# Signature-Based Traffic Classification and Mitigation for DDoS Attacks Using Programmable Network DataPlanes

IEEE ACCESS

University of Athens

Aug 2021

3 citations

---

قوانین فیلترینگ برای مقابله با حملات منع خدمت توزیع شده بر پایه IP های مبدا یا جریان می باشند اما این روش مشکل مقیاس پذیری در حملات توزیع شده و حجیم امروزی (که با استفاده از IP های مبدا زیادی صورت‌می‌گیرد) دارد. روشی که ما ارایه می‌دهیم با استفاده از یادگیری ماشین نظارتی[1] امضای[2] بسته‌های مهاجم را به دست می‌اورد و قوانین[3] فیلترینگ مربوط به امضا را نیز تولید می‌کند. برای تسریع پردازش از ماژولهای میانی که از XDP استفاده می‌کنند به عنوان DPI های برنامه‌پذیر استفاده می‌کنیم. الگوهایی که استخراج می‌کند از ترکیب چندین ویژگی[4] شاخص پکت‌ها به دست آمده اند و متقاعباً به الگوریتم‌های یادگیر ماشین داده می‌شوند تا آنها را به عناون امضای عادی یا مهاجم طبقه بندی[5] کند. امضاهای متخاصم تحت یک عملیات کاهش متناسب با بردار حمله[6] (بنا به اپلیکیشن می توانند تغییر یابند) تعیین می‌شوند و سپس براساس آن امضاها، مجموعه‌ای مختصر از قوانین فیلتر تولید کنند، که منجر به تسریع رفع مخاطره[7] می‌شود. روش مان را نیز با یک حمله حجیم DNS با معیار دقت طبقه بندی امضاها و نرخ فیلترینگ بسته‌ها ارزیابی کرده‌ایم. آزمایش‌های ما بر اساس دیتاست ترافیک‌های معمولی و مهاجم ضبط شده در محیط واقعی می باشد. رویکردما با روش های مبتنی بر مبدا با معیارهای شناسایی ترافیک‌های مهاجم، کاردینالیتی قوانین فیلتر و نرخ گذر مورد نیاز پردازش بسته‌ها در شبکه‌های امروزی مقایسه شده‌است. نتیجه می شود که روش مبتنی بر امضای ما از روش‌های شبیه به روش مبتنی بر IP دقت تشخیص بالاتری خواهد داشت.

طریقه کار: به طورخلاصه روش‌ما ترافیک را مدام مانیتور می کند و یک امضا از ترافیک براساس ویژگی‌های[8] بردار حمله (انواع حملات مختلف) ارایه می‌دهد. امضاها با استفاده از روش‌های نظارتی یادگیری ماشین که به دنبال ویژگی‌ها (فیلدها)ی متمایز هستند ، طبقه بندی میشود. این مدل‌ها نیز از قبل با یکسری دیتاست (بسته به نوع حمله و کاربرد شبکه) ترافیک مهاجم و معمولی آزمایش داده شده ، سپس این امضاها طی یک عملیات کاهش می‌یابند و سپس قوانین فیلتری را براساس آن‌ها می‌سازند و بر روی middle box های مراکز اسکراب (XDP) قرار می‌دهند تا به تسریع روند مخاطره کمک کند.

---

[1] Supervised machine learning
[2] Signature
[3] rules
[4] Packet feture
[5] classify
[6] Attack vector
[7] mitigation
[8] Packet features(fields)

تعاریف:

Programmable data-planes (eXpress Data Path): راهکارهایی مثل p4 به ما این اجازه را می‌دهند تا پایپ لاین پردازشی یک المان شبکه را در خور کاربردآن برنامه ریزی کنند (بدون سربار اضافی از لایه کنترلر).XDP یک data-plane نرم افزاری است که قبل از هرعملیات سنگین مربوط به شبکه اجرا می‌شود و امکان پردازش سریع بسته در سخت افزارهای COTS[9] (NIC های ارزان دارند) را فراهم می‌آورد (به جای استفاده از پلتفرم‌های اختصاصی) . برنامه های XDP یا با زبان C نوشته شده‌اند و یا اینکه در محیط کارت شبکه ویا به طور مستقیم روی NICها (Netronome SmartNICs) اجرا می‌شوند. اطلاعات بسته‌هایی که پردازش می‌کنند در محیطهای خاصی از حافظه به نام BPF[10] به صورت کلید-مقدار نگاشت[11] می‌شود. XDP برای هر بسته یک اقدام متناظر تعریف می‌کند:

- XDP_DROP : دورانداختن
- XDP_PASS : ارسال به استک شبکه
- XDP_REDIRECT : هدایت به یک اینترفیس دیگر
- XDP_TX : انتقال

در این مقاله از XDP برای برنامه ریزی مانیتورینگ و فیلتر بسته‌ها استفاده می کنیم. دقت شود که این برنامه‌ها باید چند شرط زیر را داشته باشند: bounded loops,fixed size data structure,4096 BPF instructions per program,limited support of ketrnel fucntions .

روشی مبتنی بر امضا:روشهای طبقه بندی و فیلتر ترافیک که تا به حال ارایه شده بر مبنای جریان یا بر مبنای امضا می‌باشند:

- مبتنی بر جریان: مبتنی بر خصوصیات جریان طبقه بندی می‌کنند
- مبتنی بر امضا: بیشتر در سامانه‌های تشخیص نفوذ استفاده می‌شود. در برابر حملات روز صفر ناایمن است. ابزاری به نام DeepDefense [22] ارایه شد که خصوصیاتی با دفعات تکراری بیش از بقیه را شناسایی می‌کرد. همچنین راهکار برمبنای RNN می‌باشد که ترافیک را در پنجره‌های لغزان زمانی جمع آوری می‌کند و در آرایه ای از خصوصیات[12] خلاصه می‌کند و سپس به RNN ارسال می‌کند. همچنین Cloudflare ابزاری برای تشخیص بی‌نظمی[13] بر پایه امضا به صورت سرویس ارایه می‌دهد. طبق اطلاعات ما این روش‌ها به صورت عمومی منتشر نشده‌اند و مقایسه با آنها امکان پذیر نمی‌باشد.

روشی که ما ارایه می‌دهیم نیز قابلیت شناسایی حملات روز صفر را دارد.

## معماری:

[9] Commercial off the Shelf
[10] Berkely Packet Filter
[11] mapping
[12] Feature vecor
[13] anomaly

از چهارتا کامپوننت اصلی تشکیل شده‌است و تمامی این چهار مرحله اصلی به صورت پیوسته در بازه های زمانی[14] ۱۰ ثانیه‌ای اجرا می‌شوند.

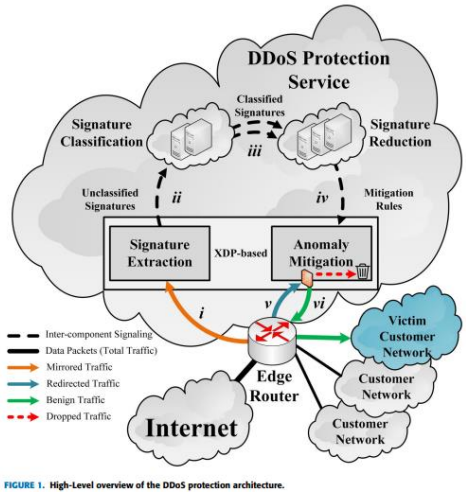

FIGURE 1. High-Level overview of the DDoS protection architecture.

Signature Extraction: بر اساس خصوصیات بردارهای ترافیکی مختلف (بعد از این که ترافیک‌های معمولی و مهاجم را بررسی کردیم و ویژگی‌ها-فیلدهای مهمشان را به دست آوردیم) ویژگی‌های مربوطه را از ترافیک عبوری استخراج می‌کند. و مدام ترافیک را توسط XDP ها پردازش می‌کند (علاوه‌بر XDP از هر متود دیگری که امکان دسترسی به فیلدهای بسته را داشته باشد می‌توان استفاده کرد).

نحوه انتخاب ویژگی‌ها درذیل توضیح داده شده است: به صورت کلی اینچنین است که بخش های اصلی سرایند بسته‌های پروتکل سواستفاده شده را انتخاب می‌کنیم.

از هر دو مجوعه داده‌های ترافیک عادی و مهاجم مربوط به یک بردار حمله برای آموزش یک Random Forest classifer استفاده می‌کنیم.

مجموعه ویژگی‌ها را با یک بردار امضای $X=[x_1x_2...x_m]^T$ نشان می‌دهیم. هر امضای واحد X متناظر یک سطر در جدول مانیتور دیتا در شکل زیر است. هر امضای بسته مشاهده شده مربوط به یک شمارنده است که در یک نقشه BPF مناسب (به عنوان مثال جدول هش) ذخیره شده است.

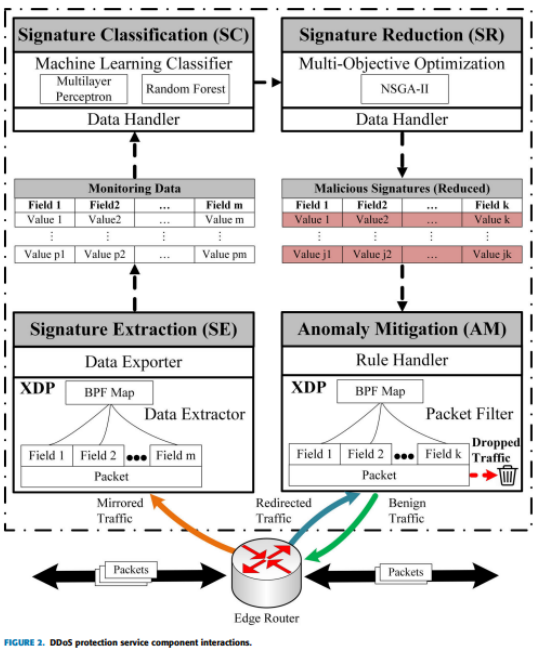

FIGURE 2. DDoS protection service component interactions.

این ماژول شامل چندین نمونه[15] می‌باشد که هریک مربوط به یک بردار حمله می‌باشند و شامل دو بخش زیر می‌باشند:

- Data Extractor: یک برنامه data-plane XDP در کرنل که مقادیر سرایند بسته‌ها برای فیلدهای انتخابی را به همراه آدرس ISP (که برای شناسایی مقتول و ارسال آن به بخش اسکراب-anomaly mitigation می‌باشد) ذخیره می‌کند.

[14] Time-inteval
[15] instance

- Data Exporter: برنامه‌ای در فضای کاربر که امضاهای نگاشت شده BPF را به صورت یکجا به SC ارسال می‌کند.

Signature Classification: یک ماژول control-plane می باشد و اطلاعات امضاها بهش داده می‌شود و آنها را به کمک مدلهای یادگیری ماشین نظارتی دسته بندی می‌کند(مهاجم یا عادی). و از امضاهای طبقه بندی شده برای تولید خط قانون استفاده می‌کند. از دوبخش data handler (بخش پیش پردازش مجموعه ویژگی‌های دریافتی توسط ماژول SE و نرمال سازی آنها) و MLclassifier تشکیل شده‌است.

Signature Reduction: این پروسه به صورت یک مسیله بهینه سازی چند هدفی (Patero) مدل می‌شود(با هدف کمترین تعداد امضا و همچنین به حداقل رساندن نرخ دورانداختن ترافیک‌های معمولی). وبه کمک الگوریتم مرتب‌سازی ژنتیک، جواب بهینه یا همان زیر مجموعه حداقلی از ویژگی‌ها را در یک زمان محدود(ایده‌ال) می‌یابد. و تعداد امضاها را کاهش می‌دهد و آن امضاها را در BPF map های XDP ذخیره می‌کند و می‌توان بدین صورت packet matching در data-plane را انجام داد.

Anomaly Mitigation:همانند یک فایروال اسکراب است که در data-plane عمل می‌کند (XDP) و مجموعه امضاهای کاهش یافته به آن داده می شود. پکت‌های با مقصد مقتول به صورت توزیع شده به آن هدایت می‌شوند(مرحله v) و آنها را دور میاندازد درحالی که بسته های عادی را به مسیریاب شبکه برمی‌گرداند (مرحلهvi). از دوبخش تشکیل شده‌است:

- Rule Handler: لیستی از امضاهای مخرب مرتبط با IP قربانی را دریافت می کند، آنها را به عنوان قوانین فیلتر در BPF map نصب می‌کند.
- packet Filter: یک برنامه XDP در فضای کرنل که شبیه به data extractor در SE می‌باشد. ترافیک‌های به مقصد IP مقتول را دریافت می‌کند و فیلدهایشان بر اساس مجموعه امضاهای کاهش یافته استخراج می‌کند و مقادیرشان را با مجوعه ویژگی‌های قوانین فیلتر در BPF مقایسه می‌شوند. اگر تطابقی رخ دهد که توسط XDP_DROPدورانداخته می‌شود ویا در غیراینصورت با XDP_TX به مسیریاب لبه شبکه ارسال می‌شود.

## ارزیابی:

اهداف ما بررسی کامپوننت‌های مختلف راهکارمان، بررسی الگوریتم طبقه بندی استفاده شده در SC و همچنین مقایسه روش مبتنی بر امضایمان با روشهای مبتنی بر IP و جریان (با معیارهای: قابلیت شناسایی و فیلتر ترافیک‌های متخاصم، تعداد - قوانین فیلتر موردنیاز، عملکرد رفع مخاطره-فیلترینگ بسته‌ها) با استفاده از ترافیک های واقعی و تولیدی شامل ترکیبی از ترافیک مهاجم و عادی، می‌باشد.

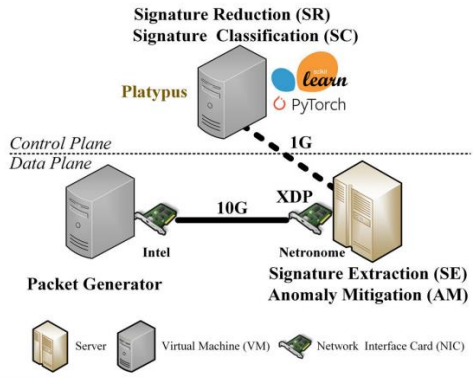


FIGURE 3. Proof-of-concept testbed setup.

**نتیجه گیری:** روش‌ما نه تنها برای DNS های حجیم بلکه برای همه حملات amplification DDoS[16] می‌تواند کارا باشد.

## کارهای آتی:

- پیاده سازی روشهای طبقه بندی که به صورت توام و مرتبط با هم می‌توانند بردارهای حمله را با استفاده از تکنیک‌های یادگیری چند-وظیفه‌ای[17] تشخیص دهند.
- پیاده سازی امکان شناسایی حملات لایه کاربرد با تمرکز بر روی ترافیک رمز شده.

سوالها، پیشنهادات و نکات:

چرا از XDP استفاده می کند؟

DPDK

[16] حملاتی که از پروتکل‌های آسیب پذیر استفاده می‌کنند تا حجم ترافیک زیادی را تولید کرده و روانه دستگاه مقتول کنند.

[17] Multi-task learning techniques