IXP Scrubber: Learning from Blackholing Traffic for ML-Driven DDoS Detection at Scale

Matthias Wichtlhuber* | Eric Strehle[+] | Daniel Kopp* | Lars Prepens* | Stefan Stegmüller* | Alina Rubina* | *Christoph Dietzel** | Oliver Hohlfeld[+]

*DE-CIX | [+]Brandenburg University of Technology, Cottbus

rnd-team@de-cix.net | oliver.hohlfeld@b-tu.de

DE CIX

b-tu Brandenburgische Technische Universität Cottbus - Senftenberg

Bundesministerium für Bildung und Forschung

Based on: M. Wichtlhuber, E. Strehle, D. Kopp, L. Prepens, S. Stegmueller, A. Rubina, C. Dietzel, and O. Hohlfeld. 2022. „IXP Scrubber: Learning from Blackholing traffic for ML-driven DDoS detection at scale“. ACM SIGCOMM 2022 Conference (SIGCOMM '22). https://doi.org/10.1145/3544216.3544268

Motivation

Distributed Denial of Service (DDoS)

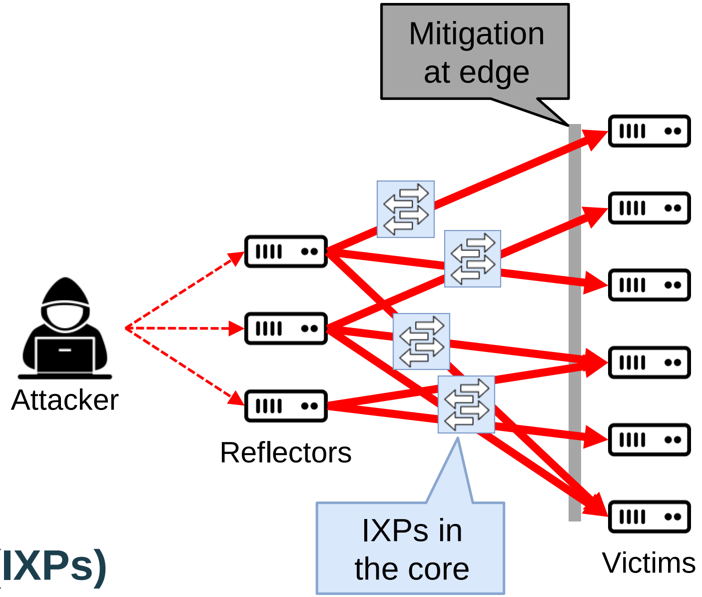
- Millions of attacks/day globally @14% compound annual growth+
- Peak is 3.5 Tbps [53], average is 1 Gbps [63]
- Frequent problem for network operators

+Cisco Annual Internet Report (2018-2023)

Motivation

Distributed Denial of Service (DDoS)

- Millions of attacks/day globally @14% compound annual growth+
- Peak is 3.5 Tbps [53], average is 1 Gbps [63]
- Frequent problem for network operators



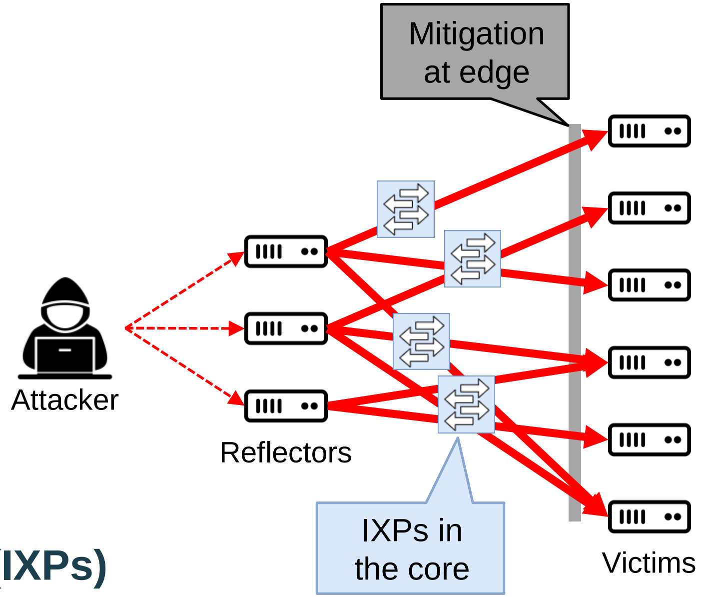
Mitigation at Internet Exchange Points (IXPs)

- Stopping DDoS at IXPs: *2+ AS hops earlier in ~55% of attacks* [55]
- Removes stress from the infrastructure, simplifies complex DDoS traffic analysis

Motivation

Distributed Denial of Service (DDoS)

- Millions of attacks/day globally @14% compound annual growth+
- Peak is 3.5 Tbps [53], average is 1 Gbps [63]
- Frequent problem for network operators



Mitigation at Internet Exchange Points (IXPs)

- Stopping DDoS at IXPs: *2+ AS hops earlier in ~55% of attacks* [55]
- Removes stress from the infrastructure, simplifies complex DDoS traffic analysis

Can we build a DDoS mitigation system fitting IXPs‘ operational requirements?

+Cisco Annual Internet Report (2018-2023)

Operational Requirements

Low cost

- no appliances, needs to work with existing hardware

Low maintenance

- no manual definition of rules and triggers, high degree of automation

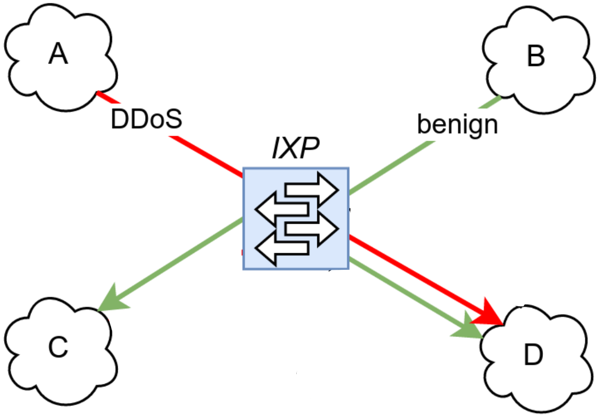
Member-driven

- IXP members define what DDoS is and what they want to filter

Controllable

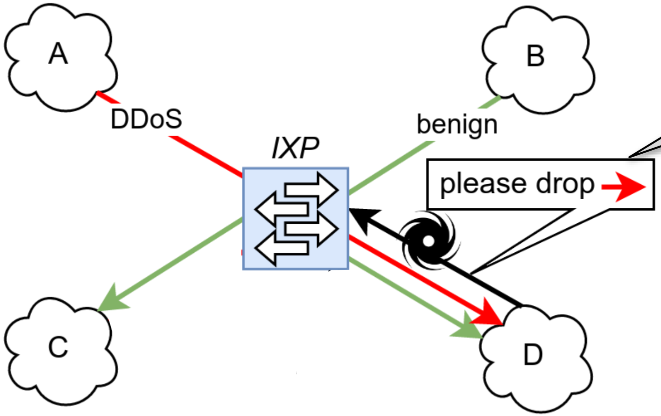
- limit possible damage of false positives, understand performance limitations

IXP Scrubber: Contributions

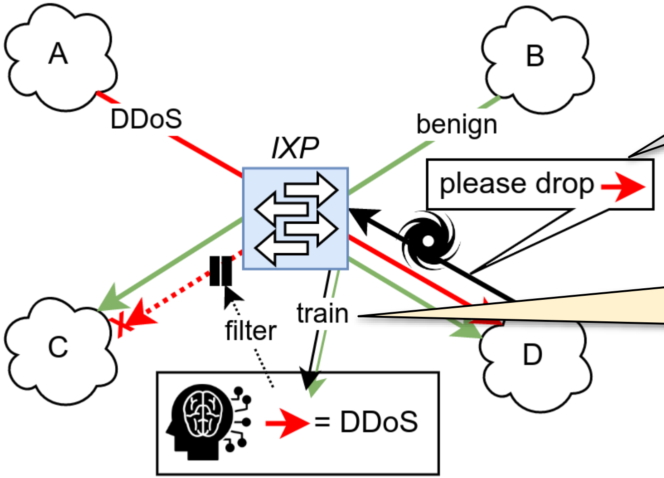


IXP Scrubber: Contributions

ASes mark DDoS traffic via BGP (blackholing [36])



IXP Scrubber: Contributions

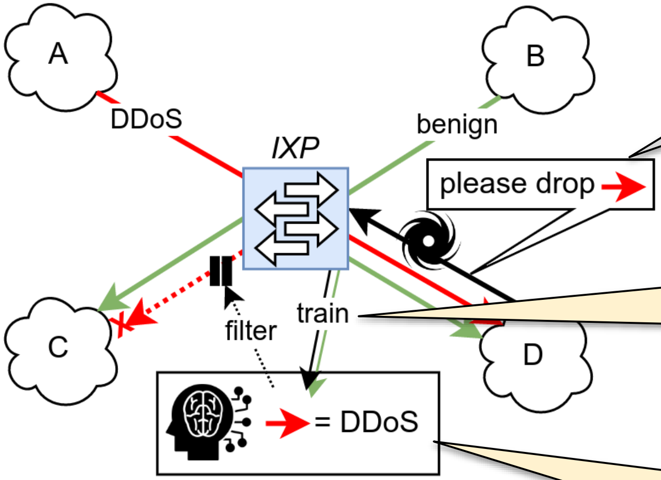


ASes mark DDoS traffic via BGP (blackholing [36])

Contribution 1: large DDoS training sets from blackholing traffic

IXP Scrubber: Contributions

ASes mark DDoS traffic via BGP (blackholing [36])

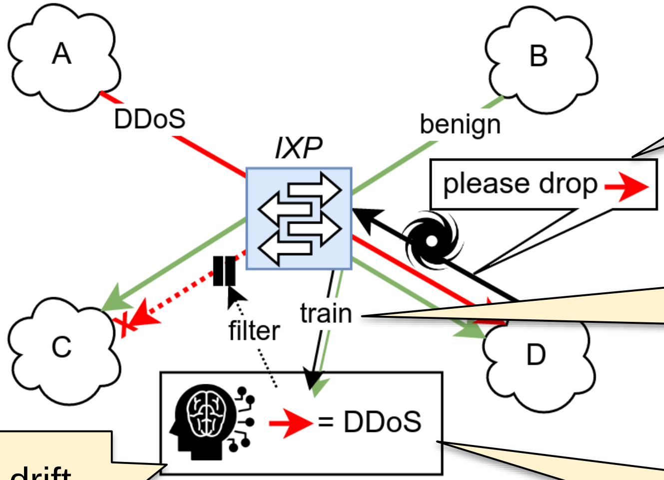


Contribution 1: large DDoS training sets from blackholing traffic

Contribution 2: self-learning, locally explainable IXP-scrubber ML-model

IXP Scrubber: Contributions

ASes mark DDoS traffic via BGP (blackholing [36])

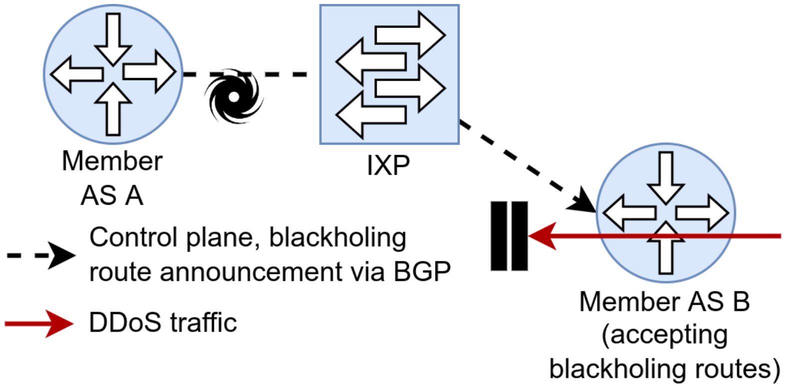


<u>Contribution 1:</u> large DDoS training sets from blackholing traffic

<u>Contribution 2:</u> self-learning, locally explainable IXP-scrubber ML-model

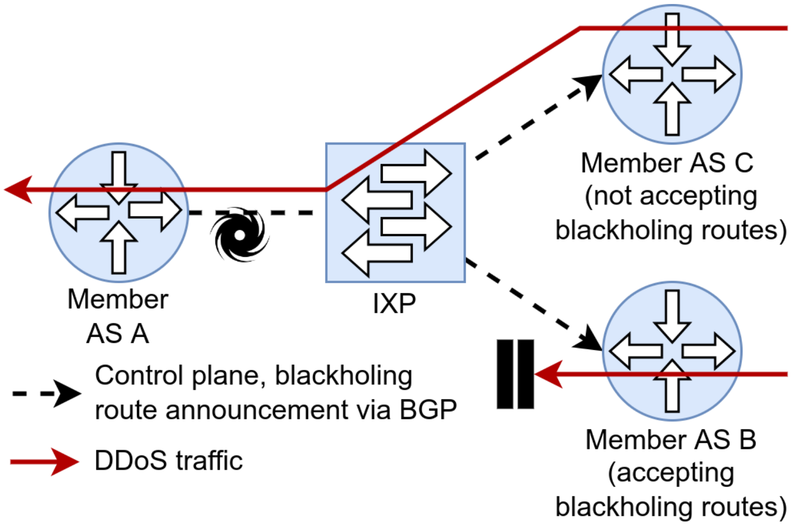
<u>Contribution 3:</u> model drift evaluation; up to 2 years of data from 5 IXPs

Crowdsourced DDoS Labeling with Blackholing



Crowdsourced DDoS Labeling with Blackholing

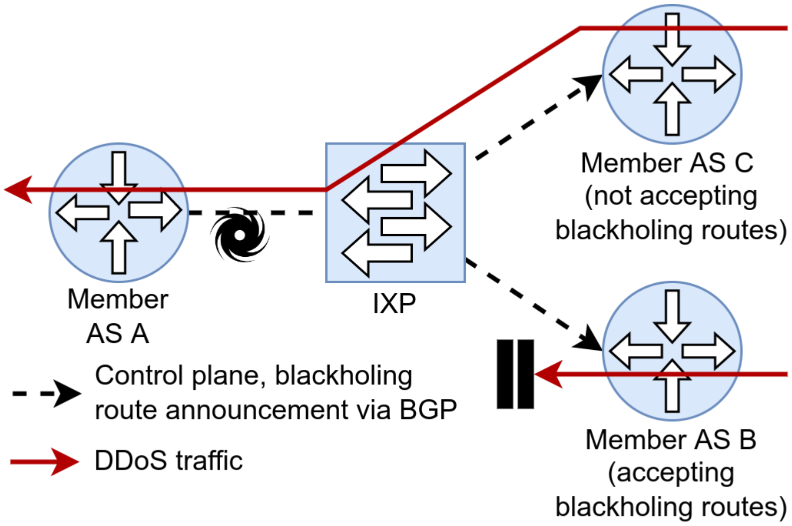
- IXP members not accepting blackholing routes send *__unfiltered and unwanted__* traffic [19]



Crowdsourced DDoS Labeling with Blackholing

- IXP members not accepting blackholing routes send *unfiltered and unwanted* traffic [19]

- Correlate BGP and flow data to *automatically* generate DDoS labels

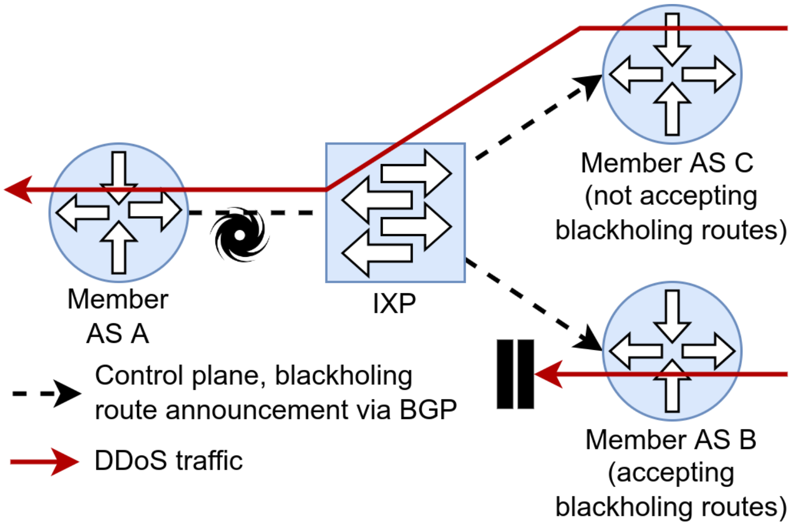


Crowdsourced DDoS Labeling with Blackholing

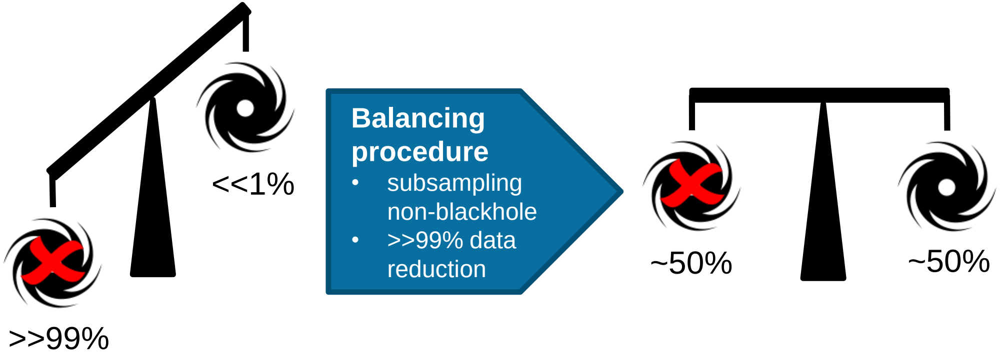
- IXP members not accepting blackholing routes send *unfiltered and unwanted* traffic [19]

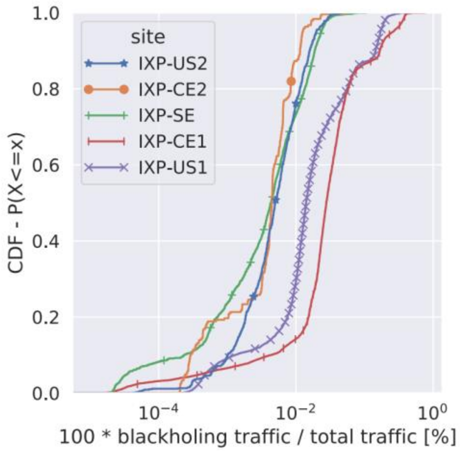
- Correlate BGP and flow data to *automatically* generate DDoS labels

→ Training set size is only limited by size of BGP/flow data



Balancing Procedure

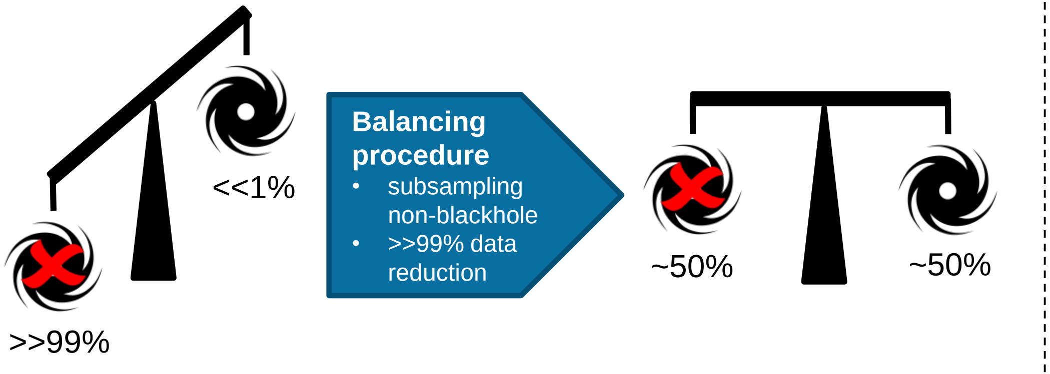


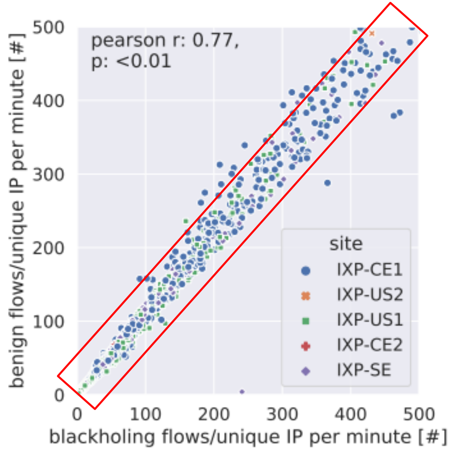


- Blackholing flows are highly underrepresented in overall flow data export (<<1%)
- We balance by subsampling non-blackholing flows
- Balancing preserves #IPs and #Flows/IP in blackholing/non-blackholing classes

→ Reduces overall raw data by >99%

Balancing Procedure





- Blackholing flows are highly underrepresented in overall flow data export (<<1%)
- We balance by subsampling non-blackholing flows
- Balancing preserves #IPs and #Flows/IP in blackholing/non-blackholing classes

→ Reduces overall raw data by >99%

Datasets from Five IXPs

→ ML training set (from BH)

- 685Bn flow records from five IXPs+BGP
  - 3-24 months of data
  - EU and US
  - Up to >800 ASes, up to >10 Tbps traffic
- 202M flow records after balancing

ML pipeline design, training, performance evaluation

train

Datasets from Five IXPs

→ ML training set (from BH)

- 685Bn flow records from five IXPs+BGP
  - 3-24 months of data
  - EU and US
  - Up to >800 ASes, up to >10 Tbps traffic
- 202M flow records after balancing

→ Self-attack set (SAS)

- Collected with different method
- Flow records from self-attacks
  - Dedicated infrastructure
  - DDoS-for-hire services [38]
- 702k flow records

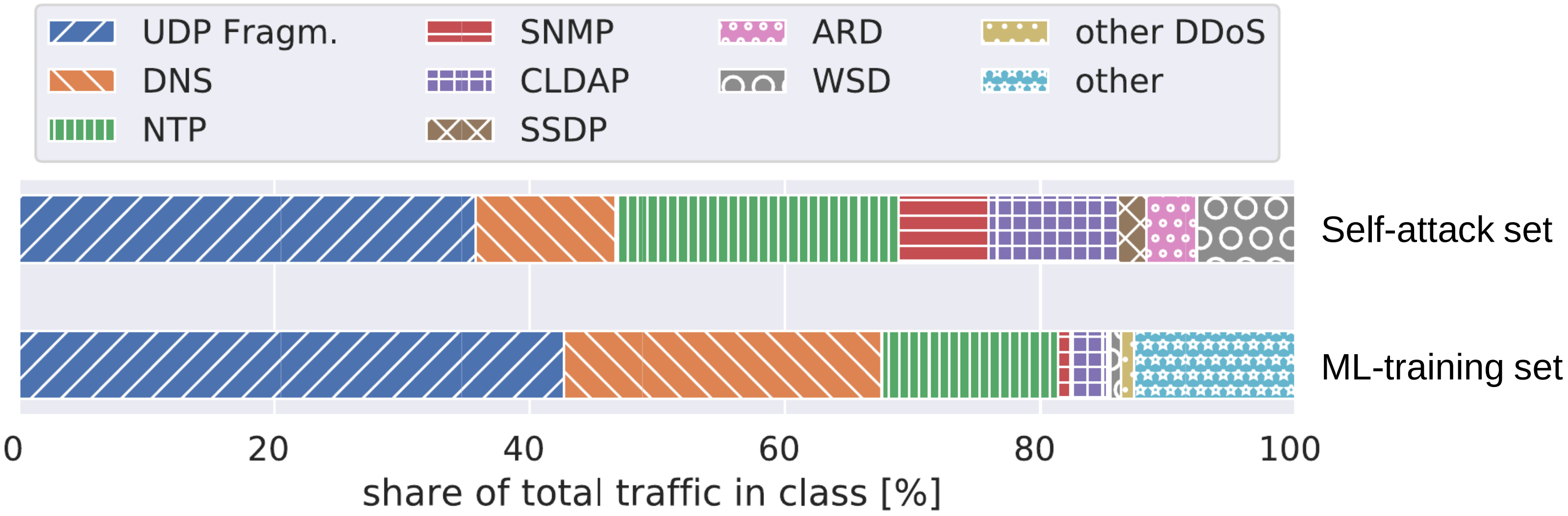
ML pipeline design, training, performance evaluation

train

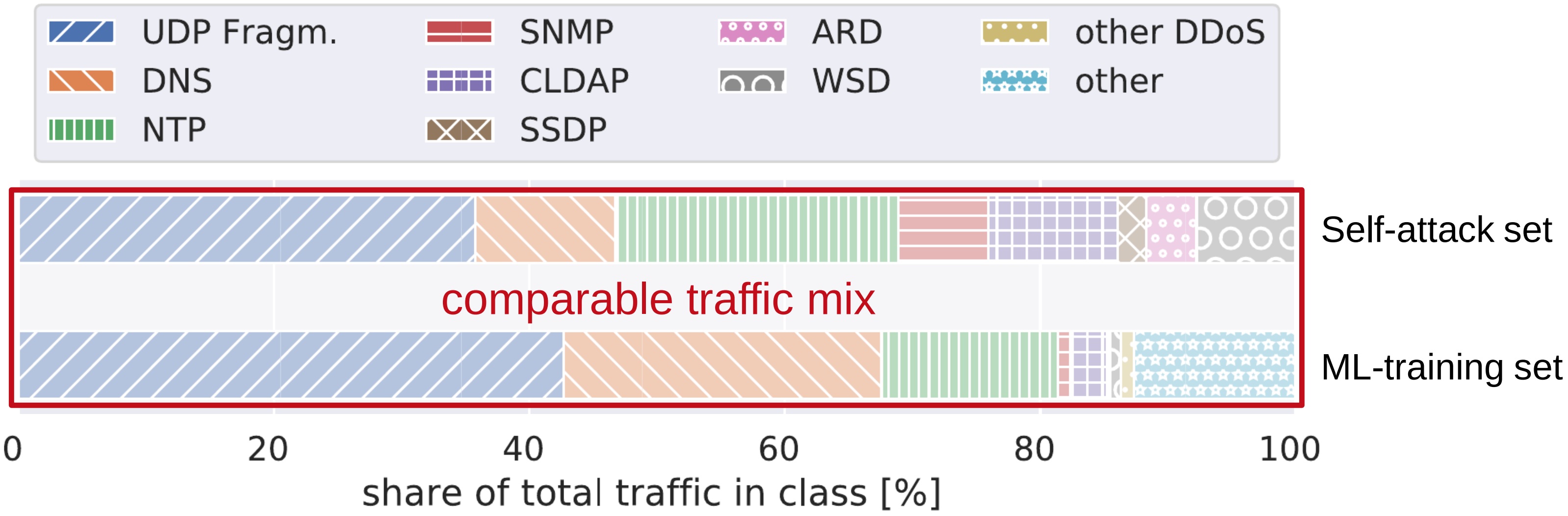
vali-date

Validate models trained on ML training set (reduces risk of bias)

Dataset Validation



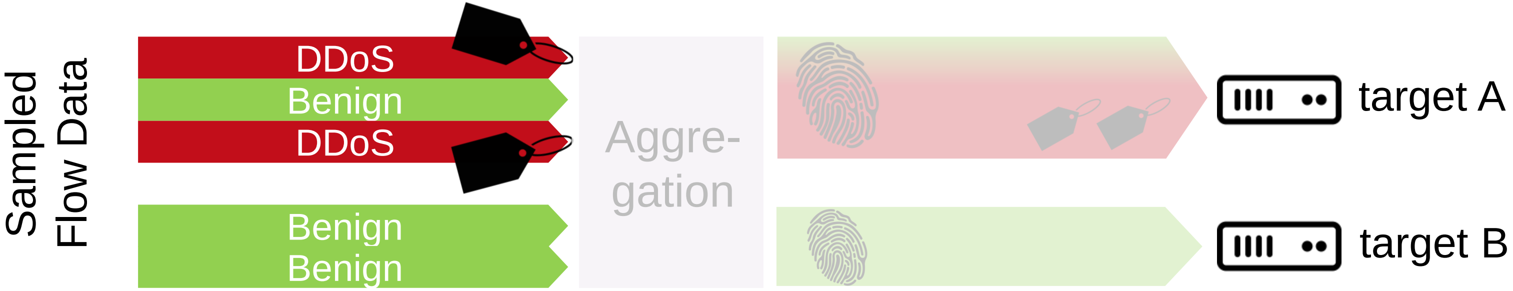
Dataset Validation



Dataset Validation

Challenge: blackholing data is impure; ~15% of possibly benign traffic

ML-Model: Classification Process



Microscopic ML-model (§ 5.2)
Tag single flows if they are likely part of an attack

→ solves impurity of blackholing data

Macrosopic ML-model (§ 5.2)
Classify targets into attacked (A) / not attacked (B)

→ if under attack: drop traffic matching tags

ML-Model: Classification Process

Microscopic ML-model (§ 5.2)
Tag single flows if they are likely part of an attack

→ solves impurity of blackholing data

Macrosopic ML-model (§ 5.2)
Classify targets into attacked (A) / not attacked (B)

→ if under attack: drop traffic matching tags

Microscopic Level (flow tagging)

→ *<u>Goal:</u>* identify blackholing prone flow clusters

- Association Rule Mining (ARM): „*customers buying milk also bought bread.*"
- *<u>Example</u>:* {src_port=389;packet_size=(1400,1500]} → {blackhole}

Microscopic Level (flow tagging)

→ *Goal:* identify blackholing prone flow clusters

- Association Rule Mining (ARM): „*customers buying milk also bought bread.*"
- *Example:* {src_port=389;packet_size=(1400,1500]} → {blackhole}

→ Manual curation by experts

- Support with UI and by minimizing possible tags

id	protocol	port_src	port_dst	packet_size	confidence	antecedent support	rule status	notes
429ce0cf	17	123	~{0,17,19,21,2…	(400,500]	0.97601	0.02598	accept	NTP reflection with typical size to random destination ports (except popular ones).
152bf00c	17	123	~{0,17,19,22,2…	*	0.99136	0.05531	staging	NTP reflection sprayed over arbitrary destination ports.
91fe9d4a	17	123	~{0,17,19,22,2…	(300,400]	0.98893	0.00042	decline staging	NTP reflection attack
43bc7f62	17	123	~{0,17,19,22,2…	(200,300]	0.98588	0.00045	accept	NTP reflection attack

DE CIX

Microscopic Level (flow tagging)

→ Goal: identify blackholing prone flow clusters

- Association Rule Mining (ARM): „*customers buying milk also bought bread.*"
- *Example*: {src_port=389;packet_size=(1400,1500]} → {blackhole}

→ Manual curation by experts

- Support with UI and by minimizing possible tags

id	protocol	port_src	port_dst	packet_size	confidence	antecedent support	rule status	notes
429ce0cf	17	123	~{0,17,19,21,2...	(400,500]	0.97601	0.02598	accept	NTP reflection with typical size to random destination ports (except popular ones).
152bf00c	17	123	~{0,17,19,22,2...	*	0.99136	0.05531	staging decline staging accept	NTP reflection sprayed over arbitrary destination ports.
91fe9d4a	17	123	~{0,17,19,22,2...	(300,400]	0.98893	0.00042		NTP reflection attack
43bc7f62	17	123	~{0,17,19,22,2...	(200,300]	0.98588	0.00045		NTP reflection attack

Study with networking experts shows our approach is understandable and useful (§5.1).

Macroscopic Level (per target)

→ *Goal:* classify targeted hosts correctly (attack/no attack)

- Independent of location and locally explainable

Macroscopic Level (per target)

→ *Goal:* classify targeted hosts correctly (attack/no attack)

- Independent of location and locally explainable

→ Weight of Evidence (WoE) encoding for categoricals [56]

- Likely to appear in blackhole → positive risk score (e.g., reflector IPs, NTP, SSDP)
- Unlikely to appear in blackhole → negative risk score (e.g., 8.8.8.8, HTTP)

Macroscopic Level (per target)

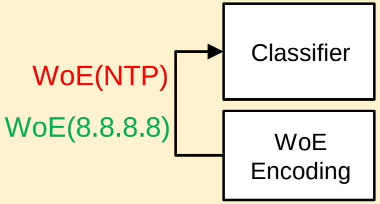
→ <u>*Goal:*</u> classify targeted hosts correctly (attack/no attack)

- Independent of location and locally explainable

→ Weight of Evidence (WoE) encoding for categoricals [56]

- Likely to appear in blackhole → positive risk score (e.g., reflector IPs, NTP, SSDP)
- Unlikely to appear in blackhole → negative risk score (e.g., 8.8.8.8, HTTP)

<u>*Classifier never sees categoricals, only WoE risk scores*</u>



DE CIX

General Performance and Retraining

→ General performance

- Evaluation of five optimized ML classifiers on all data
- XGBoost [23] has highest overall performance (F1-score > 0.98)

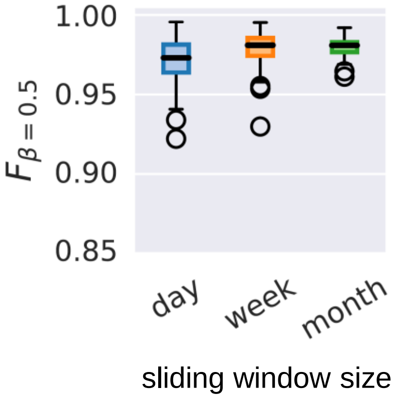
General Performance and Retraining

→ General performance

- Evaluation of five optimized ML classifiers on all data
- XGBoost [23] has highest overall performance (F1-score > 0.98)

→ Retraining

- Temporal model drift is a problem
- Daily retraining with sliding window



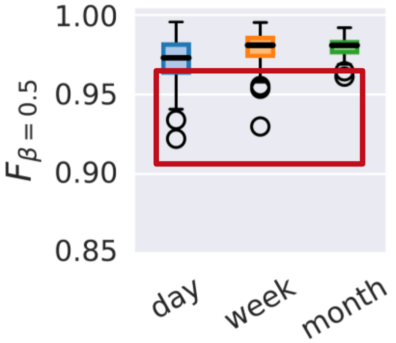
General Performance and Retraining

→ General performance

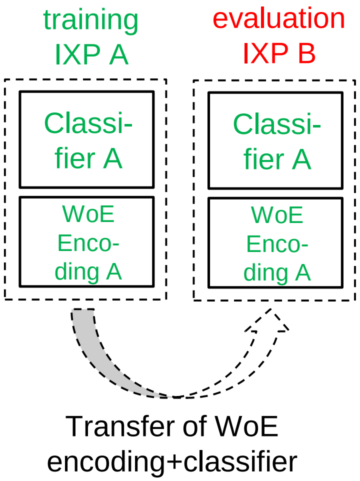
- Evaluation of five optimized ML classifiers on all data
- XGBoost [23] has highest overall performance (F1-score > 0.98)

→ Retraining

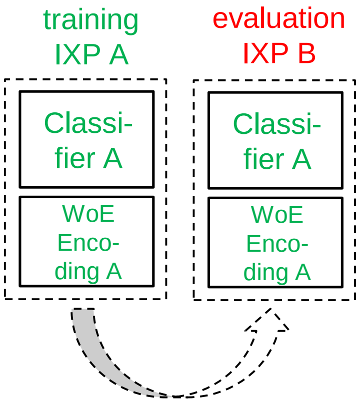
- Temporal model drift is a problem
- Daily retraining with sliding window
- Window size hardly affects median, but reduces outliers



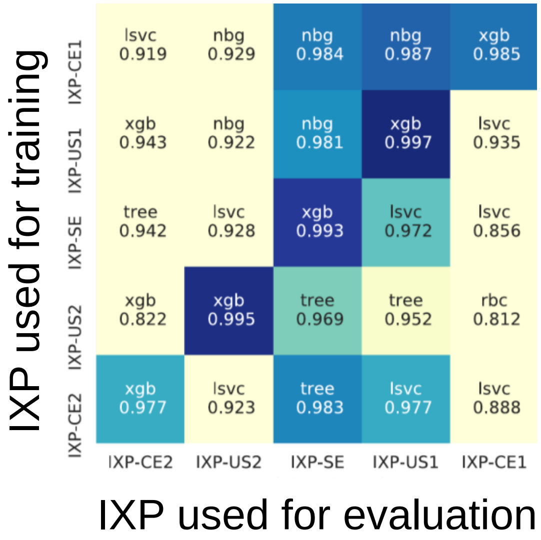
Model Transfer



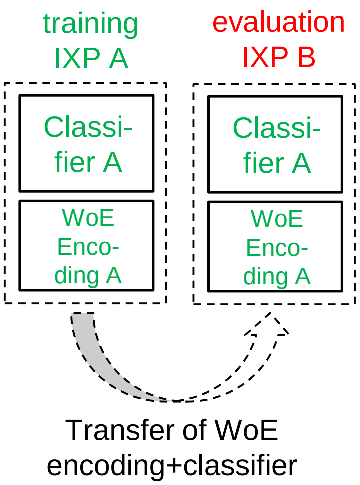
Model Transfer



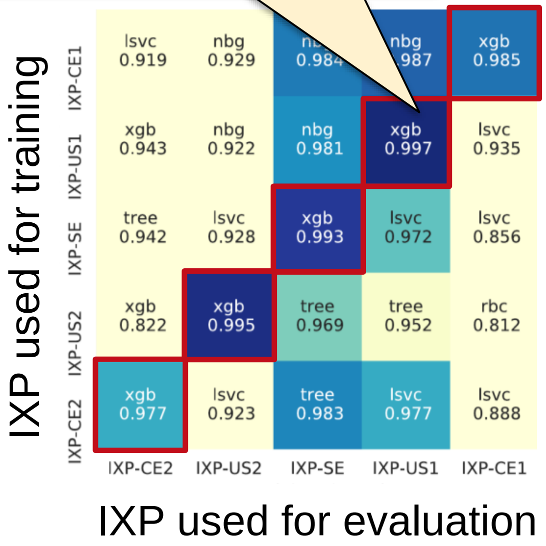
Transfer of WoE
encoding+classifier



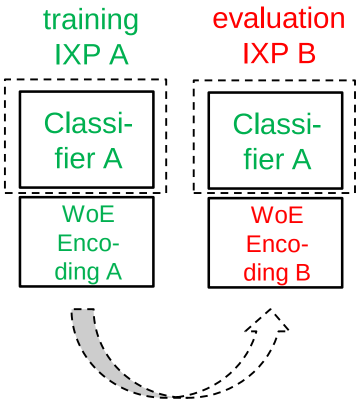
Model Transfer



__Acceptable performance only for training and evaluation at same IXP.__

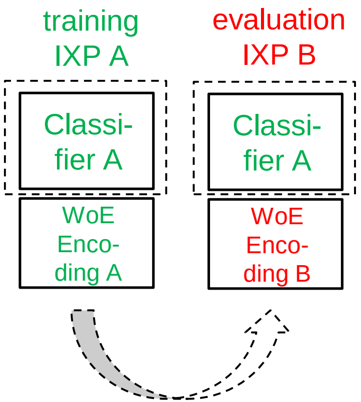


Model Transfer

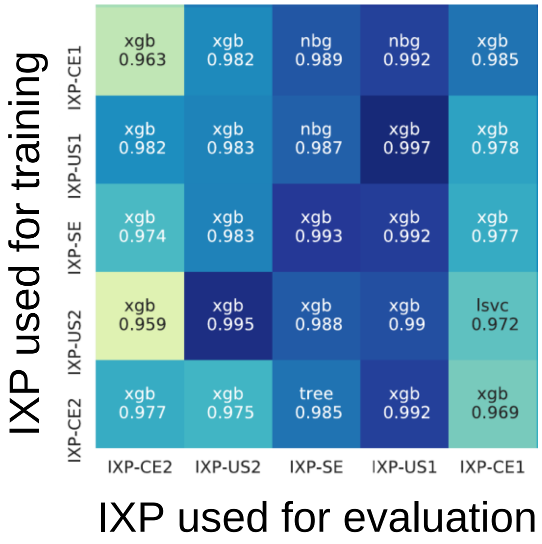


Transfer of classifier,
WoE encoding remains local

Model Transfer



Transfer of classifier,
WoE encoding remains local



Each IXP sees different DDoS vectors and attacking systems (see §6.4)
→ WoEs differ geographically and encapsulate local knowledge

Operational Requirements

Low cost ✓

- no appliances, needs to work with existing hardware

Low maintenance ✓

- no manual definition of rules and triggers, high degree of automation

Member-driven ✓

- IXP members define what DDoS is and what they want to filter

Controllable ✓

- limit possible damage of false positives, understand performance limitations

Want to know more?

Download

IXP Scrubber: Learning from Blackholing Traffic for ML-Driven DDoS Detection at Scale

Matthias Wichtlhuber[1], Eric Strehle[2], Daniel Kopp[1], Lars Prepens[1], Stefan Stegmueller[1]
Alina Rubina[1], Christoph Dietzel[1], Oliver Hohlfeld[2]
[1]DE-CIX [2]Brandenburg University of Technology

ABSTRACT

Distributed Denial of Service (DDoS) attacks are among the most critical cybersecurity threats, jeopardizing the stability of even the largest networks and services. The existing range of mitigation services predominantly filters at the edge of the Internet, thus creating unnecessary burden for network infrastructures. Consequently, we present IXP Scrubber, a Machine Learning (ML) based system for detecting and filtering DDoS traffic at the core of the Internet at Internet Exchange Points (IXPs) which see large volumes and varieties of DDoS. IXP Scrubber continuously learns DDoS traffic properties from neighboring Autonomous Systems (ASes). It utilizes BGP signals to drop traffic for certain routes (blackholing) to sample DDoS and can thus learn new attack vectors without the operator's intervention and on unprecedented amounts of training data. We present three major contributions: *i)* a method to semi-automatically generate arbitrarily large amounts of labeled DDoS training data from IXPs' sampled packet traces, *ii)* the novel, controllable, locally explainable and highly precise two-step IXP Scrubber ML model, and *iii)* an evaluation of the IXP Scrubber ML model, including its temporal and geographical drift, based on data from 5 IXPs covering a time span of up to two years.

CCS CONCEPTS

• **Security and privacy** → **Denial-of-service attacks**; • **Networks** → **Wide area networks**; *Network monitoring; Public Internet.*

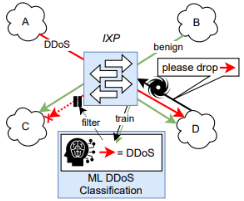


Figure 1: IXP Scrubber applies an ML DDoS classifier at IXPs at the Internet's core and filters DDoS traffic for connected networks. It learns continuously from ASes (A-D) marking unwanted traffic (blackholing).

One of the most prevalent threats to online services to date are DDoS attacks [17, 35, 39, 46, 47, 52, 59]. DDoS attacks aim at consuming more critical resources than available to a service, e.g., network bandwidth, which makes protection against DDoS hard for victims. They are frequent (e.g., thousands of attacks can be observed at certain vantage points every single day [16, 37]), they can be conducted without technical expertise [38], and can generate attack volumes (e.g., of up to 3.5 Tbit/s observed in late 2021 [53])

Paper was published at ACM SIGCOMM'22; available for download from DE-CIX

Thank You for Your attention!

DE-CIX Management GmbH | Lindleystr. 12 | 60314 Frankfurt | Germany
Phone + 49 69 1730 902 0 | sales@de-cix.net | www.de-cix.net