به نام خدا

عنوان پروژه:

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

ابزارهای مورد استفاده:

Python (pandas ،scipy ،statsmodels ،matplotlib)

منبع داده:

در این پروژه از داده‌های شبیه‌سازی‌شده (Synthetic) استفاده شده است که با هدف 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**5000** :
- 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: مدت زمان جریان (ثانیه)
- avg_pkt_len: میانگین طول بسته (بایت)

آمار توصیفی

میانگین‌ها به تفکیک برچسب

flow_pkts_s (Packets/s)

- Normal: میانگین **77.10** ≈ ، انحراف معیار **36.07** ≈
- DDoS: میانگین **549.86** ≈ ، انحراف معیار **257.72** ≈

flow_byts_s (Bytes/s)

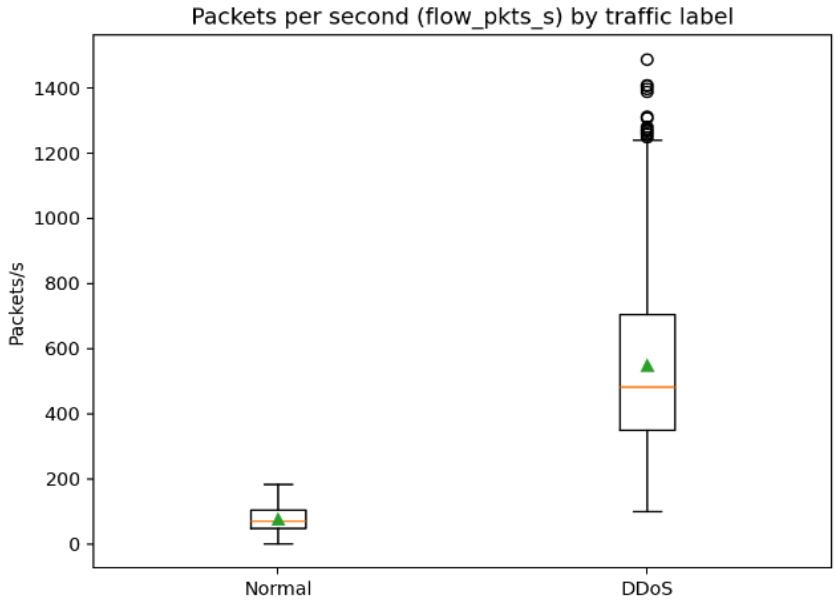
- Normal: میانگین **45,127.26** ≈ ، انحراف معیار **32,508.98** ≈
- DDoS: میانگین **231,225.52** ≈ ، انحراف معیار **179,584.81** ≈

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

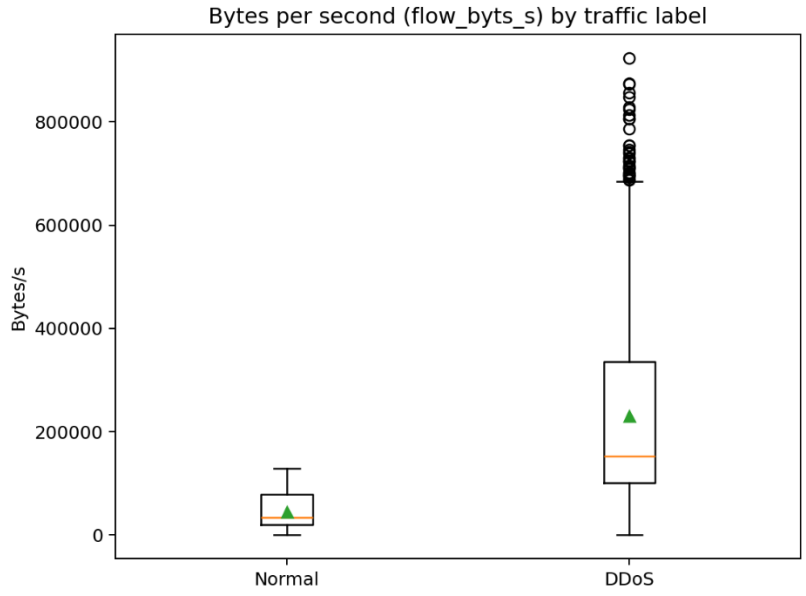
متغیر	گروه	میانگین	انحراف معیار
flow_pkts_s	Normal	77.10	36.07
flow_pkts_s	DDoS	549.86	257.72
flow_byts_s	Normal	45127.26	32508.98
flow_byts_s	DDoS	231225.52	179584.81

تفسیر جدول 1:
این جدول نشان می‌دهد که میانگین و پراکندگی شدت ترافیک شبکه در حالت DDoS به‌طور قابل توجهی بیشتر از حالت عادی است که بیانگر رفتار غیرعادی ترافیک در زمان حمله می‌باشد.

شکل 1 – توزیع نرخ بسته‌ها (Packets/s) در ترافیک Normal وDDoS



شکل 1 .این نمودار نشان می‌دهد که توزیع نرخ بسته‌ها در حملات DDoS به‌طور چشم‌گیری بالاتر از ترافیک عادی است و تفاوت بین دو حالت به‌وضوح قابل مشاهده است.



شکل 2 .افزایش محسوس Bytes/s در حالت DDoS نشان‌دهنده فشار شدید به شبکه و رفتار سازگار با حملات انکار سرویس می‌باشد.

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_sدر 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)**

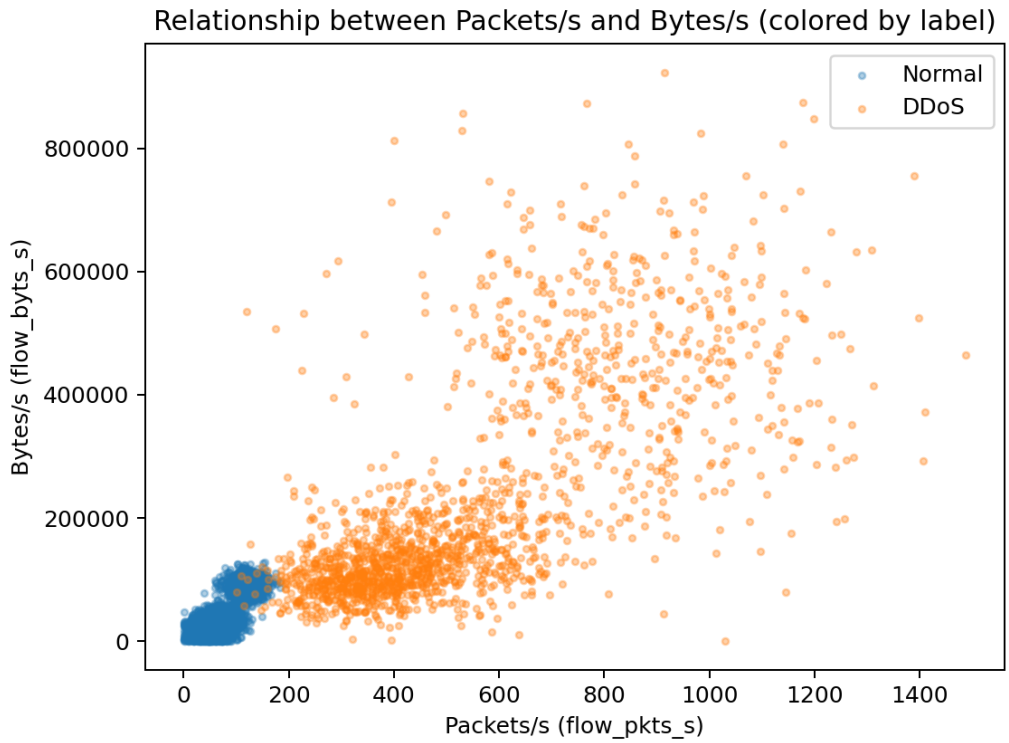
نتیجه: اختلاف بسیار **معنی‌دار** است.

جدول 2 – 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 2:
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 قابل مشاهده است. **این الگو نشان می‌دهد که شدت ترافیک شبکه می‌تواند مبنای مناسبی برای تشخیص حملات DDoS باشد و انگیزه‌ای برای استفاده از مدل‌های رگرسیونی فراهم می‌کند.**

تحلیل واریانس(ANOVA)

سؤال:

آیا میانگین flow_pkts_s در 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_s در DDoS:

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

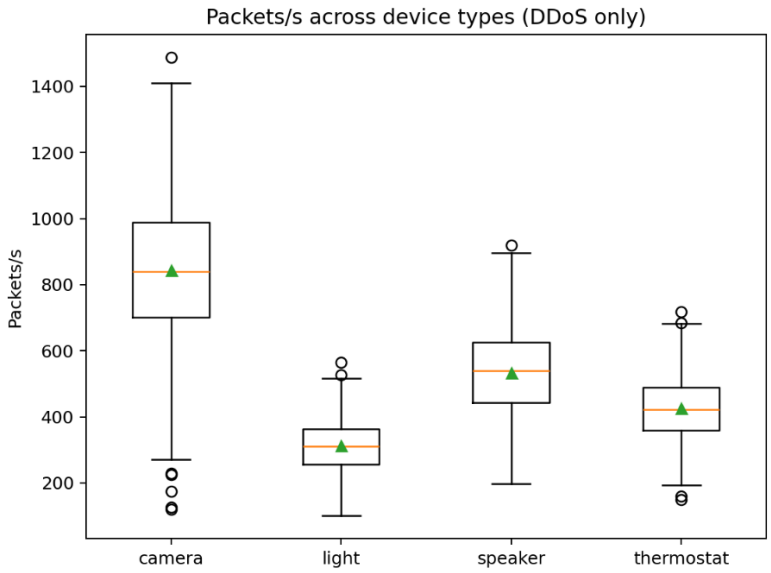
آزمون ANOVA:

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

جدول 3 – نتایج تحلیل واریانس (ANOVA)

آماره F	p-value
1129.21	< 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 قابل فهم(Odds Ratio)

برای افزایش **10واحد** در flow_pkts_s:

- OR ≈ **3.2636**

یعنی اگر نرخ بسته‌ها 10 واحد بیشتر شود، **شانس DDoS شدن** حدود **3.26برابر** می‌شود (با ثابت بودن سایر متغیرها).

جدول 4 – 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.

جمع‌بندی و نتیجه‌گیری

1. ترافیک DDoS از نظر **Packets/s** و **Bytes/s** به‌صورت بسیار معنی‌دار از ترافیک Normal بیشتر است-p value‌ها تقریباً صفر
2. برای `flow_pkts_s` در DDoS فاصله اطمینان 95٪ نشان داد میانگین نرخ بسته‌ها حدود **537** تا **562** است.
3. ANOVA نشان داد در شرایط DDoS ، نوع دستگاه روی میانگین نرخ بسته‌ها اثر دارد و بین دستگاه‌ها اختلاف معنی‌دار وجود دارد.
4. رگرسیون لجستیک نشان داد `flow_pkts_s` یک پیش‌بین بسیار قوی برای رخداد DDoS است.