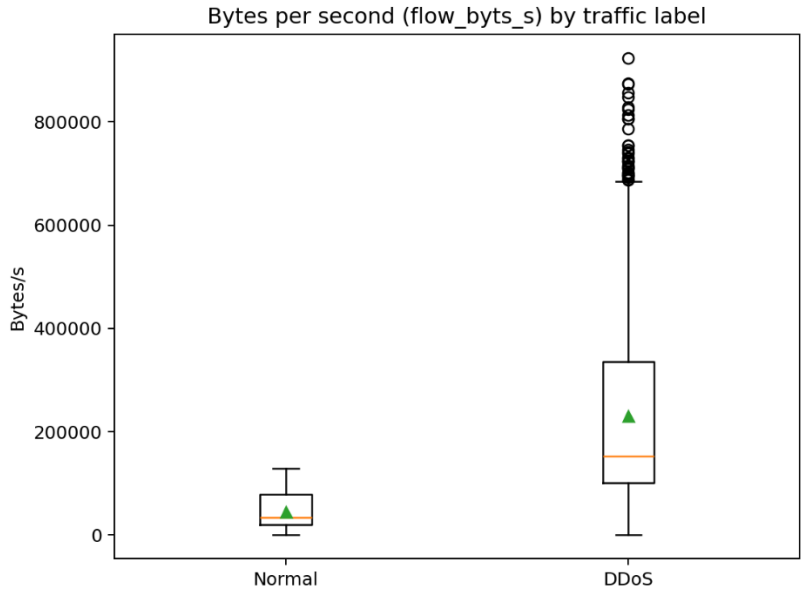
متغیر	گروه	میانگین	انحراف معیار
flow_pkts_s	Normal	۷۷.۱۰	۳۶.۰۷
flow_pkts_s	DDoS	۵۴۹.۸۶	۲۵۷.۷۲
flow_byts_s	Normal	۴۵۱۲۷.۲۶	۳۲۵۰۸.۹۸
flow_byts_s	DDoS	۲۳۱۲۲۵.۵۲	۱۷۹۵۸۴.۸۱

تفسیر جدول ۱:
این جدول نشان می‌دهد که میانگین و پراکندگی شدت ترافیک شبکه در حالت DDoS به طور قابل‌توجهی بیشتر از حالت عادی است که بیانگر رفتار غیرعادی ترافیک در زمان حمله است.

شکل ۱ - توزیع نرخ بسته‌ها (Packets/s) در ترافیک Normal وDDoS



شکل ۱ .این نمودار نشان می‌دهد که توزیع نرخ بسته‌ها در حملات DDoS به طور چشمگیری بالاتر از ترافیک عادی است و تفاوت بین دو حالت به‌وضوح قابل‌مشاهده است.



شکل ۲ .افزایش محسوس Bytes/s در حالت DDoS نشان‌دهنده فشار شدید به شبکه و رفتار سازگار با حملات انکار سرویس است.

بررسی نرمالیتی(Normality)

از آزمون **Shapiro–Wilk**استفاده شد.(α=0.05)

نتایج کلیدی

- برای داده‌های ترکیبی :(All) بیشتر متغیرها نرمال نیستند p-value) بسیار کوچک.(
- به تفکیک گروه‌ها:
  - avg_pkt_lenدر هر دو گروه **Normal**و **DDoS نرمال** است:
    - DDoS: p ≈ **0.596**
    - Normal: p ≈ **0.572**
  - اما flow_pkts_sو flow_byts_sدر بسیاری موارد **نرمال نیستند** p-value)خیلی کوچک.(

تفسیر:
در نمونه‌های بزرگ، Shapiroبسیار حساس است و حتی انحراف‌های کوچک را رد می‌کند. بنابراین در کنار آزمون، بررسی نمودار هیستوگرام QQ-plot/توصیه می‌شود. با این حال، با توجه به حجم نمونه بالا، آزمون‌های) t خصوصاً (Welch معمولاً مقاوم هستند.

فاصله اطمینان(Confidence Interval)

4.1فاصله اطمینان 95٪ برای میانگین $flow\_pkts\_s$ در حالتDDoS

- میانگین**549.8592** :
- CI 95%: **(537.4447 , 562.2737)**
- n=1658

تفسیر :با اطمینان 95٪ میانگین نرخ بسته‌ها در حمله DDoS بین حدود **537تا 562 بسته بر ثانیه** است.

آزمون فرض یک جامعه(One-sample)

سؤال:

آیا میانگین flow_pkts_sدر ترافیک Normal برابر **75**است؟

فرضیه‌ها

- H0: μ = 75
- H1: μ ≠ 75

نتیجه

- آماره**3.3656** ≈ t
- p-value ≈ **0.0007725**

تصمیم

چون **H0** ⇒ p-value < 0.05 **رد می‌شود.**

تفسیر :میانگین نرخ بسته‌ها در ترافیک عادی، به‌صورت معنی‌دار با مقدار 75 تفاوت دارد.

آزمون فرض دو جامعه(Two-sample)

از **Welch t-test**(عدم فرض برابری واریانس‌ها) استفاده شد.

مقایسه flow_pkts_s بین DDoS وNormal

فرضیه‌ها

- H0: μ_DDoS − μ_Normal = 0
- H1: μ_DDoS − μ_Normal ≠ 0

نتیجه

- t ≈ **74.3320**
- p-value: **< 1e-300** (عملاً نزدیک صفر)
- اختلاف میانگین(DDoS − Normal) ≈ **472.7591**
- 95% CIاختلاف: **(460.2846 , 485.2336)**

نتیجه: اختلاف بسیار **معنی‌دار** است.

مقایسه flow_byts_s بین DDoS وNormal

- t ≈ **41.8565**
- p-value ≈ **2.99×10^-264**
- اختلاف میانگین(DDoS − Normal) ≈ **186,098.2612**
- 95% CIاختلاف: **(177,377.8982 , 194,818.6242)**

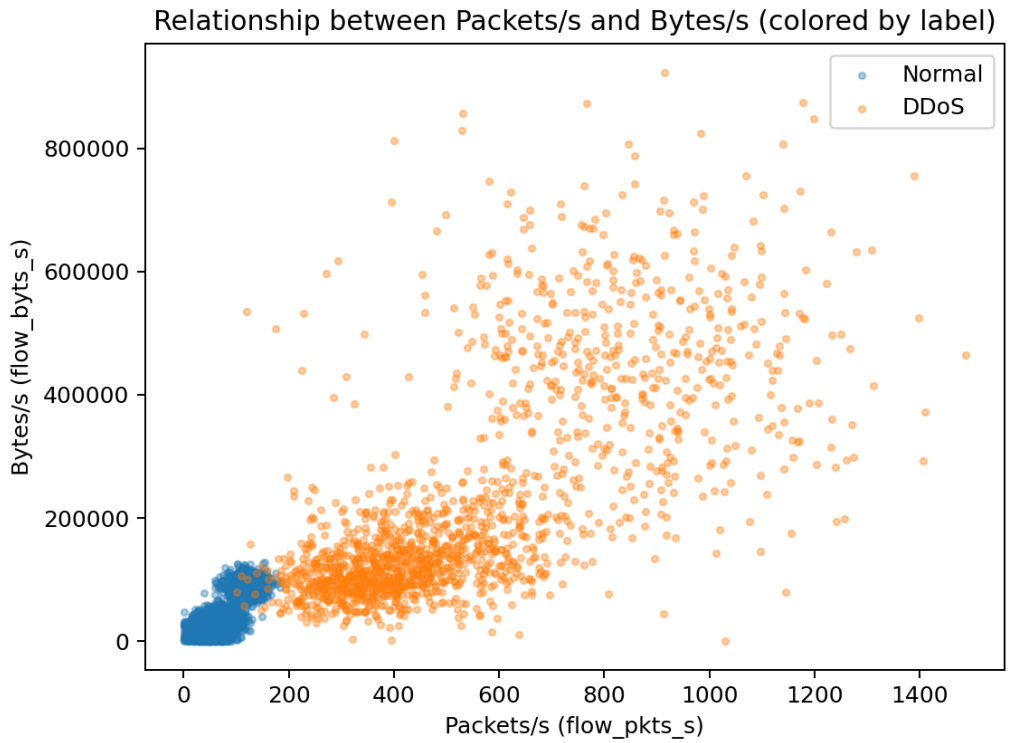
نتیجه: اختلاف بسیار **معنی‌دار** است.

جدول ۲ – نتایج آزمونWelch t-test

متغیر	t	p-value	اختلاف میانگین
flow_pkts_s	74.33	< 1e-300	472.76
flow_byts_s	41.86	2.99×10^{-264}	186098.26

تفسیر جدول ۲:
مقادیر بسیار کوچک p-value نشان می‌دهد که اختلاف بین ترافیک عادی و DDoS از نظر آماری کاملاً معنی‌دار است و احتمال تصادفی بودن این اختلاف عملاً صفر است.

نمودار پراکنش Packets/s و Bytes/s به تفکیک نوع ترافیک



این نمودار رابطه بین نرخ بسته‌ها (Packets/s) و حجم داده ارسالی (Bytes/s) را نشان می‌دهد. مشاهده می‌شود که نقاط مربوط به ترافیک DDoS عمدتاً در مقادیر بالاتر هر دو متغیر قرار گرفته‌اند و تفکیک نسبی بین ترافیک عادی و حمله قابل‌مشاهده است. **این الگو نشان می‌دهد که شدت ترافیک شبکه می‌تواند مبنای مناسبی برای تشخیص حملات DDoSباشد و انگیزه‌ای برای استفاده از مدل‌های رگرسیونی فراهم می‌کند.**

تحلیل واریانس(ANOVA)

سؤال:

آیا میانگین flow_pkts_sدر حالت DDoS بین انواع دستگاه‌ها (device_type) متفاوت است؟

فرضیه‌ها

- H0:میانگین همه گروه‌ها برابر است.
- H1:حداقل یک میانگین متفاوت است.

نتایج DDoS فقط

میانگین flow_pkts_sدر:DDoS

- camera: **843.1161**
- speaker: **534.2162**
- thermostat: **426.4276**
- light: **312.1506**

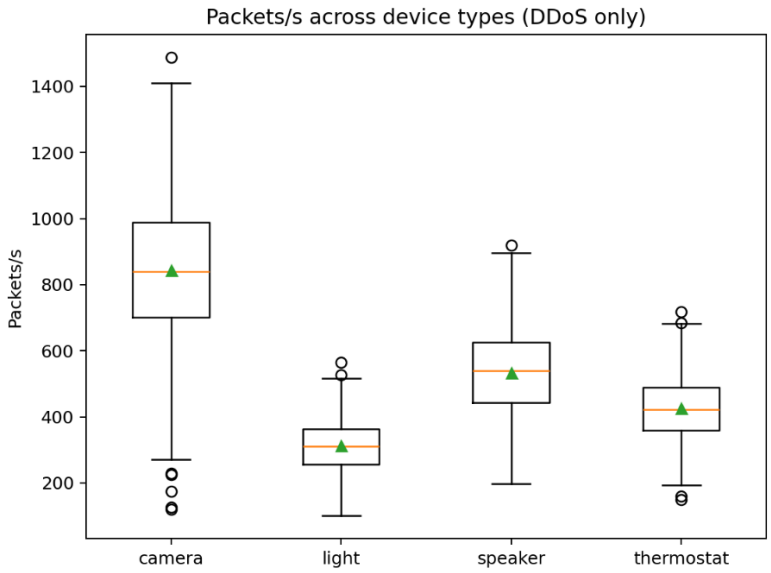
آزمون:ANOVA

- F ≈ **1129.2150**
- p-value: **< 1e-300**

نتیجه **H0 :رد می‌شود** ⇒میانگین‌ها بین دستگاه‌ها یکسان نیست.

آزمون تعقیبی(Tukey HSD)

آزمون Tukey نشان می‌دهد اختلاف‌ها بین اکثر جفت‌های device_type معنی‌دار است



شکل ۳ .این نمودار نشان می‌دهد که شدت ترافیک در حملات DDoS به نوع دستگاه IoT وابسته است و برخی دستگاه‌ها ترافیک سنگین‌تری تولید می‌کنند.

جدول ۳ – نتایج تحلیل واریانس (ANOVA)

آماره F	p-value
۱۱۲۹.۲۱	< 1e-300

نتایج ANOVA نشان می‌دهد میانگین Packets/s در انواع مختلف دستگاه‌ها یکسان نیست و نوع دستگاه نقش معنی‌داری در شدت حمله دارد.

رگرسیون

رگرسیون لجستیک برای پیش‌بینیDDoS

مدل :
[
is_ddos \sim flow_pkts_s + avg_pkt_len + device_type
]

نتایج کلیدی(p-value)

- flow_pkts_s:ضریب**0.1183** ≈ ، **7.48×10^-11** ≈ pمعنی‌دار
- avg_pkt_len: p ≈ **0.109** در این مدل معنی‌دار نیست
- برخی سطوح device_typeنیز معنی‌دار شدند.

تفسیر قابل‌فهم(Odds Ratio)

برای افزایش **۱۰ واحد** در flow_pkts_s:

- OR ≈ **3.2636**

یعنی اگر نرخ بسته‌ها ۱۰ واحد بیشتر شود، **شانس DDoS شدن** حدود **۳.۲۶ برابر** می‌شود (با ثابت‌بودن سایر متغیرها).

جدول ۴ - ضرایب مدل رگرسیون لجستیک

متغیر	ضریب	p-value
flow_pkts_s	0.1183	7.48×10^{-11}
avg_pkt_len	—	0.109

نتایج نشان می‌دهد نرخ بسته‌ها مهم‌ترین عامل پیش‌بینی وقوع حمله DDoS است.

خلاصه آزمون‌های آماری و فرض‌های مورد بررسی:

آزمون آماری	متغیر(ها)	هدف آزمون	فرض صفر (H_0)	فرض مقابل (H_1)	نتیجه
Shapiro–Wilk	flow_pkts_s، flow_byts_s، avg_pkt_len	بررسی نرمال بودن داده‌ها	داده‌ها از توزیع نرمال پیروی می‌کنند	داده‌ها نرمال نیستند	برای برخی متغیرها H_0 رد شد
One-sample t-test	flow_pkts_s (Normal)	مقایسه میانگین با مقدار مرجع	$\mu = 75$	$\mu \neq 75$	H_0رد شد
Welch t-test	flow_pkts_s	مقایسه میانگین Normalو DDoS	$\mu DDoS = \mu Normal$	$\mu DDoS \neq \mu Normal$	H_0رد شد
Welch t-test	flow_byts_s	مقایسه میانگین Normalو DDoS	$\mu DDoS = \mu Normal$	$\mu DDoS \neq \mu Normal$	H_0رد شد
ANOVA	flow_pkts_s (DDoS)	مقایسه میانگین بین انواع دستگاه	میانگین همه گروه‌ها برابر است	حداقل یک میانگین متفاوت است	H_0رد شد
Tukey HSD	flow_pkts_s (DDoS)	شناسایی تفاوت بین گروه‌ها	اختلاف میانگین‌ها صفر است	اختلاف معنادار وجود دارد	اختلاف معنادار مشاهده شد
Logistic Regression	flow_pkts_s، avg_pkt_len، device_type	پیش‌بینی وقوعDDoS	ضرایب = 0	حداقل یک ضریب ≠ 0	flow_pkts_s معنادار

تفسیر: این جدول خلاصه‌ای از آزمون‌های آماری مورد استفاده در پروژه، اهداف هر آزمون و نتایج حاصل از آن‌ها را ارائه می‌دهد. استفاده از این جدول به درک بهتر مسیر تحلیل آماری و ارتباط بین آزمون‌ها و فرضیات پژوهش کمک می‌کند

محدودیت ها:

داده‌های مورد استفاده در این پژوهش به‌صورت شبیه‌سازی‌شده (Synthetic) تولید شده‌اند و هدف اصلی از به‌کارگیری آن‌ها، تمرین مفاهیم آماری و تحلیل رفتار ترافیک شبکه در سناریوهای کنترل‌شده بوده است. هرچند تلاش شده توزیع متغیرها و الگوهای ترافیکی تا حد ممکن به شرایط واقعی شبکه‌های IoT نزدیک باشد، اما این داده‌ها لزوماً تمام پیچیدگی‌ها، نویزها و رفتارهای غیرقابل پیش‌بینی موجود در داده‌های واقعی را منعکس نمی‌کنند.

بنابراین، نتایج به‌دست‌آمده در این پروژه بیشتر جنبه آموزشی و تحلیلی داشته و تعمیم مستقیم آن‌ها به محیط‌های عملیاتی واقعی باید با احتیاط انجام شود. در پژوهش‌های آتی، استفاده از داده‌های واقعی شبکه می‌تواند اعتبار و کاربردپذیری نتایج را به‌طور قابل توجهی افزایش دهد.

جمع‌بندی و نتیجه‌گیری:

در این پروژه، رابطه بین شدت ترافیک شبکه و وقوع حملات DDoS در محیط دستگاه‌های IoT خانه هوشمند مورد بررسی آماری قرار گرفت. نتایج به‌دست‌آمده از آمار توصیفی، آزمون‌های فرض آماری و تحلیل‌های پیشرفته نشان داد که متغیرهای مرتبط با شدت ترافیک، به‌ویژه نرخ بسته‌ها (flow_pkts_s) و نرخ بایت‌ها(flow_byts_s) ، در حالت وقوع حمله DDoS به‌صورت معناداری افزایش می‌یابند.

آزمون‌های مقایسه‌ای بین ترافیک عادی و ترافیک تحت حمله نشان دادند که این اختلاف‌ها از نظر آماری بسیار معنادار هستند و احتمال تصادفی بودن آن‌ها عملاً ناچیز است. همچنین نتایج تحلیل واریانس بیانگر آن بود که نوع دستگاه IoT می‌تواند بر شدت ترافیک در شرایط حمله تأثیرگذار باشد، به‌طوری‌که برخی دستگاه‌ها ترافیک سنگین‌تری را تجربه می‌کنند.

علاوه بر این، مدل رگرسیون لجستیک نشان داد که نرخ بسته‌ها یکی از مهم‌ترین متغیرهای پیش‌بینی‌کننده وقوع حمله DDoS است و افزایش آن می‌تواند به‌طور قابل توجهی احتمال وقوع حمله را افزایش دهد. این یافته‌ها نشان می‌دهند که شاخص‌های شدت ترافیک شبکه می‌توانند مبنای مناسبی برای طراحی سیستم‌های تشخیص حمله در شبکه‌های IoT باشند.

با این حال، با توجه به شبیه‌سازی‌شده بودن داده‌ها، تعمیم نتایج به محیط‌های واقعی شبکه باید با احتیاط انجام شود. در پژوهش‌های آینده، استفاده از داده‌های واقعی و به‌کارگیری مدل‌های یادگیری ماشین پیشرفته می‌تواند به بهبود دقت تشخیص و افزایش کاربردپذیری نتایج کمک کند.
