Deterrence of Intelligent DDoS via Multi-Hop Traffic Divergence

Yuanjie Li*, Hewu Li*, Zhizheng Lv*, Xingkun Yao*, Qianru Li†, Jianping Wu*
* Institute for Network Sciences and Cyberspace, BNRist, Tsinghua University, Beijing 100084, China
† Department of Computer Science, University of California at Los Angeles, CA 90095, USA

ABSTRACT

We devise a simple, provably effective, and readily usable deterrence against *intelligent, unknown* DDoS threats: Demotivate adversaries to launch attacks via multi-hop traffic divergence. This new strategy is motivated by the fact that existing defenses almost always lag behind numerous emerging DDoS threats and evolving intelligent attack strategies. The root cause is if adversaries are smart and adaptive, no single-hop defenses (including optimal ones) can perfectly differentiate unknown DDoS and legitimate traffic. Instead, we formulate intelligent DDoS as a game between attackers and defenders, and prove how multi-hop traffic divergence helps bypass this dilemma by reversing the asymmetry between attackers and defenders. This insight results in EID, an **E**conomical **I**ntelligent DDoS **D**emotivation protocol. EID combines local weak (yet divergent) filters to provably null attack gains *without* knowing exploited vulnerabilities or attack strategies. It incentivizes multi-hop defenders to cooperate with boosted local service availability. EID is resilient to traffic dynamics and manipulations. It is readily deployable with random-drop filters in real networks today. Our experiments over a 49.8 TB dataset from a department at Tsinghua campus network validate EID's viability against rational and irrational DDoS with negligible costs.

CCS CONCEPTS

• **Security and privacy → Denial-of-service attacks**; **Economics of security and privacy**; • **Theory of computation → Algorithmic game theory and mechanism design**.

KEYWORDS

Intelligent DDoS, traffic divergence, game theory, adversarial machine learning, random drop, cyber security economics.

ACM Reference Format:
Yuanjie Li, Hewu Li, Zhizheng Lv, Xingkun Yao, Qianru Li, Jianping Wu. 2021. Deterrence of Intelligent DDoS via Multi-Hop Traffic Divergence. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21), November 15–19, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3460120.3484737

Hewu Li is the corresponding author.

CCS '21, November 15–19, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8454-4/21/11…$15.00
https://doi.org/10.1145/3460120.3484737

1 INTRODUCTION

Can we fight against network attacks that we do not know? Network security is fundamentally asymmetric warfare between attackers and defenders: The evolution of security defenses almost always lag behind the emergence of numerous new threats and fast-evolving attack strategies. It is usually too late to come up with remedies before new threats have already caused negative impacts [1, 2]. This long-standing phenomenon is exacerbated by recent advances in adversarial machine learning, which empowers criminals to automatically discover more unknown vulnerabilities and refine their attack strategies. Thus, it is open to question if defenders still have decent chances to win this race against intelligent attackers.

This paper studies this question in one of the most serious threats to the Internet: Distributed Denial of Services (DDoS). In DDoS attacks, adversaries disrupt victims' service by exhausting their network resources with compromised robot networks (botnets). In the past decades, we have seen numerous DDoS to many institutes such as Google [3], GitHub [4] and Amazon [5] with more than 1 Tbps malicious traffic [6–8] and huge financial loss [9].

As a result, DDoS is probably one of the most studied topics in network security. Numerous DDoS defenses have been proposed, spanning on ingress filtering [10–13], source validation [14, 15], anomaly detection [16–18], anycast [19, 20], traceback [21, 22], Internet architecture change [23–26], and many more. As for commercial products, huge scrubbing centers have been widely deployed [6, 27, 28] to absorb global DDoS traffic. Despite these extensive excellent efforts, however, the frequency and intensity of DDoS still continue to grow without signs of stoping.

We note that, DDoS is fundamentally hard to eliminate because attackers and defenders are asymmetric in at least three aspects:

(1) **Massive attack sources:** It is much easier to gain attack capacity than defense capacity. Attackers can hijack numerous bots (e.g., vulnerable IoT devices [29]) to form huge botnets at low costs. Instead, the defense capacity is usually expensive, e.g., $150,000/year for ≤600Gbps DDoS scrubbing [30].
(2) **Numerous unknown threats:** The attack surface is always larger than defense coverage. From various emergent protocols and applications, attackers can easily discover and exploit new vulnerabilities that defenders do not recognize.
(3) **Evolving intelligent attack strategies:** Disguising DDoS is always easier than detecting it. With the recent advances in generative adversarial learning [31–33] and layer 7 attacks [34, 35], smart attackers can easily imitate legitimate usage behaviors to bypass or defeat defenders.

As we will prove in §3, these asymmetries ensure strategic attackers can always benefit from DDoS even under optimal intelligent defenses. The root cause is that, from generative adversarial learning's view, the strategically mimicked DDoS traffic is *indistinguishable* from legitimate traffic. Such indistinguishability is exacerbated by

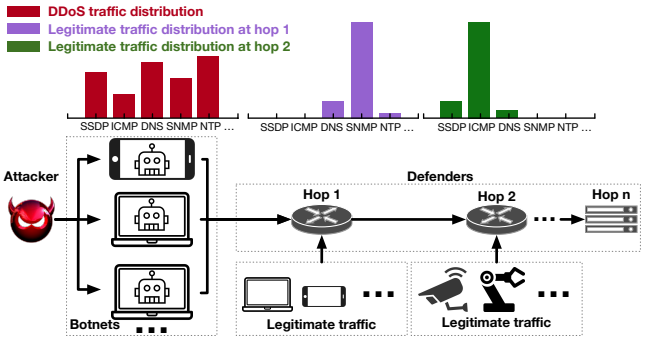


Figure 1: An example of threat model in intelligent DDoS.

numerous unknown threats and traffic dynamics (as "noises") and is unlikely to be resolved by the aforementioned existing solutions.

To this end, we believe the ultimate mitigation of intelligent DDoS should not just passively defend it after it occurs. Instead, defenders should also seek to proactively prevent attackers from launching DDoS in the first place, which is possible when the attack costs exceed gains. This strategic deterrence complements existing passive defenses. It bypasses the above asymmetries with a more economical, affordable solution against intelligent, unknown DDoS.

In this work, we formulate intelligent unknown DDoS as an evolutionary game between attackers and defenders, and devise strategic deterrence against DDoS by nulling attackers' benefit-cost ratio. Our key observation is that, *multi-hop traffic divergence offers a natural mechanism to reverse the attacker-defender asymmetry in this DDoS game.* In real distributed networks, traffic distributions from different nodes are naturally divergent due to heterogeneous user behaviors and network capabilities. Suppose multi-hops on the path enable smart filters (inspired by GAN [32], detailed in §5.2). In that case, they can collaboratively force the strategic attacker to face the dilemma of *whose traffic* to mimic (Figure 3). Imitating one hop's legitimate traffic leads to significant deviance from other hops' traffic, thus lowering the success of passing all nodes' DDoS filters. As we will see in §5, this gaming mechanism provably nulls attackers' gains *without* prior knowledge of new threats or attack strategies, and yields a win-win situation between defenders. It is resilient to malicious manipulations and traffic dynamics, and readily deployable with random-drop filters in real networks.

Following this insight, this paper makes three contributions:

(I) We formulate intelligent DDoS in reality as a game between attackers and defenders (§2), derive optimal single-hop filters via generative adversary nets (GAN), and prove the legitimacy-DDoS indistinguishability even for optimal filters (§3). This motivates us to go beyond existing single-hop DDoS defenses;

(II) We devise EID, a distributed **E**conomical **I**ntelligent DDoS **D**emotivation protocol (§4–5). EID formulates multi-hop traffic divergence as an extension of f-divergence in statistics [36]. Based on traffic divergence, EID recursively combines weak (yet divergent) local filters from multi-hop defenders to form a strong global DDoS demotivation. This meta-policy provably nulls DDoS attackers' benefit-cost ratio, and offers built-in incentives for multi-hop defenders to contribute. EID supports scalable on-demand demotivation via composable filters and hop pruning. It is incrementally deployable with random-drop filters in commodity hosts and routers.

Table 1: Notations and definitions.

Notation	Definition
k	Network traffic type index: $k = 1, 2, ...K$
μ_n, μ_a	Network capacity at a legitimate network node (hop) n and attacker
$\mathbf{p_a}, \mathbf{p_n}$	DDoS and hop n's traffic distribution: $\mathbf{p_n} = (\mu_n, p_n^1, ...p_n^k, ...)$, $\mathbf{p_a} = (\mu_a, p_a^1, ...p_a^k, ...), \sum_k p_n^k = 1, \sum_k p_a^k = 1$
$\mathbf{q_n}$	Equivalent legitimate traffic distribution from hop 1 to n
$d_n, d(\mathbf{p_1}, \mathbf{p_2}, ...\mathbf{p_n})$	Multi-hop traffic divergence between distribution $\mathbf{p_1}, \mathbf{p_2}, ...\mathbf{p_n}$
$D_{1,n}(k)$	Hop n's local DDoS filter for traffic type k: $D_{1,n}(k) \in [0, 1]$
$D_n(k)$	Accumulative filter from hop 1 to n: $D_n(k) = D_{n-1}(k)D_{1,n}(k)$
$U(\mathbf{p_a}, \mathbf{D})$	Adversary's attack gain under DDoS strategy $\mathbf{p_a}$ and defenses $\mathbf{D}$
$\eta(\mathbf{p_a}, \mathbf{D})$	Adversary's benefit-cost ratio: $\eta(\mathbf{p_a}, \mathbf{D}) = U(\mathbf{p_a}, \mathbf{D})/\mu_a$
$V_n(\mathbf{p_a}, \mathbf{D})$	GAN-inspired utility in n for DDoS-legitimacy differentiation
o_n^k	The observable total traffic volume of packet type k at hop n
c_k	Bandwidth amplification factor for protocol/application k

(III) We prototype EID in the PA-7050 firewall (§6) and evaluate it with a 49.8 TB real dataset from a department at Tsinghua campus network. Without knowing attack strategies or exploited threats, EID demotivates both real and optimal DDoS attacks with < 1 benefit-cost ratio for *rational* adversaries, and retains ≥99.9% (≥84.7%) legitimate service availability under *irrational* real (optimal) DDoS attacks with negligible network and system costs (§7).

While EID currently focuses on DDoS, we believe its core ideas can be extended to broader intelligent network attacks.

2 THREAT MODEL: INTELLIGENT DDOS AS A GAME

This paper studies intelligent DDoS, in which both attackers and defenders are smart and strategic. The defenders seek to accurately detect and block attackers' DDoS traffic. The attackers aim to bypass defenders' protections by exploiting *new* threats to defenders, and mimicking legitimate usage behaviors. Attackers and defenders mutually impact each other and iteratively refine their strategies to maximize their own merits. This forms an evolutionary game [37, 38] between attackers and defenders that is consistent with most DDoS in reality [6–8]. Figure 1 illustrates the threat model and Table §1 summarizes its notations.

Attack model: We consider a strong attacker that forms a massive botnet to exhaust the victims[1]. The attacker can systematically monitor victims' usage behaviors, discover new vulnerabilities, and refine its strategy to bypass defenses by exploiting new threats and mimicking legitimate network behaviors (e.g., via generative adversary learning [32, 33]). It has various options to craft attack traffic, including choosing attack protocols, packet header fields or payloads (Figure 4). To combat smart defenders, it can also exploit multiple vulnerabilities, and distribute its traffic among diverse protocols and applications (i.e., *mixed DDoS attacks* [39, 40]). The attacker aims for *successful* DDoS. Its benefits beyond successful attacks (e.g., extorting victims via *unsuccessful attacks*) are out of this paper's scope, although our core idea of deterrence still holds and is generally extensible to these scenarios as discussed in §8.

To tackle unknown DDoS threats, our model does not assume specific attack methodologies or exploited vulnerabilities. Instead, we focus on the attacks' **benefit-cost ratio**, which motivates most real adversaries' decisions in the cybercrime economy [41–43]. For example, a large-scale study from [8] shows global DDoS becomes less frequent when Bitcoin price increases, because more attackers

[1] In this paper, we use "attackers", "adversaries", and "criminals" interchangeably. We also use "defenders", "victims", and "hops" interchangeably.

will utilize their botnets to mine Bitcoins rather than run DDoS. The recent rise of "DDoS-as-a-Service" [44–46] on the darknet market also renders the importance of the benefit-cost ratio for criminals.

We next formalize the adversary's benefit-cost ratio. As attacks' "benefits" and "costs" can be diverse in reality, we aim for generic and extensible definitions of benefit-cost ratio. On the benefit side, both attack and defense strategies affect the adversary's gains. The adversary gains more if more DDoS traffic reaches victims, and gains zero if defenders block all DDoS traffic [34, 35, 47–50]. Let $\mathbf{p_a} = (\mu_a, p_a^1, ..., p_a^k, ...)$ be the adversary's DDoS traffic distribution (attack strategy), with μ_a as its total attack capacity and p_a^k as the probability of traffic type $k = 1, 2...., K$ in this DDoS attack portfolio[2]. So $\mu_a p_a^k$ is DDoS traffic k's portion. Meanwhile, DDoS traffic can be blocked by defenders' filters $\mathbf{D} = (D(1), ...D(k), ...)$, with $D(k) \in [0, 1]$ as defenders' probability of forwarding traffic k. We assume the attack gain $U(\mathbf{p_a}, \mathbf{D}) \geq 0$ satisfies:

(i) *More DDoS, more gains:* $U(\mathbf{p_a}, \mathbf{D})$ increases monotonically with DDoS traffic that reaches victim $\mu_a p_a^k D(k)$ for each type k[3];

(ii) *No DDoS, no gains:* $U(\mathbf{p_a}, \mathbf{D}) = 0$ if $\mu_a p_a^k D(k) = 0, \forall k$.

On the cost side, the adversary should employ massive attack capacity (μ_a) for successful DDoS. Such attack capacity is not free (though usually cheap): The adversary should spend efforts hijacking large botnets, rent botnets from DarkNet [41–43], or purchase the DDoS service [44–46]. To this end, we formulate the costs as attack capacity μ_a consumed to initiate DDoS, which can be employed from botnets, open resolvers, self-purchased hosts, or others. So its benefit-cost ratio η quantifies how much attack capacity it should pay for the targeted DDoS gains:

$$\eta(\mathbf{p_a}, \mathbf{D}) \triangleq \frac{U(\mathbf{p_a}, \mathbf{D})}{\mu_a} \tag{1}$$

As a concrete instance, in the amplification DDoS attack (which contributes more than 80% large DDoS with >1 Tbps attack capacity today [6–8]), the benefit-cost ratio η equals the bandwidth amplification factor [47, 48, 51, 52], a well-known metric to quantify the usefulness of this attack (detailed in §5.4). We assume an adversary is more motivated to run DDoS with higher benefit-cost ratio η.

Defense model: In a network, both the end hosts and network nodes on the path are victims and suffer from DDoS attacks. They seek to maximize their network service availability with minimal costs by strategically refining their defenses against DDoS. Their local network service availability (utility) is defined as the percentage of successfully forwarded legitimate traffic under DDoS attacks. We will derive and explain its formulation in §3.1 (Equation 3). Victims can be threatened by *unknown attacks*, i.e., they do not know adversaries' attack strategies, malicious traffic distribution, or exploited vulnerabilities. Each node on the path only knows its legitimate *local* traffic distributions (via standard traffic monitoring or offline profiling) and systematically monitors its runtime local traffic (which aggregates legitimate and DDoS traffic). Our model allows for nodes from multiple paths. We assume they are willing to cooperate if beneficial to themselves [24, 53]. They use readily available mechanisms to mitigate DDoS, without requiring sophisticated software/hardware modifications. We explore a simple and popular mechanism in almost all commodity network nodes: random packet drop. Despite its simplicity, this mechanism suffices to provably void attack gains under intelligent and unknown DDoS, with negligible impacts on legitimate traffic (§5).

Algorithm 1 Optimal single-hop DDoS filter inspired by GAN.

Input: Runtime traffic rate $\{o_n^k\}_k$ and configurable maximal load threshold $\{\mu_n^k = \mu_n p_n^k\}_k$ for each traffic type k

1: **for** each packet belonging to traffic type k **do**
2: **if** $o_n^k \leq \mu_n^k$ **then** Forward the packet; ▷ *Underloaded: Forward all traffic.*
3: **else then** Forward the packet with probability μ_n^k / o_n^k; ▷ *Overloaded: Stateless random drop.*
4: **end for**

Interplay between attackers and defenders: For their own merits, both attackers and defenders iteratively refine their strategies based on the other's behaviors. To understand their mutual impacts, we study their Nash Equilibrium, during which neither can further improve their benefits if the other's strategy remains unchanged. Note our model does **not** assume both players are entirely rational or perfect; they may make mistakes or adopt imperfect strategies (e.g., due to incomplete information). Instead, the Nash Equilibrium sheds light on the *stable state* as smart attackers and defenders evolve. As we will see in §5.2, minimizing the maximal benefit-cost ratio for adversaries in this Nash Equilibrium offers a practical path to demotivate intelligent, unknown DDoS.

3 WHY IS INTELLIGENT DDOS HARD TO ELIMINATE?

DDoS is a decades-old security threat to the Internet. The challenges for eliminating DDoS have been extensively discussed and validated from network architecture and system mechanism perspectives (§9). This section complements them with an orthogonal view from evolutionary gaming and generative adversarial learning policies [32, 33, 54]. We investigate how intelligent attackers and defenders interact and evolve for their own merits (§3.1), and derive their stable gains at Nash Equilibrium (§3.2). We prove that, due to the asymmetric nature in §1, any single defender's optimal mitigation cannot eliminate smart adversaries' gains from intelligent DDoS.

3.1 A Single Defender's Optimal Filters

We start from a single defender's perspective to explore the optimal DDoS defenses. As introduced in §2, a smart adversary may exploit new threats and carefully-crafted DDoS traffic to mimic legitimate behaviors and bypass the defenses adaptively. Moreover, the "legitimacy" of network traffic sometimes depends on the environment and context, especially for layer-7 DDoS attacks [34, 35]. Both make the static defenses (e.g., rule-based filters) ineffective against unknown DDoS vulnerabilities.

To combat unknown DDoS attacks, a defender's most viable choice so far is to adopt *probabilistic* filters to block DDoS traffic with minimal hurt for legitimate traffic. For maximal legitimate network availability with low cost, the defender should adopt filters that are *accurate* to maximize the differentiation between legitimate and DDoS traffic, *efficient* to process traffic with marginal overhead, and *readily deployable* today with minimal changes (if possible).

[2]Our threat model supports flexible traffic classification granularities $k = 1, 2, \ldots K$ (depending on attack strategies). For example, k can indicate the protocols to exploit (DNS, NTP, ICMP, etc), different variants of vulnerabilities in each protocol (e.g., exploiting open resolvers *or* authoritative name servers in DNS-based DDoS [47]), or application-layer content types. We support all of them with seamless tradeoff (§5.3).

[3]Besides DDoS traffic, it is also possible to extend attackers' utility to consider mistakenly dropped legitimate traffic in $U(\mathbf{p_a}, \mathbf{D})$ as we will see in §8.

To this end, we devise the optimal DDoS filters that can be realized in commodity network nodes, and prove its optimality from the generative adversarial network (GAN) perspective. In the classical GAN model [32, 33, 54], the generator (adversary) and discriminator (defender) contest with each other in a game. The generator is trained to fool the discriminator, during which the discriminator is also updated dynamically. At the equilibrium, the discriminator cannot differentiate legitimate and generated data. In our context, consider a defender n with legitimate traffic distribution $\mathbf{p_n} = (\mu_n, p_n^1, ..., p_n^k, ...)$, where $\mu_n > 0$ is defender's maximal network capacity and p_n^k is the probability that its legitimate traffic belongs to type k (i.e., the maximal legitimate traffic rate belonging to k is $\mu_n p_n^k$). As explained in §2, we assume the defender knows its legitimate traffic distribution $\mathbf{p_n}$ as prior knowledge. Without DDoS, the legitimate traffic rate o_n will not exceed the capacity μ_n.

Now consider a DDoS attack with malicious traffic distribution $\mathbf{p_a} = (\mu_a, p_a^1, ..., p_a^k, ...)$. In this case, the runtime traffic would be a sum of legitimate and DDoS traffic: $o_n^k = \mu_n p_n^k + \mu_a p_a^k, \forall k$. To disrupt the defender's service, this DDoS aims to exhaust its capacity $o_n > \mu_n$. Then the defender has to drop (or delay the processing of) some packets. To this end, we propose a simple random drop filter:

$$D_{1,n}^*(k) \triangleq \min\left(1, \frac{\mu_n p_n^k}{o_n^k}\right) \tag{2}$$

Without DDoS, the runtime traffic $o_n^k \le \mu_n p_n^k$ so $D_n^*(k) = 1$, i.e., no packets are dropped. Otherwise, $o_n^k = \mu_n p_n^k + \mu_a p_a^k$ and therefore

$$D_{1,n}^*(k) = \frac{\mu_n p_n^k}{\mu_n p_n^k + \mu_a p_a^k} \tag{3}$$

Algorithm 1 illustrates the filter's random drop policy. At runtime, an overloaded node n randomly drops packets of traffic type k with probability $1 - D_{1,n}(k)$. In this way, the filter guarantees the total traffic is always no more than $\mu_n^k = \mu_n p_n^k$. The filter is efficient since it simply randomly drop packets, without complex per-packet processing or maintaining any states in the host. It is also readily deployable since random-drop policy has been an de facto mechanism in commodity hosts, routers, and firewalls (§6).

We next prove that, from GAN's perspective, this filter also maximizes the discrimination between legitimate and DDoS traffic, thus facilitating high network service availability. For readers unfamiliar with GAN, we recommend them to read [32] for a high-level review and its widespread influence on machine learning security (e.g., Deepfake [55]). In our context, the adversary crafts its DDoS traffic to bypass defender's filters, while the defender seeks to maximize the differentiation between legitimate traffic $\mathbf{p_n}$ and DDoS $\mathbf{p_a}$. To this end, GAN suggests the defender to train its filter $\mathbf{D_{1,n}}$ to maximize the following divergence variance function [32, 54]:

$$V_n(\mathbf{p_a}, \mathbf{D}) \triangleq \mu_n \mathbb{E}_{k\sim\mathbf{p_n}} \log D_{1,n}(k) + \mu_a \mathbb{E}_{k\sim\mathbf{p_a}} \log(1 - D_{1,n}(k))$$

If $V_n(\mathbf{p_a}, \mathbf{D})$ is maximized, the "distance" between $\mathbf{p_n}$ and $\mathbf{p_a}$ (a variant of f-divergence in statistics [54], detailed in §5.1) is also provably maximized in GAN, thus helping defenders accurately detect and drop DDoS traffic with minimal hurt for legitimate traffic. The following result confirms our filters indeed achieve so:

Theorem 1 (Optimal single-hop filter)**.** *Under DDoS attack* $\mathbf{p_a}$*, each hop* n *maximizes* $V_n(\mathbf{p_a}, \mathbf{D})$ *with the filter* $\mathbf{D_{1,n}^*}$ *in Equation 2.*

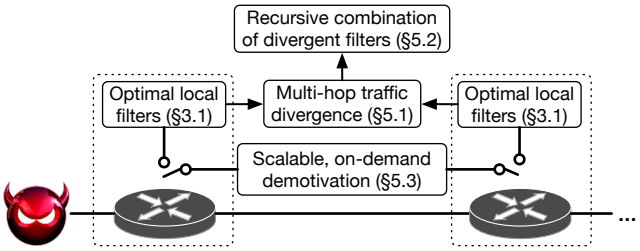


Figure 2: Overview of EID.

Theorem 1 is proved in Appendix A and is a variant of Proposition 1 in the original GAN paper [32]. It shows that Equation 2 is the best random-drop filter the defender can adopt to combat intelligent, unknown DDoS attacks. To understand filters' merits for service availability, we follow the statistics community to define their accuracy as $\frac{TP+TN}{TP+FP+TN+FN}$, where TP is the true positive (i.e., legitimate packets are correctly forwarded), TN is the true negative (DDoS packets are correctly dropped), FP is the false positive (DDoS packets are mistakenly forwarded), and FN is the false negative (legitimate packets are mistakenly dropped). It is easy to verify that local filter's accuracy in Equation 3 equals itself $D_{1,n}^*(k)$.

3.2 Are Optimal Filters Effective for Attackers?

We next switch to the strategic attackers' view to explore their reactions under the optimal defenses in §3.1. As an asymmetric war, intelligent DDoS is inherently advantageous to attackers rather than defenders. We show a single defender's optimal filters still fail to eliminate intelligent DDoS due to three fundamental limits:

(1) DDoS-legitimacy indistinguishability: In the presence of unknown vulnerabilities and mimicked DDoS traffic, no traffic filters can 100% accurately differentiate DDoS and legitimate traffic. As shown in §3.1, the optimal filters' accuracy equals $D_{1,n}^*(k)$, which decreases monotonically with the adversary's attack capacity μ_a. Under heavier DDoS attacks, the single defender's optimal filters will misclassify more legitimate and DDoS traffic, thus mistakenly disrupting the legitimate network service. Advanced anomaly detections or more expensive stateful firewalls may help improve the accuracy, but still cannot always succeed due to attackers' exploited unknown threats and mimicked DDoS traffic.

(2) Undiminished attack gains: Most adversaries are willing to launch intelligent DDoS when the attack gains exceed costs (§2). Even under the defender's optimal filters in §3.1, it is still possible for the smart adversary to adapt its attack strategy to retain appealing benefit-cost ratio. Specifically, given a fixed attack capacity μ_a, the attacker's optimal DDoS strategy $\mathbf{p_a^*}$ under defender's optimal filters $\mathbf{D_{1,n}^*}$ in Equation 2 is as follows:

$$\mathbf{p_a^*} = \arg\max_{\mathbf{p_a}} \ \eta(\mathbf{p_a}, \mathbf{D_{1,n}^*}) \qquad s.t. \sum_k p_a^k = 1, 0 \le p_a^k \le 1, \forall k$$

Note $(\mathbf{p_a^*}, \mathbf{D_{1,n}^*})$ forms the *Nash Equilibrium* in the intelligent DDoS, because neither attacker nor defender can gain more if the other's strategy remains unchanged. $\mathbf{p_a^*}$ is adversaries' best attack strategy when defenders have adopted the optimal filters. In §5.4, we will showcase the concrete Nash Equilibrium in the popular amplification attacks in Lemma 1. Here we show the adversary always has a guaranteed lower bound of benefit-cost ratio at Nash Equilibrium, thus always motivated to launch intelligent DDoS attacks if attack capacity is cheap (which is mostly true in reality [56, 57]):

Theorem 2 (Lower bound of attack gains). *At Nash Equilibrium, at least* $\frac{1}{2}\sum_k \min\left(\mu_a p_a^k, \mu_n p_n^k\right)$ *DDoS traffic will bypass the optimal filters in §3.1, and yield a lower bound of benefit-cost ratio* $\eta(\mathbf{p_a^*}, \mathbf{D_{1,n}^*}) \geq \eta(\mathbf{p_n}, \mathbf{D_{1,n}^*}) > 0$ *if legitimate traffic* $\mu_n > 0$.

Theorem 2 is proved in Appendix B. This guaranteed benefit-cost ratio is achieved when the adversary fully mimics the legitimate traffic $\mathbf{p_a} = \mathbf{p_n}$ (e.g., via generative models like GAN). Such lower bound is non-negligible as the DDoS traffic that reach victims grows monotonically to attacker and defender's capacity. In most cases, the adversary gains even higher benefit-cost ratio (§5.4), thus motivating them to launch DDoS. Fundamentally, such non-disappearing attack incentives arise from DDoS-legitimacy indistinguishability.

(3) Optimal filters' vulnerability to traffic dynamics: The analysis so far assumes the defender *can* always achieve the optimal filters, which requires accurate prior knowledge legitimate traffic distribution $\mathbf{p_n}$. In reality, however, both the legitimate and DDoS traffic varies over time, thus incurring "noises" for the defender to approximate the optimal filters[4]. In the presence of defender's imperfect filters, the adversary can further refine its attack strategies for higher benefit-cost ratio and increase the DDoS severity.

3.3 Problem Statement

We aim to overcome these limitations of existing defenses against intelligent, unknown DDoS. We consider the threat model in §2 and seek a generic solution with

(1) **Provable deterrence against intelligent DDoS:** Even for strategic and unknown attacks, it can provably null adversaries' benefit-cost ratio to demotivate DDoS;
(2) **Built-in incentives of deployment:** Defenders are self-motivated to adopt this solution for their own benefits;
(3) **Affordability:** The solution should be incrementally deployable in today's networks with marginal costs.

4 EID OVERVIEW

We devise EID, an **E**conomical **I**ntelligent DDoS **D**emotivation that achieves all the goals in §3.3. Figure 2 overviews EID. EID is a distributed, signaling-free protocol among legitimate network nodes (defenders). To overcome the limitations of single-hop defenses in §3, EID explores *multi-hop traffic divergence* to provably null the adversaries' attack gains in intelligent, unknown DDoS. It designs a meta-policy to recursively combine their local weak (yet divergent) filters to form a strong global defense *without* knowing the attack strategies or new threats. This cooperative meta-policy forces the adversaries to be trapped into the Nash Equilibrium with a negligible benefit-cost ratio, thus demotivating them to launch attacks. EID inherently incentivizes network nodes to participate with improved network service availability. EID can scale up to massive filters via aggregation and scale out to massive network nodes via distributed on-demand DDoS demotivation. It is incrementally deployable with random-drop filters in commodity nodes today. We next describe EID's key intuitions. The detailed design will be presented in §5.

Multi-hop traffic divergence (§5.1). Intuitively, to demotivate all intelligent DDoS attacks, EID should minimize (or even void) the adversaries' maximal gains regardless of their attack strategies or (unknown) vulnerabilities. This is deemed impossible for a single defender: As proved in Theorem 2, due to the DDoS-legitimacy indistinguishability, even the optimal single-hop filters do not suffice to prevent adversaries from launching DDoS. To this end, EID seeks to bypass this fundamental limit via multi-hop mitigations.

We observe that, multi-hop traffic divergence offers a ubiquitous and strong paradigm to help trap smart attackers. As we will show in §7.1, real network traffic from different nodes is highly diverse due to heterogeneous user behaviors and network capabilities. Consider the case when nodes on the path adopt local optimal traffic filters in §3. Since each hop's filter in Equation 3 is based on its local legitimate traffic distribution, the adversary now faces the dilemma of *whose traffic* it should mimic (exemplified in FIgure 3). With large traffic divergence, mimicking one hop's traffic causes significant difference from other hops', thus lowering the success of passing all hops' filters for effective DDoS. This dilemma persists for the adversary regardless of its strategies or exploited (new) threats. By taking advantage of their traffic divergence, defenders can always cooperatively nulls adversaries' attack gains. Note that unlike traditional single-hop defenses, such multi-hop divergence does *not* seek to help explicitly distinguish between malicious and normal traffic. Instead, it bypasses "indistinguishability" by demotivating adversaries to run DDoS in the first place.

Clearly, the success of aforementioned DDoS demotivation largely relies on how "divergent" multi-hop traffic distributions are. In §5.1, we will define the multi-hop traffic divergence by generalizing the f-divergence [36] in our context. We will elaborate on its intuitions in intelligent DDoS, basic properties, and the relation to f-divergence and our optimal DDoS filters in §3.

Divergence-boosted demotivation policy (§5.2): With traffic divergence, EID recursively combines weak (yet divergent) local filters from multi-hop defenders to form a strong global DDoS demotivation. This meta-policy offers three appealing properties:
(i) Global DDoS demotivation: With sufficient traffic divergence, the adversary's benefit-cost ratio asymptotically converges to 0, regardless of its attack strategies or exploited vulnerabilities;
(ii) Local incentives of cooperation: As the adversary's optimal DDoS traffic holistically adapts to *all* hops' traffic distributions for maximal gains, it will significantly deviate from each hop's *local* legitimate traffic. This facilitates DDoS-legitimacy classification at each hop, thus motivating defenders to join EID for boosted local service availability (asymptotically 100%).
(iii) Resiliency to traffic dynamics: Multi-hop traffic divergence persists despite temporal traffic dynamics. Therefore, EID is tolerant to traffic dynamics and retains its effectiveness against DDoS. It does not mandate accurate implementations of local optimal filters.

Scalable on-demand demotivation protocol (§5.3) Although hop-by-hop filters demotivate intelligent DDoS via traffic divergence, they are also expensive with additional per-packet processing overhead. To this end, EID offers a fully distributed protocol that minimizes the number of hops needed to demotivate ongoing DDoS attacks. This protocol leverages each hop's local incentives to participate in the DDoS demotivation on demand, *without* requiring centralized coordination or additional signaling overhead.

[4]From adversarial machine learning perspective, it is well known that the training of GAN discriminators (i.e., traffic filters in our context) are vulnerable to noises [58, 59].

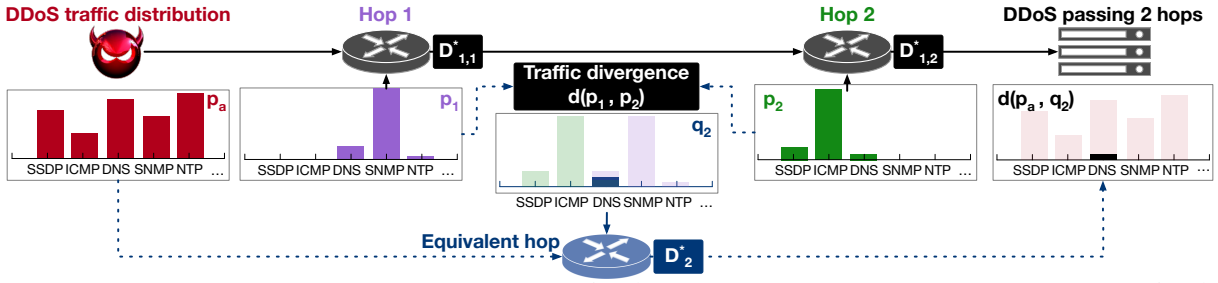


Figure 3: EID's intuition: Multi-hop traffic divergence (§5.1) and recursive combination of divergent filters (§5.2).

Showcase: Amplification DDoS (§5.4) We next showcase EID's powerfulness with a concrete example of amplification DDoS attacks, one of the most popular DDoS attacks in reality [6–8]. We detail adversaries' optimal strategies at the Nash Equilibrium, and demonstrate how EID demotivates the adversaries by compressing their bandwidth amplification factors (i.e., benefit-cost ratio).

5 THE EID DESIGN

We next elaborate on each solution component in EID.

5.1 Multi-hop Traffic Divergence

EID's core idea is to exploit *multi-hop traffic divergence* to demotivate *any* intelligent DDoS. We next define the traffic divergence, explain its intuitions in the DDoS context, list its key properties, and clarify its relationship with f-divergence in statistics.

Definition: Figure 3 visualizes the multi-hop traffic divergence in a network. We start with two hops. As shown in §3.1, each hop's traffic distribution $\mathbf{p} = (\mu_p, p_1, ..., p_k, ...)$ is a tuple of its total capacity μ_p and probability of traffic types $k = 1, 2, ...K$ with $\sum_k p_k = 1$[5]. Given two hops with traffic distributions $\mathbf{p} = (\mu_p, p_1, ..., p_k, ...)$ and $\mathbf{q} = (\mu_q, q_1, ..., q_k, ...)$, we define their traffic divergence $d(\mathbf{p}, \mathbf{q})$ as

$$d(\mathbf{p}, \mathbf{q}) \triangleq \sum_k \left(\frac{1}{\mu_p p_k} + \frac{1}{\mu_q q_k} \right)^{-1} \tag{4}$$

We next generalize this to n hops. Given n hops with distributions $\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_n} (n \geq 2)$, we define their traffic divergence as

$$d_n = d(\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_n}) \triangleq \sum_k \left(\sum_{j=1}^{n} \frac{1}{\mu_j p_j^k} \right)^{-1} = \sum_k d_n^k \tag{5}$$

where $d_n^k = (\sum_{j=1}^{n} \frac{1}{\mu_j p_j^k})^{-1}$ is the divergence for each traffic type k. Note this n-hop traffic divergence is equivalent to a 2-hop traffic divergence in Equation 4 by the following recursion:

$$d_{n+1} = d(\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_{n+1}}) = d(\mathbf{p_{n+1}}, \mathbf{q_n})$$

where $\mathbf{q_n} = (d_n, q_n^1, ...q_n^k)$ is an *equivalent traffic distribution* with total capacity as $d_n = d(\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_n})$ and

$$q_n^k = \frac{1}{d_n} \left(\frac{1}{\mu_n p_n^k} + \frac{1}{d_{n-1} q_{n-1}^k} \right)^{-1} = \frac{d_n^k}{d_n} \tag{6}$$

as probability for traffic type k. Such recursive definition ensures all properties in 2-hop divergence also apply to n-hop divergence.

Intuitions behind this definition: The traffic divergence in Equation 4 and 5 reflect two traffic distributions' "distance" in our optimal DDoS filters in §3.1. As visualized in Figure 3, assume a hop with legitimate traffic $\mathbf{p_2}$ seeks to filter malicious traffic passing the previous hop with distribution $\mathbf{p_1}$. With the optimal filters in §3.1, the remained total malicious traffic equals their traffic divergence:

$$\sum_k \mu_1 p_1^k D_{1,2}^*(k) = \sum_k \frac{\mu_1 p_1^k \cdot \mu_2 p_2^k}{\mu_1 p_1^k + \mu_2 p_2^k} = d(\mathbf{p_1}, \mathbf{p_2})$$

Recall that from GAN's perspective, the optimal filters in §3.1 maximize the discrimination between $\mathbf{p_1}$ and $\mathbf{p_2}$ (Theorem 1). The more divergent $\mathbf{p_1}$ and $\mathbf{p_2}$ are, the less $\mathbf{p_1}$'s traffic remains after passing $\mathbf{p_2}$'s optimal filters, and therefore the smaller $d(\mathbf{p_1}, \mathbf{p_2})$ is. After the filtering, the remained malicious traffic for each type k forms a new traffic distribution $\mathbf{q_2} = (d_2, q_2^1, ...q_2^k)$ as defined in Equation 6. This new distribution is passed to the next hop for filtering, and results in the recursive n-hop traffic divergence definition in Equation 5.

Basic properties of multi-hop traffic divergence: The following theorem lists key properties for later designs (proved in Appendix C). As we will see in §5.2, these properties enable a powerful paradigm to demotivate intelligent, unknown DDoS attacks.

Theorem 3 (Properties of Traffic divergence). *The traffic divergences in Equation 4 and 5 always guarantee the following properties:*

- **Symmetry:** $d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p})$;
- **Bounded & non-negativity:** $0 \leq d(\mathbf{p}, \mathbf{q}) \leq \frac{1}{4}(\mu_p + \mu_q)$;
- **Identity:** $d(\mathbf{p}, \mathbf{q})$ *is maximized if and only if* $\mathbf{p} = \mathbf{q}$, *i.e.,* $\mu_p = \mu_q$ *and* $p_k = q_k, \forall k$; *and*
- **Monotonicity:** $d_n \leq \min(d_{n-1}, \mu_n) \leq \min_{i \in [1,n]} \mu_i$, *and "=" holds if and only if* $d_{n-1} = d(\mathbf{p_1}, ..., \mathbf{p_{n-1}}) = 0$.

Relationship with f-divergence: Equation 4 and 5 can be viewed as an extension of f-divergence [36], a classical metric to measure the "distance" between two probability distributions in statistics. Given a parameterized convex function $f(x)$ with $f(1) = 0$, the f-divergence is defined as $\text{Div}_f(\mathbf{p}||\mathbf{q}) = \sum_k q_k \cdot f\left(\frac{p_k}{q_k}\right)$. In the DDoS context, EID generalizes f-divergence in two aspects, while f-divergence cannot quantify the divergence due to heterogeneous network capacity. First, EID considers network capacity rather than probability distribution only. As a special case, when $\mu_p = \mu_q = \mu$ (identical total traffic), EID's traffic divergence is equivalent to f-divergence: $d(\mathbf{p}, \mathbf{q}) = \mu \sum_k \frac{p_k q_k}{p_k + q_k} = \mu \left[\frac{1}{2} - \text{Div}_f(\mathbf{p}||\mathbf{q})\right]$, where $f(x) = \frac{1}{2} - \frac{x}{x+1}$. Second, EID generalizes its traffic divergence to arbitrary n distributions that f-divergence cannot.

5.2 Divergence-Boosted Demotivation Policy

With multi-hop traffic divergence, EID empowers defenders with a meta-policy to demotivate intelligent, unknown DDoS attacks. This

[5] As we will detail in §5.3, the granularity of EID's traffic type space K is flexible, e.g., per-protocol level, per packet header level, and packet header+payload level (Figure 4).

meta-policy combines weak (yet divergent) local filters from multi-hop defenders to form a strong global DDoS mitigation. We prove how it nulls the attacker's benefit-cost ratio at Nash Equilibrium, explain how it incentivizes multi-hop defenders to participate, and discuss its resiliency to network traffic dynamics.

Recursive combination of divergent filters: Consider the multi-hop DDoS filters in Figure 3. For each hop n, its local traffic is an aggregation of legitimate and DDoS traffic. With multi-hop filters, the DDoS traffic has been filtered by the previous $n-1$ hops before reaching n. So hop n's incoming rate of traffic type k is

$$o_n^k = \mu_n p_n^k + \mu_a p_a^k D_{n-1}(k)$$

where $\mathbf{p_n} = (\mu_n, p_n^1, ..., p_n^k, ...)$ is hop-n's local legitimate traffic, $\mathbf{p_a}$ is adversary's DDoS traffic, and $D_{n-1}(k)$ is the accumulative traffic filter for k from previous $n-1$ hops and is derived recursively as

$$D_n(k) = D_{n-1}(k) \cdot D_{1,i}(k)(n > 1), D_1(k) = D_{1,1}(k)$$

If all n hops adopt local optimal filters in §3.1, the following theorem shows the accumulative filter $D_n^*(k)$ is equivalent to an optimal *local* DDoS filter in Theorem 1 over an equivalent legitimate traffic distribution $\mathbf{q_n}$ (proved in Appendix D)

Theorem 4 (Recursive combination of divergent filters)**.** *The accumulative n-hop optimal filter $D_n^*(k)$ is equivalent to a single-hop optimal filter in Equation 3 with legitimate traffic $\mathbf{q_n} = (d_n, q_n^1, ..., q_n^k, ...)$:*

$$D_n^*(k) = \frac{d_n q_n^k}{d_n q_n^k + \mu_a p_a^k} \tag{7}$$

$$q_n^k = \frac{1}{d_n}\left(\frac{1}{\mu_n p_n^k} + \frac{1}{d_{n-1} q_{n-1}^k}\right)^{-1} = \frac{d_n^k}{d_n} \tag{8}$$

$$d_n = d(\mathbf{p_n}, \mathbf{q_{n-1}}) = d(\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_n}) \tag{9}$$

Theorem 4 bridges EID's multi-hop DDoS filters with traffic divergence. Figure 3 visualizes its intuition. As explained in §5.1, multi-hop traffic divergence d_n and equivalent legitimate traffic distribution $\mathbf{q_n}$ reflect the remaining traffic after passing all hops' DDoS filters. The remaining DDoS traffic would thus be the divergence between the original DDoS traffic distribution $\mathbf{p_a}$ and equivalent legitimate traffic distribution $\mathbf{q_n}$, thus resulting in Theorem 4. Although each local filter may be weak (§3.1), their combination can be strong due to traffic divergence. To this end, EID adopts Theorem 4 to combine divergent multi-hop filters for DDoS mitigation.

Global demotivation for intelligent DDoS: With EID's multi-hop filters, we next analyze how much an intelligent adversary can gain at most from launching DDoS. To optimize its attack strategy under multi-hop filters, the adversary should follow §3.2 under the accumulative optimal filter D_n^*. To pass multi-hop filters, the adversary must holistically adapt its DDoS traffic to all hops' legitimate traffic. If multi-hop defenders' traffic divergence is huge, the adversary faces the dilemma of whose traffic to mimic. We show that, its benefit-cost ratio will decrease with the traffic divergence (proved in Appendix E):

Theorem 5 (Demotivation with traffic divergence)**.** *Regardless of its attack strategy and capacity μ_a, any adversary in §2 satisfies*

- *its benefit-cost ratio $\eta(\mathbf{p_a}, \mathbf{D_n^*})$ and attack gain $U(\mathbf{p_a}, \mathbf{D_n^*})$ decrease monotonically with traffic divergence $d_n^k, \forall k$ and*

$$\lim_{d_n \to 0} \eta(\mathbf{p_a}, \mathbf{D_n^*}) = 0, \lim_{d_n \to 0} U(\mathbf{p_a}, \mathbf{D_n^*}) = 0$$

- *its benefit-cost ratio $\eta(\mathbf{p_a}, \mathbf{D_n^*})$ and attack gain $U(\mathbf{p_a}, \mathbf{D_n^*})$ decrease monotonically with hop count n and*

$$\lim_{n \to \infty} \eta(\mathbf{p_a}, \mathbf{D_n^*}) = 0, \lim_{n \to \infty} U(\mathbf{p_a}, \mathbf{D_n^*}) = 0$$

Theorem 5 unveils two powerful options in EID to demotivate *any* intelligent DDoS. First, if two hops' traffic divergence is large, Theorem 5 ensures no adversary can gain from DDoS. Second, in case two-hop traffic divergence is small, EID guarantees to enlarge the traffic divergence with more hops (Theorem 3), thus still provably demotivating intelligent DDoS. Of course, more hops imply higher traffic processing latency and costs. In §5.3, we will show how EID adapts the active defenders for on-demand demotivation.

Local incentives to join EID: We next switch to defenders' perspective and investigate their incentives of adopting EID. To maximize its attack gain, the adversary should holistically adapt its DDoS traffic $\mathbf{p_a}$ to the equivalent legitimate traffic distribution $\mathbf{q_n}$. But from each hop's perspective, this optimal strategy results in significant divergence between DDoS traffic $\mathbf{p_a}$ and local legitimate traffic $\mathbf{p_n}$. This helps each defender refine its local filter accuracy and network service availability. Specifically, at Nash Equilibrium of the intelligent DDoS game, once the adversary is demotivated to launch attacks, each defender will end up with 100% local filter accuracy and network availability. The following theorem quantifies EID's improvement of local filter's accuracy (proved in Appendix F):

Theorem 6 (Boosted local filters with EID)**.** *At Nash Equilibrium, for each hop m (m=1,2,...,n), its local filter satisfies $\lim_{d_n \to 0} D_{1,m}^*(k) = 1$ and $\lim_{n \to \infty} D_{1,m}^*(k) = 1, \forall k$ (i.e., no traffic blocked).*

Natural support for multiple attack paths: The results so far are presented with same attack path for simplicity, but EID naturally supports multiple paths. The key is that, multi-hop traffic divergence remains for every target's path. Attacking multiple targets with the same infrastructure increases adversaries' benefits, but is still mitigable by EID with each path's multi-hop divergence. All results thus still hold for adversaries/victims on different paths.

Resilience to network traffic dynamics: EID is more robust to network dynamics than single-hop defenses. Recall traffic dynamics make it difficult to accurately estimate legitimate traffic distribution that the single-hop filters in §3.1 mandate. Instead, multi-hop traffic divergence relaxes the reliance on accurate traffic estimation. Despite the temporal dynamics of multi-hop traffic, their divergence remains and persists. This suffices to demotivate intelligent DDoS in EID, as we will evaluate in §7.2. Each hop's local traffic dynamics only affect its own filter's accuracy, which can also be compensated by traffic divergence according to Theorem 6.

5.3 On-Demand Demotivation Protocol

We next convert EID's meta-policy in §5.2 to a scalable, distributed protocol. We require the EID protocol should *scale up* to excessive exploited protocols and applications inside each node, and *scale out* to massive nodes (hops) in the large network. In achieving so, we

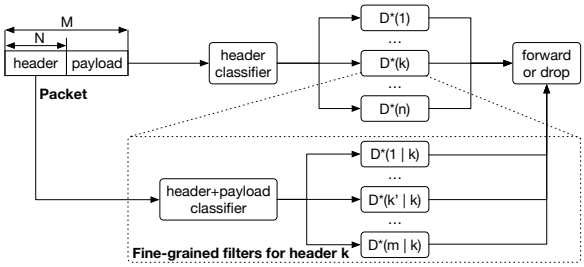


Figure 4: An example of EID's composable filters.

Algorithm 2 EID's distributed, on-demand hop pruning.

Input: Runtime traffic $o_n^k(t)$ and maximal load μ_n^k at hop n
Output: Whether filter $D^*_{1,n}(k)$ be activated at time slot t

$T \leftarrow 0$; ▷ *Initialization at $t = 0$.*
if $D^*_{1,n}(k, t) < D^*_{1,n}(k, t-1)$ **then** ▷ *Local filter is less accurate. Enable it when necessary.*
  $T \leftarrow \max(\mu_{\max}, T+1)$; ▷ *Wait other hops to enable their local filters*
  if $T \geq \mu_n^k$ **then return true**; ▷ *Now n is the inactive hop with the smallest μ_n^k*
else ▷ *Local filter is more accurate. Disable it when necessary.*
  $T \leftarrow \min(0, T-1)$; ▷ *Wait other hops to disable their local filters*
  if $T = 0$ **then return false**; ▷ *Now n is the active hop with the largest μ_n^k*
end if

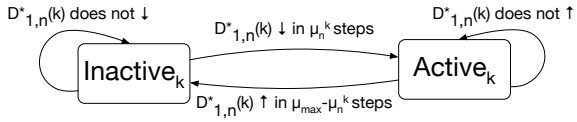


Figure 5: EID's protocol state machine for hop pruning.

face the challenge of balancing the DDoS mitigation efficiency and system overhead. This tradeoff spans on both dimensions:

∘ *Scale-up: Finer-grained filters at each hop.* With the rise of layer 7 intelligent DDoS attacks, it is not always sufficient to detect malicious traffic by checking the packet header only. Instead, deep packet inspections (DPIs) have been widely adopted to check more contents in each packet for DDoS detection. However, finer-grained detection incurs more filters (larger K) and per-packet processing costs that some nodes cannot afford (e.g., low-end IoT devices and latency-sensitive edge nodes). To scale up inside each hop, EID should seamlessly balance filter granularity (accuracy) and costs.

∘ *Scale-out: Multi-hops in large networks.* As shown in Theorem 5, EID is more effective with more hops (thus larger divergence). But hop-by-hop filtering can be expensive: The per-packet processing costs accumulate with more EID nodes. Moreover, it is not necessary to activate all nodes all the time. Large traffic divergence among fewer nodes also suffices to deter DDoS. To scale to large networks, EID should adapt the active hops for low-cost demotivation.

To this end, EID devises *on-demand* DDoS demotivation. To scale up to finer-grained filters at each hop, EID supports *composable filters* to seamlessly balance the filter granularity (accuracy) and cost. To scale out to large networks, EID runs distributed *hop pruning* based on runtime DDoS severity. We next elaborate on each.

Scaling-up: Composable filters at each hop. EID supports finer-grained detection by customizing the traffic type space K (e.g., from header space to entire payload, as exemplified in Figure 4). The cost, however, is more per-packet processing delay and system overhead. The granularity-cost tradeoff depends on each node's capability and demand. For example, some delay-sensitive edge and resource-constrained IoT devices may prefer coarse-grained filter.

EID supports seamless latency-accuracy tradeoff with *composable optimal filter.* As exemplified in Figure 4, any coarse-grained filter (e.g., based on header space only) can be decomposed into multiple fine-grained filters (e.g., based on header and payload), without hurting the resilience to DDoS. Specifically, for all packets with header k, if finer-grained filters are used, the aggregated coarse-grained filter for k can be derived from the Bayesian rule:

$$D_{1,n}(k) = \frac{\sum_{k'.hdr=k} D^*_{1,n}(k'|k) \cdot (\mu_i p_i^{k'|k} + \mu_a p_a^{k'|k})}{\sum_{k'.hdr=k} (\mu_i p_i^{k'|k} + \mu_a p_a^{k'|k})}$$
$$= \frac{\sum_{k'.hdr=k} \mu_i p_i^{k'|k}}{\sum_{k'.hdr=k} (\mu_i p_i^{k'|k} + \mu_a p_a^{k'|k})} = \frac{\mu_i p_i^k}{\mu_i p_i^k + \mu_a p_a^k} = D^*_{1,n}(k)$$

which equals to the coarse-grained optimal filter based on header only. All properties in §5.1 still hold. So each node can customize optimal filters with flexible, hybrid granularity and seamless latency-accuracy tradeoff (e.g., coarse-grained filter for delay-sensitive edge applications, and fine-grained ones for reliability-sensitive traffic).

Scaling-out: Distributed, on-demand hop pruning. To scale out to massive network nodes, EID adaptively prunes active hops, while retaining demotivation against DDoS. This results in two questions: (a) What are the right criteria for "sufficient" demotivation? (b) Which nodes should be (de)activated under this criteria?

∘ *Criteria for "sufficient" divergence for demotivation.* Intuitively, EID should invoke more nodes to defend against serious DDoS attacks, and inactivate them if there are no attacks. In our protocol, each node locally signals the DDoS severity and thus required divergence based on its local service availability $D^*_{1,n}(k)$ (§3.1), and decides whether to join EID for its own merits (i.e., boosted local accuracy for higher service availability).

∘ *On-demand hop (de)activation.* For effective DDoS mitigation with low costs, EID incrementally activates (deactivates) a node that maximizes (minimizes) the increment (decrement) of traffic divergence. Consider a path with n nodes, m of which have been activated for traffic type k. Under high DDoS threats, the next node to activate EID for k is the one with the smallest $\mu^k_{m+1} = \mu_{m+1} p^k_{m+1}$, such that the increment of traffic divergence is maximized according to Equation 5. Similarly, under low DDoS threats, the next node to deactivate is the one with the largest μ^k_{m-1}, such that the decrement of divergence is minimized. In this way, EID retains deterrence against DDoS with minimal active nodes (thus lowest costs).

Algorithm 2 and Figure 5 illustrate how EID realizes this on-demand node pruning in a fully distributed, signaling-free fashion. We assume a discrete-time model with a per-defined time slot size (agreed by all nodes). At each time slot, each node locally signals the DDoS severity based on Equation 2. When DDoS threat raises, each inactive node m starts to wait for μ^k_{m+1} time slots before activating its local filter for k. Therefore, nodes with smaller μ^k_{m+1} will activate its local filter earlier and increase *all* nodes' local service availability. If m observes the increment of its local accuracy before timeout, it means other better nodes have joined and successfully mitigated DDoS. Then it does not need to join. Otherwise, upon timeout, m would be the next best node to join. Similarly, when DDoS becomes less serious, this backoff mechanism ensures nodes with larger μ^k_{m+1} leaves earlier, thus ensuring on-demand pruning. This protocol is fully distributed without additional signaling costs. Moreover, it is incentive-compatible: Each node is self-motivated to join/leave EID based on its own merits (i.e., local service availability).

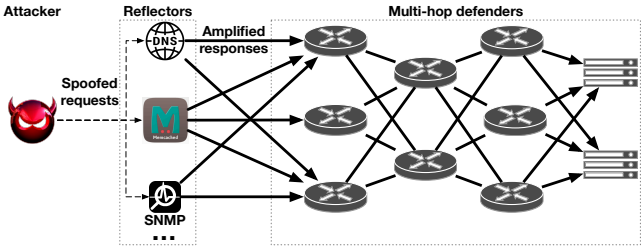


Figure 6: Intelligent amplification DDoS attacks.

5.4 Showcase: Amplification DDoS

We last use a concrete example to demonstrate the powerfulness of EID: Amplification DDoS. Since 2016, Amplification DDoS has contributed more than 80% large-scale DDoS attacks (>1 Tbps) in the world [6–8]. Figure 6 illustrates how it works. An adversary sends spoofed requests (with victim's IP address as the source) to a huge number of benign reflectors (amplifiers). Upon receiving these requests, the reflectors will reply much larger responses (usually 10–500× larger than request [47]) to the spoofed address (i.e., victim), exhaust the victim's network bandwidth or computing resource.

Amplification DDoS is popular today because adversaries can easily exhaust the victims with the help of reflectors at low costs. A well-known metric to quantify the benefits from the amplification DDoS is the bandwidth amplification factor: $\eta = \frac{\text{Response packet size}}{\text{Request packet size}}$. If $\eta > 1$, the adversaries have the motivation to leverage reflectors due to their amplification effect; otherwise direct attacks without reflectors are more beneficial. For example, attacks via open reflectors (e.g., DNS resolvers) have lower but non-zero costs (bandwidth consumed by attackers' spoofed requests to reflectors), thus unfavorable if reflectors' responses are smaller than spoofed requests ($\eta < 1$). Under EID, the expected amplification factor becomes

$$\eta(\mathbf{p_a}, \mathbf{D}) = \frac{U(\mathbf{p_a}, \mathbf{D})}{\mu_a} = \sum_k c_k \cdot p_a^k D_n^*(k) \tag{10}$$

$$U(\mathbf{p_a}, \mathbf{D}) = \sum_k c_k \cdot \mu_a p_a^k D_n^*(k) \tag{11}$$

where $U(\mathbf{p_a}, \mathbf{D})$ is the eventual amplified DDoS traffic (by reflectors) to the victim after passing EID's multi-hop filters, η is the average bandwidth amplification factor, c_k is the amplification factor for protocol k (exemplified in Figure 10), and $D_n(k)$ is EID's multi-hop filter in §5.2. Note that, such bandwidth amplification factor is a special case of the benefit-cost ratio in EID's threat model (§2). We first derive the adversary's optimal attack strategies under EID:

Lemma 1 (Optimal Amplification DDoS at Nash Equilibrium)**.** *Given the attack capacity μ_a, the adversary's optimal attack policy that maximizes $U(\mathbf{p_a}, \mathbf{D})$ and $\eta(\mathbf{p_a}, \mathbf{D})$ under EID is as follows:*
• **Mixed DDoS attack:** *If the amplification factors satisfy*

$$\max_k \frac{q_n^k \sqrt{c_k}}{\omega_n q_n^k + 1 - \omega_n} \le \mathbb{E}_{q_n}\sqrt{c} \le \min_k \frac{\sqrt{c_k}}{\omega_n} \tag{12}$$

where $\omega_n = \frac{d_n}{d_n + \mu_a}$ and $\mathbb{E}_{q_n}\sqrt{c} = \sum_k q_n^k \sqrt{c_k}$. Then the adversary's optimal amplification DDoS policy is to distribute the attack traffic with the following probability:

$$p_a^k = \frac{\omega_n q_n^k}{1 - \omega_n}\left(\frac{\sqrt{c_k}}{\omega_n \mathbb{E}_{q_n}\sqrt{c}} - 1\right) \tag{13}$$

and adversary's maximal amplification factor is

$$\max_{p_a} U(\mathbf{p_a^*}, \mathbf{D}^*) = d_n \mathbb{E}_{q_n}[c] - d_n \omega_n (\mathbb{E}_{q_n}[\sqrt{c}])^2 \tag{14}$$

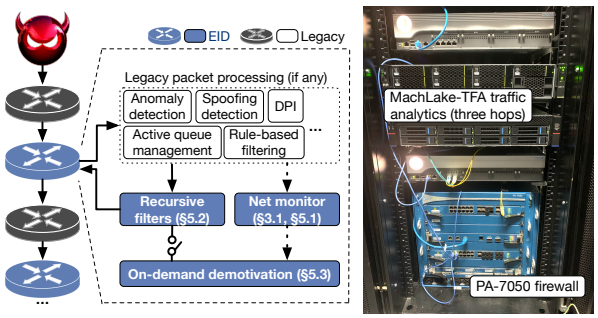
• **Simple DDoS attack:** *Otherwise, the adversary's optimal amplification DDoS policy is to choose only one protocol as*

$$p_a^k = \begin{cases} 1 & \text{if } k = \arg\max_j \frac{d_n q_n^k \cdot \mu_a c_k}{d_n q_n^k + \mu_a} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

and adversary's maximal amplification factor is

$$\max_{p_a} U(\mathbf{p_a^*}, \mathbf{D}^*) = \frac{d_n q_n^k \cdot \mu_a c_k}{d_n q_n^k + \mu_a} \tag{16}$$

In both cases, the optimal amplification factor $\max_{p_a} \eta(\mathbf{p_a^*}, \mathbf{D}^*) = \max_{p_a} U(\mathbf{p_a^*}, \mathbf{D}^*)/\mu_a$ and $(\mathbf{p_a^*}, \mathbf{D}^*)$ forms the Nash Equilibrium between attackers and defenders.



(a) Components in hybrid deployment **(b) Prototype setup**

Figure 7: Implementation of EID.

Lemma 1 uncovers adversaries' best amplification attack traffic under EID. In reality, both simple and mixed DDoS attack strategies are commonly observed [6–8, 60]. We next show that, both optimal attacks are bounded by traffic divergence (proved in Appendix H):

Theorem 7 (Amplification DDoS bound by traffic divergence)**.** *For any amplification DDoS strategy, EID always limits the adversary's attack gain and benefit-cost ratio with the following bound*

$$\max_{p_a} U(\mathbf{p_a}, \mathbf{D}^*) \le \begin{cases} d_n \mathbb{E}_{q_n}[c] & \text{if (12) holds} \\ d_n c_k & \text{otherwise} \end{cases} \tag{17}$$

$$\max_{p_a} \eta(\mathbf{p_a}, \mathbf{D}^*) \le \begin{cases} \frac{d_n}{\mu_a} \mathbb{E}_{q_n}[c] & \text{if (12) holds} \\ \frac{d_n}{\mu_a} c_k & \text{otherwise} \end{cases} \tag{18}$$

both bounds are monotonic to traffic divergence d_n^k for each k, and $\lim_{d_n \to 0} \max_{p_a} U(\mathbf{p_a}, \mathbf{D}^*) = 0$, $\lim_{d_n \to 0} \max_{p_a} \eta(\mathbf{p_a}, \mathbf{D}^*) = 0$.

Theorem 7 shows the maximal amplification factor decreases with traffic divergence. EID suffices to demotivate amplification DDoS when multi-hop traffic divergence $d_n \le \min(\frac{\mu_a}{E_{q_n}[c]}, \frac{\mu_a}{c_k})$ for $\max_{p_a} \eta(\mathbf{p_a}, \mathbf{D}^*) \le 1$. We will experimentally validate this in §7.2.

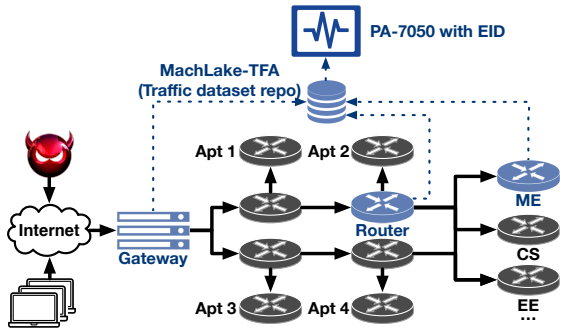


Figure 8: Simplified campus network topology in tests.

6 IMPLEMENTATION

We focus on three aspects of EID implementation: (1) How to realize EID in existing commodity nodes? (2) How can EID and state-of-the-art DDoS defenses coexist and mutually benefit each other? (3) How to incrementally deploy EID in today's networks?

To this end, our implementation follows three principles:

(I) *EID based on random-drop filters:* We realize EID as random-drop policies that have been widely supported by commodity hosts [61, 62], routers [63, 64], and firewalls [65]. This allows EID to be readily implemented in today's networks.

(II) *EID as a post-processor:* To coexist with existing defenses and take their advantages, EID is packaged as a post-processor of traffic *after* existing defenses (e.g., spoofing detection, queue management, rule-based filters, etc). This approach ensures EID works coherently with existing defenses. It takes advantage of existing defenses against well-known threats, and complements them to defend *unknown, intelligent* DDoS threats.

(III) *Hybrid EID deployment:* EID can coexist with legacy network nodes today. Due to EID protocol's distributed, signaling-free nature (§5.3), legacy nodes do not need to make any changes to work in concert with EID nodes.

Figure 7a shows EID's logical components at commodity nodes in hybrid deployments. At each EID-aware node, we add three modules as post-processors after existing defenses: The traffic load monitor that tracks the severity of DDoS (§3.1), the traffic divergence-based DDoS filters on top of readily-available random-drop filters (§5.1–5.2), and on-demand DDoS demotivation via filter reconfiguration (§5.3). At each hop n, these modules work as follows: (1) Estimate legitimate traffic $\mathbf{p_n}$. This can be either profiled offline or estimated online with standard traffic monitoring functions in commodity hosts/routers. Different from existing defenses in §3.1, EID does not need 100% accurate traffic distributions since multi-hop traffic divergence-based DDoS demotivation is resilient to traffic dynamics (§5.2); (2) For each traffic type k, add a new random-drop filter with activation threshold $\mu_n p_n^k$ and forward probability $D^*_{1,n}(k)$ in Equation 2. With multi-hops running EID, this achieves the optimal demotivation in §5.1–5.2; (3) For each rule in (b), the on-demand demotivation follows Algorithm 2 to (de)activate it based on runtime load monitor (§5.3); If the rule should be activated, concatenate it to the end of existing DDoS defense rules (*post-processing* principle). This takes advantage of existing defenses, and ensures incremental deployment and seamless rollback.

Prototype: We follow the above methodology to prototype EID on Palo Alto Networks Enterprise Firewall PA-7050 [66] in our campus network testbed (Figure 8). PA-7050 has built-in support for random-early drop policies, thus facilitating our EID implementation [65]. With its hardware processing offloading capability (6 network processing cards, each having 64 processing cores), this firewall supports up to 396 Gbps throughput with threat prevention, and up to 4M new sessions per second. For fair comparisons in experiments (§7), we use MachLake-TFA traffic analytics platform to remotely monitor and store three hops' pcap traces, redirect them to the firewall, implement one EID instance per hop, and concatenate their filters based on the topology (Figure 8).

7 EVALUATION

We evaluate EID with trace-driven experiments in an operational campus network. We first characterize multi-hop traffic divergence in reality (§7.1). Then we evaluate EID's effectiveness and costs from attacker (§7.2) and defender's (§7.3) perspectives.

Experiment environment: We evaluate EID in Tsinghua University's campus network, as shown in Figure 8. This network serves 59 departments and 45,000+ concurrent users. It experiences tens of noticeable DDoS attacks every day from U.S., China, Europe, and elsewhere. The DDoS attacks are diverse and mixed, including UDP flood, TCP SYN flood, amplification attacks (DNS, NTP, ICMP, SNMP, SSDP, Memcache, etc.), layer-7 DDoS, to name a few. To defend them, the network operator deploys a PA-7050 firewall [66] at the campus gateway, where we prototype EID as detailed in §6.

Dataset: We run a 24-hour data collection at three hops (blue nodes in Figure 8): The campus gateway, the intermediate core router, and the end department (Mechanical Engineering or ME). Within 24 hours, the ME department typically experiences 10s of noticeable DDoS within 24 hours, with 200-400 active victims in the department. We concurrently log pcap packet traces at all hops, resulting in a 49.8 TB dataset with 48,760,584 packets from 168 network protocols or applications. Moreover, to evaluate real DDoS attacks, we also collect a 24-hour amplification DDoS pcap trace from Alibaba cloud DDoS HoneyPot [67] (external open reflectors) . The honeypot voluntarily makes itself vulnerable to attract and lure hackers. The adversaries use their tools to exploit this "vulnerable" reflector, which allow us to monitor their attack traffic as ground truth.

Ethical evaluation: This work does not raise ethical issues. We responsibly conduct data collections and experiments. The dataset from all nodes was collected under the operator's approval. To avoid user privacy leakage, we anonymized raw pcap packets by removing privacy-sensitive payloads and anonymizing IP addresses. Moreover, we bear in mind that imperfect DDoS filters over the operational network can mistakenly block legitimate traffic, thus detrimental to user experiences (§3.2). Instead, we conduct an offline evaluation by replaying the above three-hop legitimate and DDoS traces in EID on the firewall.

7.1 Traffic Divergence and Dynamics in Reality

We characterize the multi-hop traffic divergence in the campus network. We replay the 24-hour traces from three nodes in Figure 8, estimate their traffic distributions every 10 minutes, and follow §5.1 to compute the traffic divergence d_3 and d_3^k for each protocol. We compare the runtime traffic divergence with per-hop traffic rate.

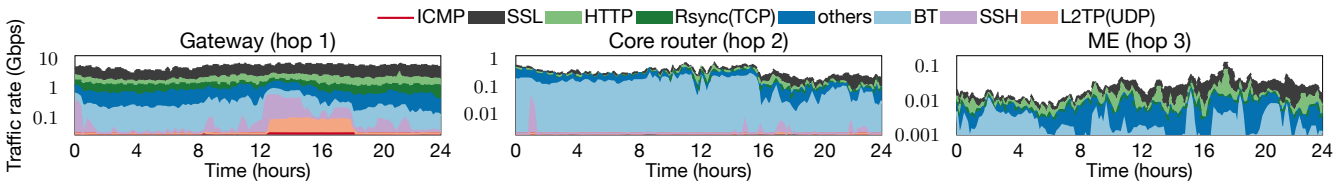


Figure 9: Temporal network traffic dynamics in our tested operational campus network.

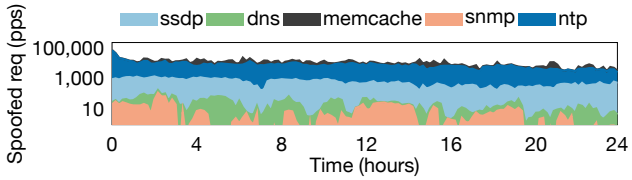


Figure 10: Amplification DDoS traffic captured by honeypot.

Figure 9 and Figure 10 plot the runtime traffic from our dataset. They validate that both traffic distributions vary dynamically over time. Moreover, Figure 10 shows adversaries exploit various protocols to run DDoS and adapt their attack strategies dynamically. As explained in §3.2, both incur "noises" to accurately differentiate legitimate and DDoS traffic and thus prevents the realization of optimal single-hop defenses. Meanwhile, the legitimate traffic between different nodes exhibits large diversity. In our tests, the campus gateway, core router, and ME department exhibit approximately 5.806 Gbps, 407.778 Mbps, and 35.535 Mbps total traffic, respectively. Their traffic distributions also differ and vary over time, which implies significant traffic divergence.

Figure 11 and Figure 12 show the temporal dynamics and statistics of multi-hop traffic divergence (normalized by the peak traffic rate for the ease of comparison). We make three observations. First, we confirm significant traffic divergence between the campus gateway, intermediate core router, and end department. Figure 12 shows the traffic divergence d_3 is 30% smaller than the smallest per-hop traffic rate (Lemma 3). Second, traffic divergences differ between protocols. For example, we note for SNMP and DNS, the end department (3rd hop)'s traffic rate is sometimes higher than the gateway (1st hop) due to the traffic exchange inside the campus. Such internal traffic proliferates the traffic diversity among network nodes. Third, similar to traffic dynamics, the multi-hop traffic divergence also varies over time. But as the harmonic mean of three hops' traffic (Equation 5), traffic divergence is usually smoother than per-hop traffic. All these observations imply that traffic divergence is a promising paradigm for DDoS mitigation.

7.2 Effectiveness Against Intelligent Attackers

We next evaluate EID from the attacker's perspective. We assess EID's effectiveness against intelligent DDoS with two concrete examples: reflective amplification DDoS (§5.4) and direct UDP flooding. We replay the three-hop legitimate traffic in Figure 9 in EID (§6). To attack them, we generate and test two DDoS attacks strategies: (1) **Optimal DDoS strategies** against EID in §5.4; (2) **Real DDoS strategies** observed in our dataset from the BotNet (Figure 10). For amplification DDoS, we evaluate adversaries' benefit-cost ratio in Equation 11, based on the real bandwidth amplification factors from CISA [47, 68](30.8 for SSDP, 28.7 for DNS, 51,000 for Memcache, 6.3 for SNMP, 556.9 for NTP, 3.8 for BitTorrent, 5.5 for Steam). For direct UDP flooding, Kaspersky reports each DDoS's profit margin can reach 95% [69, 70]. So we define its benefit-cost ratio $\eta = \sum_k c_k \cdot p_a^k D_n^*(k)$ where $c_k = \frac{\text{DDoS profits}}{\text{DDoS price}} \approx 1.95$ according to [69, 70] We repeat this experiment under different attack capacities μ_a (from 100Mbps to 1Tbps) and varying number of hops joining EID (from 1 to 3, with 1-hop filter being the state-of-the-art optimal defenses in §3.1). To evaluate EID against *unknown* threats, all nodes in these experiments have *no prior knowledge* of adversaries' DDoS distributions or strategies.

Overall effectiveness against intelligent DDoS: Figure 14 and Figure 16a plot the adversary's benefit-cost ratio under amplification and direct DDoS, respectively. We make three observations. First, for both optimal and real DDoS attacks, EID's optimal filters in §5.2 reduce most of adversaries' benefit-cost ratio to be $\eta < 1$ under tested attack capacities. Recall when $\eta < 1$, the adversary has no motivation to exploit reflectors to amplify the DDoS traffic (§5.4). So EID successfully demotivates the adversaries to launch such DDoS attacks. Second, as shown in Figure 14b, adversaries' benefit-cost ratio decreases as more hops join EID due to larger traffic divergence. Third, adding more attack capacity μ_a does not help the adversary gain more. Instead, as shown in Theorem 7, Figure 14 and Figure 16a, more attack capacity raises adversaries' costs and lowers benefit-cost ratios.

Comparison of DDoS attack strategies: Figure 14 and 16a confirm that the optimal DDoS strategy in Lemma 1 outperforms the real DDoS in Figure 10 in all tested scenarios. Note that the real attack strategy in Figure 10 always adopts mixed DDoS, while the optimal strategy in Lemma 1 may adopt simple DDoS sometimes. Figure 13 plots the frequency of mixed DDoS in the optimal strategy under different attack capacities and EID hop counts in our 24-hour experiment. The optimal strategy adopts mixed DDoS more frequently with larger attack capacity μ_a *or* higher traffic divergence (from 1 to 3 hops in this case), because both raise the adversaries' risk of being detected by EID if all DDoS traffic comes from a single protocol (i.e., simple DDoS). This phenomenon also explains the similar trend of benefit-cost ratio under 100Mbps-1Tbps attack capacities in Figure 14a, in which the adversary will always adopt mixed DDoS according to Figure 13. Even so, EID still guarantees the adversaries are demotivated even under the optimal attack.

Resiliency to traffic dynamics: Figure 14a and 16a confirm EID remain effective against DDoS under dynamic traffic. With traffic dynamics as "noises", each hop's local filters may deviate from the optimal ones (§3.2). But the traffic divergence persists despite traffic dynamics, thus ensuring always-on DDoS demotivation.

7.3 Efficiency and Overhead for Defenders

We next switch to the victims' perspective to evaluate their benefits and costs in adopting EID. When the adversaries are rational, §7.2 has shown that they will stop attacking, resulting in 100% legitimate

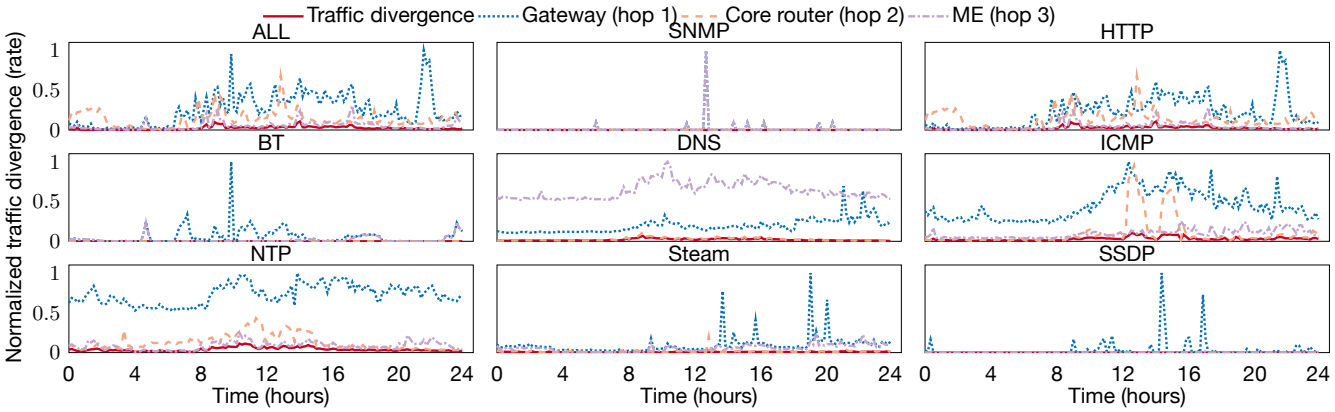


Figure 11: Temporal dynamics of multi-hop traffic divergence and rates in operational campus networks.

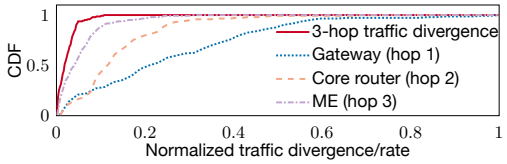


Figure 12: Comparison of traffic divergence and rates.

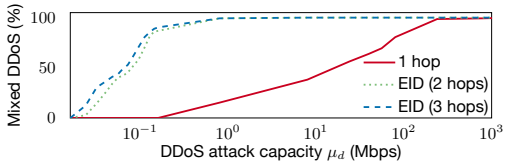


Figure 13: Frequency of mixed DDoS in the optimal attacks.

service availability for the victims and defenders. Instead, here we consider (irrational) adversaries that insist on attacking despite no merits for themselves. In this case, we repeat the experiments in §7.2 to quantify the defender's benefits and costs of joining EID.

Local incentives to participate EID: Figure 15 and Figure 16b plot the ME department (hop 3)'s local legitimate service availability (Equation 3 as proved in §3.1) under different attack capacities, attack strategies, and number of hops joining EID. We make two observations. First, as more nodes join EID, each hop's local service availability is also improved as explained in Theorem 6. This offers a strong local incentive for each node to join EID. Second, despite irrational adversaries, EID still retains ≈99.99% local service availability under real DDoS attacks. Under optimal DDoS attacks, EID ensures ≈84.78% service availability with 3 hops on average, which can be further improved with more hops and traffic divergence.

Scalable, on-demand demotivation: Following §5.3, we next evaluable EID's scalability for defenders in two aspects:

∘ *Scale-up: Composable filters at each hop.* As explained in §5.3, EID's composable optimal filters support seamless accuracy-cost tradeoff. Figure 17 evaluates the local legitimate service availability with local optimal filters based on total traffic volume ($K = 1$) and protocol type (K = num. protocols). The per-protocol filter achieves the highest local legitimate service availability. The filter based on total traffic volume is too coarse-grained, thus mistakenly blocks more legitimate traffic. Note finer-grained filters do not always necessarily offer higher service availability. In reality, many DDoS attacks exploit random spoofed source address and port numbers, resulting in insufficient samples for each filter. This raises statistical biases and reduce the accuracy of each filter. We suggest the EID users to carefully decide the filter granularity and traffic categories, e.g., based on the standard DDoS feature engineering.

∘ *Scale-out: On-demand hop pruning.* As shown in Figure 14b, two hops usually suffice to fully demotivate adversaries with 100Mbps–1Tbps attack capacity. The single-hop optimal filters in §3.1 cannot always demotivate adversaries (e.g., under optimal DDoS with μ_a=100Mbps, the bandwidth amplification factor can be more than 1), while three hops are more than necessary. In this case, EID's on-demand demotivation will activate the second hop according to Algorithm 2 to fully demotivate adversaries via traffic divergence.

CPU, memory, and signaling costs: EID incurs negligible system overhead for defenders. In our 24-hour tests, the PA-7050 firewall running always retains ≤0.5% CPU and ≤6.2GB with/without EID. The reason is that, EID reuses the mature and free random-drop filters in commodity devices. Moreover, EID has no signaling between nodes by design, thus affordable by large networks.

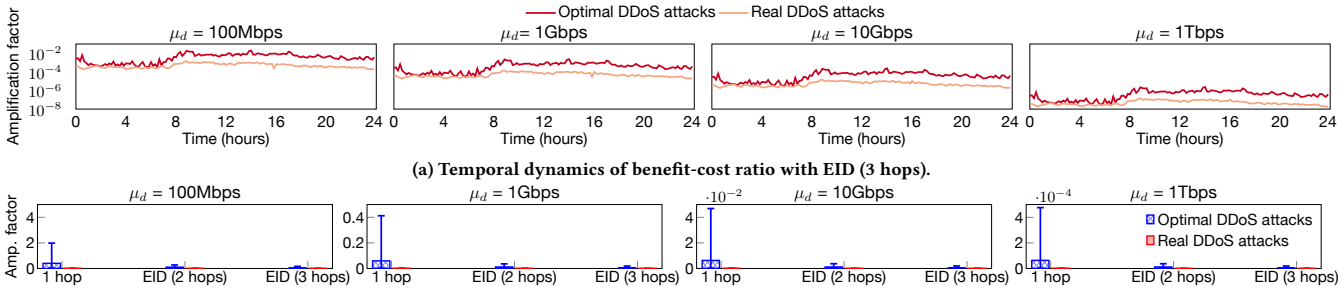
8 DISCUSSION

EID is our first step toward strategically deterring intelligent, unknown attacks. Although its results are encouraging, we believe EID can be further enhanced in at least three aspects:

Extended benefit-cost ratio: The current adversaries' utility $U(\mathbf{p_a}, \mathbf{D})$ in §2 only accounts for the DDoS traffic that bypasses filters. Beyond that, it is also possible to extend attackers' utility to consider mistakenly dropped legitimate traffic in $U(\mathbf{p_a}, \mathbf{D})$ by adding them to (i) and (ii) in §2. The key is that, EID's symmetry in Equation 3–8 ensures dropped legitimate traffic volume equals bypassed malicious traffic $\mu_d p_d^k D(k)$. All results still hold with minor constant factor updates (e.g., $2\eta(\mathbf{p_d}, \mathbf{D})$ in Theorem 7).

Beyond DDoS: While the experiments in §7 focus on volumetric DDoS, EID's core ideas of deterrence apply to other DDoS attacks (e.g., low rate attacks [49, 50]) since the threat model in §2 generally holds. Beyond DDoS, EID can be extended to deter other intelligent, unknown attacks with similar utilities in §2, since multi-hop traffic divergence offers a generic mechanism to null attack gains.

Beyond successful attacks: The current EID assumes adversaries only benefit from *successful* attacks. In reality, sometimes adversaries can also gain from *unsuccessful attacks* (e.g., extorting victims with fear, uncertainty, and doubt). In this case, EID needs generalizations to deter these attacks, but its core idea of lowering gains with multi-hop traffic divergence still holds.



(a) Temporal dynamics of benefit-cost ratio with EID (3 hops).

(b) Statistics of benefit-cost ratio under different attack capacities μ_a and number of hops joining EID.

Figure 14: Adversaries' benefit-cost ratio (bandwidth amplification factor for amplification DDoS) in EID.

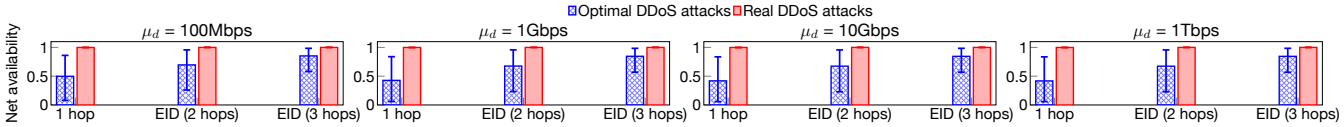
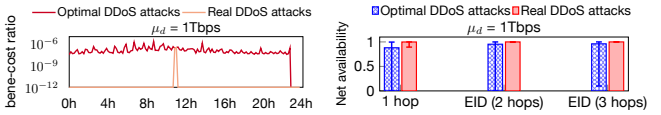


Figure 15: Victim's network service availability under different attack capacities μ_a in EID.



(a) Adversaries' benefit-cost ratio **(b) Victims' network service availability**

Figure 16: EID under direct flooding attacks.

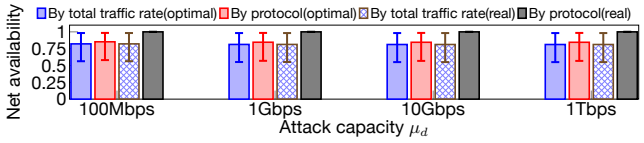


Figure 17: EID filter accuracy under different granularities.

9 RELATED WORK

DDoS has been a hot topic in network security for decades; see [56, 57] for a historical review. Numerous variants of DDoS threats have been revealed, including the flooding [12, 71], amplification DDoS [47, 48, 51, 52], low-rate attacks [49, 50], application-layer DDoS [34], to name a few; more taxonomies are available at [72, 73]. Therefore, diverse DDoS defenses have been proposed at *mechanism* and *policy* level. At the defense mechanism level, state-of-the-art DDoS defenses mainly adopt scrubbing centers [6, 27, 28], CDNs [74], rule-based filters [10–12], queue/congestion management [13, 75], pushback [76], spoof detection [15], bandwidth reservation [77], path traceback [78, 79], destination-defined defense [71, 80], BGP FlowSpec and Blackholing [81], to name a few. Meanwhile, some future network architectures are proposed with inherent resiliency to DDoS, including accountable IP [24, 25], SAVI [14], TVA [26], SCION [23, 82], ICN [83, 84], XIA [85], and more. EID is orthogonal to all these efforts: It strategically prevents attackers from launching intelligent DDoS based on multi-hop traffic divergence.

At the defense policy level, commercial DDoS mitigations [6, 27, 28, 65] mostly adopt rule-based filtering based on prior knowledge of threat features, which fall short in defending unknown threats (§3). Advanced mitigations can be realized by machine learning, such as anomaly detection [16–18, 86–88], reinforcement learning [34], entropy-based detection [21, 22], neural packet classification [89], context-specific feature engineering IoT [90], etc. Besides machine learning, game theory can also be adopted to analyze potential DDoS threats [91], design mechanisms to incentivize cooperative defenses [92], and mitigate specific DDoS scenarios such as bitcoin mining [93] and flooding [35, 94]. But as shown in §3.1, these defenses cannot eliminate strategic attacker's gains from launching smart attacks. Instead, EID leverages multi-hop traffic divergence for a game-theoretic deterrence against intelligent, unknown DDoS.

10 CONCLUSION

We propose EID, a strategic deterrence protocol against intelligent, unknown DDoS via multi-hop traffic divergence. As a game-theoretical solution, EID deters intelligent attackers by nulling their benefit-cost ratio, and motivates multi-hop defenders to collaborate. To achieve so, it exploits traffic divergence to recursively combine weak (yet divergent) filters from multi-hop defenders to form a provably strong deterrence against attackers. EID does not require prior knowledge of exploited vulnerabilities or attack strategies. It is scalable with distributed on-demand DDoS demotivation, and incrementally deployable in real networks with negligible costs.

EID is our first step to showcase the promises of deterring intelligent network attacks powered by unknown threats and adversarial machine learning. Instead of lagging behind numerous new threats and evolving attack strategies, a more practical solution should reverse the attacker-defender asymmetry to void the attack benefits. Beyond DDoS and traffic divergence in this work, we believe more opportunities lay ahead in this direction that could eventually stimulate endogenous security in future networked systems.

ACKNOWLEDGMENT

We greatly appreciate our shepherd, Mattijs Jonker, and anonymous reviewers for their valuable feedback. We thank QI-ANXIN Technology Research Institute for the help with malicious DDoS traffic analysis. This work is sponsored by the National Key Research and Development Plan of China (2018YFB1800301) and National Natural Science Foundation of China (61832013).

REFERENCES

[1] Frank Li and Vern Paxson. A large-scale empirical study of security patches. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 2201–2215, 2017.

[2] Frederico Araujo, Kevin W Hamlen, Sebastian Biedermann, and Stefan Katzenbeisser. From patches to honey-patches: Lightweight attacker misdirection, deception, and disinformation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (CCS)*, pages 942–953, 2014.

[3] Cloudflare. The largest DDoS attacks of all time. https://www.cloudflare.com/learning/ddos/famous-ddos-attacks/, 2020.

[4] WIRED. A 1.3-Tbs DDoS Hit GitHub, the Largest Yet Recorded. https://www.wired.com/story/github-ddos-memcached, 2018.

[5] A10. Five Most Famous DDoS Attacks and Then Some. https://www.a10networks.com/blog/5-most-famous-ddos-attacks/, 2020.

[6] Cloudflare. DDoS attack trends for Q4 2020. https://blog.cloudflare.com/network-layer-ddos-attack-trends-for-q4-2020/, Jan 2021.

[7] Cloudflare. DDoS attack trends for Q3 2020. https://blog.cloudflare.com/network-layer-ddos-attack-trends-for-q3-2020/, Nov 2020.

[8] China Telecom. DDoS Attack Landscape in 2019. https://www.nsfocus.com.cn/index.php?m=content&c=index&a=show&catid=222&id=162&template=download, 2020.

[9] MSSP Alert. Kaspersky Lab Study: Average Cost of Enterprise DDoS Totals $2M. https://www.msspalert.com/cybersecurity-research/kaspersky-lab-study-average-cost-of-enterprise-ddos-attack-totals-2m/, 2018.

[10] Paul Ferguson and Daniel Senie. rfc2827: network ingress filtering: defeating denial of service attacks which employ ip source address spoofing, 2000.

[11] Abraham Yaar, Adrian Perrig, and Dawn Song. Stackpi: New packet marking and filtering mechanisms for ddos and ip spoofing defense. *IEEE Journal on Selected Areas in Communications*, 24(10):1853–1863, 2006.

[12] Abraham Yaar, Adrian Perrig, and Dawn Song. Siff: A stateless internet flow filter to mitigate ddos flooding attacks. In *IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004*, pages 130–143. IEEE, 2004.

[13] Xin Liu, Xiaowei Yang, and Yong Xia. Netfence: preventing internet denial of service from inside out. *ACM SIGCOMM Computer Communication Review*, 40(4):255–266, 2010.

[14] Jianping Wu, Jun Bi, Marcelo Bagnulo, Fred Baker, and Christian Vogt. Source address validation improvement (savi) framework. *RFC7039*, 2013.

[15] Xin Liu Ang Li Xiaowei Yang and David Wetherall. Passport: Secure and adoptable source authentication. In *USENIX NSDI*, 2008.

[16] Fernando Silveira, Christophe Diot, Nina Taft, and Ramesh Govindan. Astute: Detecting a different class of traffic anomalies. *ACM SIGCOMM Computer Communication Review*, 40(4):267–278, 2010.

[17] Anukool Lakhina, Mark Crovella, and Christophe Diot. Diagnosing network-wide traffic anomalies. *ACM SIGCOMM computer communication review*, 34(4):219–230, 2004.

[18] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.

[19] Giovane CM Moura, Ricardo de O Schmidt, John Heidemann, Wouter B de Vries, Moritz Muller, Lan Wei, and Cristian Hesselman. Anycast vs. ddos: Evaluating the november 2015 root dns event. In *Proceedings of the 2016 Internet Measurement Conference*, pages 255–270, 2016.

[20] Cloudflare. How Anycast mitigates DDoS attacks. https://www.cloudflare.com/learning/cdn/glossary/anycast-network/, 2021.

[21] Shui Yu, Wanlei Zhou, Robin Doss, and Weijia Jia. Traceback of ddos attacks using entropy variations. *IEEE transactions on parallel and distributed systems*, 22(3):412–425, 2010.

[22] Jun Li, Minho Sung, Jun Xu, and Li Li. Large-scale ip traceback in high-speed internet: Practical techniques and theoretical foundation. In *IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004*, pages 115–129. IEEE, 2004.

[23] David Barrera, Laurent Chuat, Adrian Perrig, Raphael M Reischuk, and Pawel Szalachowski. The scion internet architecture. *Communications of the ACM*, 60(6):56–65, 2017.

[24] Andersen, David G and Balakrishnan, Hari and Feamster, Nick and Koponen, Teemu and Moon, Daekyeong and Shenker, Scott. Accountable Internet Protocol (AIP). In *ACM SIGCOMM Computer Communication Review*, volume 38, pages 339–350. ACM, 2008.

[25] David Naylor, Matthew K Mukerjee, and Peter Steenkiste. Balancing Accountability and Privacy in the Network. In *ACM SIGCOMM Computer Communication Review*, volume 44, pages 75–86. ACM, 2014.

[26] Xiaowei Yang, David Wetherall, and Thomas Anderson. A DoS-limiting network architecture. *ACM SIGCOMM Computer Communication Review*, 35(4):241–252, 2005.

[27] Akamai Cloud Security for DDoS Protection. https://www.akamai.com/us/en/products/security/ddos-protection-service.jsp, 2021.

[28] Amazon AWS Shield. https://aws.amazon.com/shield/, 2021.

[29] CPO. IoT-based DDoS attacks are growing and making use of common vulnerabilities. https://www.cpomagazine.com/cyber-security/iot-based-ddos-attacks-are-growing-and-making-use-of-common-vulnerabilities/, 2020.

[30] Alibaba Cloud. Pricing for DDoS protection. https://cn.aliyun.com/price/detail/ddos, 2021.

[31] Networkworld. The rise of artificial intelligence DDoS attacks. https://www.networkworld.com/article/3289108/the-rise-of-artificial-intelligence-ddos-attacks.html.

[32] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[33] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning (ICML)*, pages 214–223. PMLR, 2017.

[34] Yebo Feng, Jun Li, and Thanh Nguyen. Application-layer ddos defense with reinforcement learning. In *2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)*, pages 1–10. IEEE, 2020.

[35] Michael Walfish, Mythili Vutukuru, Hari Balakrishnan, David Karger, and Scott Shenker. DDoS Defense by Offense. In *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 303–314, 2006.

[36] Wikipedia. f-divergence. https://en.wikipedia.org/wiki/F-divergence, 2021.

[37] Jörgen W Weibull. *Evolutionary game theory*. MIT press, 1997.

[38] Herbert Gintis. *Game theory evolving*. Princeton university press, 2009.

[39] CNCERT. DDoS Attack Trend Report in China: 2020 Q1. https://www.cert.org.cn/publish/main/upload/File/DDoS2020-1.pdf, 2020.

[40] Alibaba Cloud and FreeBuf. DDoS Trend Report in 2020 Q1 and Q2. https://www.calder-systems.com/articles/paper/249963.html, 2020.

[41] Armor. Cybercrime-as-a-Service: Selling DDoS on the Dark Web. https://www.armor.com/resources/blog/cybercrime-as-a-service-selling-ddos-dark-web/, 2018.

[42] Radware. Malware and Botnet Attack Services Found on the Darknet. https://blog.radware.com/security/2016/07/malware-and-botnet-attack-services-found-on-the-darknet/, 2016.

[43] Mission Critical. The Dark Web: DDoS Attacks Sell for as Low as $10 per Hour. https://www.missioncriticalmagazine.com/articles/93185-the-dark-web-ddos-attacks-sell-for-as-low-as-10-per-hour, 2020.

[44] Mohammad Karami and Damon McCoy. Understanding the emerging threat of ddos-as-a-service. In *6th USENIX Workshop on Large-Scale Exploits and Emergent Threats)*, 2013.

[45] José Jair Santanna and Anna Sperotto. Characterizing and mitigating the ddos-as-a-service phenomenon. In *IFIP International Conference on Autonomous Infrastructure, Management and Security*, pages 74–78. Springer, 2014.

[46] José Jair Santanna, Roland van Rijswijk-Deij, Rick Hofstede, Anna Sperotto, Mark Wierbosch, Lisandro Zambenedetti Granville, and Aiko Pras. Booters: An analysis of ddos-as-a-service attacks. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 243–251. IEEE, 2015.

[47] Christian Rossow. Amplification hell: Revisiting network protocols for ddos abuse. In *NDSS*, 2014.

[48] Kührer, Marc and Hupperich, Thomas and Rossow, Christian and Holz, Thorsten. Exit from Hell: Reducing the Impact of Amplification DDoS Attacks. In *23rd USENIX Security Symposium*, pages 111–125, 2014.

[49] Aleksandar Kuzmanovic and Edward W Knightly. Low-rate tcp-targeted denial of service attacks and counter strategies. *IEEE/acm transactions on networking*, 14(4):683–696, 2006.

[50] Mina Guirguis, Azer Bestavros, Ibrahim Matta, and Yuting Zhang. Reduction of quality (roq) attacks on internet end-systems. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 1362–1372. IEEE, 2005.

[51] Soo-Jin Moon, Yucheng Yin, Rahul Anand Sharma, Yifei Yuan, Jonathan M Spring, and Vyas Sekar. Accurately measuring global risk of amplification attacks using ampmap. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*, 2021.

[52] Vern Paxson. An analysis of using reflectors for distributed denial-of-service attacks. *ACM SIGCOMM Computer Communication Review*, 31(3):38–47, 2001.

[53] Marianne Shaw. Leveraging good intentions to reduce unwanted network traffic. In *Proc. USENIX Steps to Reduce Unwanted Traffic on the Internet workshop*, page 8, 2006.

[54] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. *arXiv preprint arXiv:1606.00709*, 2016.

[55] Wikipedia. Deepfake. https://en.wikipedia.org/wiki/Deepfake, 2021.

[56] Eric Osterweil, Angelos Stavrou, and Lixia Zhang. 21 years of distributed denial-of service: Current state of affairs. *Computer*, 53(7):88–92, 2020.

[57] Eric Osterweil, Angelos Stavrou, and Lixia Zhang. 21 years of distributed denial-of-service: A call to action. *Computer*, 53(8):94–99, 2020.

[58] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *ICLR*, 2015.

[59] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (S&P)*, pages 39–57. IEEE, 2017.

[60] Mattijs Jonker, Alistair King, Johannes Krupp, Christian Rossow, Anna Sperotto, and Alberto Dainotti. Millions of targets under attack: a macroscopic characterization of the dos ecosystem. In *Proceedings of the 2017 Internet Measurement Conference (IMC)*, pages 100–113, 2017.

[61] NVIDIA Mellanox BlueField-2 Data Processing Unit (DPU). https://www.mellanox.com/files/doc-2020/pb-bluefield-2-dpu.pdf, 2020.

[62] Pensando DSC-25 Distributed Services Card. https://pensando.io/wp-content/uploads/2020/03/Pensando-DSC-25-Product-Brief.pdf, 2020.

[63] Cisco. Configuring Weighted Random Early Detection. https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/qos_conavd/configuration/xe-16/qos-conavd-xe-16-book/qos-conavd-cfg-wred.html, 2018.

[64] Juniper. Managing Congestion Using RED Drop Profiles and Packet Loss Priorities. https://www.juniper.net/documentation/us/en/software/junos/cos/topics/concept/red-drop-profile-overview-cos-config-guide.html, 2021.

[65] Palo Alto Networks (PAN) OS Administrator's Guide (v8.1). https://docs.paloaltonetworks.com/pan-os/8-1/pan-os-admin.html, 2021.

[66] Palo Alto Networks. PA-7000 Series Datasheet. https://www.paloaltonetworks.com/resources/datasheets/pa-7000-series, 2021.

[67] Alibaba cloud HoneyPot. https://help.aliyun.com/document_detail/196044.html, 2021.

[68] CISA. TA14-017A: UDP-Based Amplification Attacks. https://us-cert.cisa.gov/ncas/alerts/TA14-017A, 2014.

[69] Kaspersky. Criminal Benefits: Profit Margin of a DDoS Attack Can Reach 95%. https://www.kaspersky.com/about/press-releases/2017_criminal-benefits, 2017.

[70] Sourcelist. The cost of launching a DDoS attack. https://securelist.com/the-cost-of-launching-a-ddos-attack/77784/, 2017.

[71] Avi Yaar, Adrian Perrig, and Dawn Song. An endhost capability mechanism to mitigate ddos flooding attacks. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2004.

[72] Saman Taghavi Zargar, James Joshi, and David Tipper. A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks. *IEEE communications surveys & tutorials*, 15(4):2046–2069, 2013.

[73] Jelena Mirkovic and Peter Reiher. A taxonomy of ddos attack and ddos defense mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2):39–53, 2004.

[74] Mattijs Jonker, Anna Sperotto, Roland van Rijswijk-Deij, Ramin Sadre, and Aiko Pras. Measuring the Adoption of DDoS Protection Services. In *Proceedings of the 2016 Internet Measurement Conference (IMC)*, pages 279–285, 2016.

[75] Xinzhe Fu and Eytan Modiano. Fundamental limits of volume-based network dos attacks. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(3):1–36, 2019.

[76] Ratul Mahajan, Steven M Bellovin, Sally Floyd, John Ioannidis, Vern Paxson, and Scott Shenker. Controlling high bandwidth aggregates in the network. *ACM SIGCOMM Computer Communication Review*, 32(3):62–73, 2002.

[77] Cristina Basescu, Raphael M Reischuk, Pawel Szalachowski, Adrian Perrig, Yao Zhang, Hsu-Chun Hsiao, Ayumu Kubota, and Jumpei Urakawa. SIBRA: Scalable Internet Bandwidth Reservation Architecture. In *NDSS*, 2016.

[78] Abraham Yaar, Adrian Perrig, and Dawn Song. Pi: A path identification mechanism to defend against ddos attacks. In *2003 Symposium on Security and Privacy, 2003.*, pages 93–107. IEEE, 2003.

[79] Minho Sung and Jun Xu. Ip traceback-based intelligent packet filtering: A novel technique for defending against internet ddos attacks. *IEEE Transactions on parallel and Distributed Systems*, 14(9):861–872, 2003.

[80] Zhuotao Liu, Hao Jin, Yih-Chun Hu, and Michael Bailey. Middlepolice: Toward enforcing destination-defined policies in the middle of the internet. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1268–1279, 2016.

[81] Vasileios Giotsas, Georgios Smaragdakis, Christoph Dietzel, Philipp Richter, Anja Feldmann, and Arthur Berger. Inferring BGP Blackholing Activity in the Internet. In *Proceedings of the 2017 Internet Measurement Conference (IMC)*, pages 1–14, 2017.

[82] Xin Zhang, Hsu-Chun Hsiao, Geoffrey Hasker, Haowen Chan, Adrian Perrig, and David G Andersen. Scion: Scalability, control, and isolation on next-generation networks. In *2011 IEEE Symposium on Security and Privacy*, pages 212–227. IEEE, 2011.

[83] Van Jacobson, Diana K Smetters, James D Thornton, Michael F Plass, Nicholas H Briggs, and Rebecca L Braynard. Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 1–12, 2009.

[84] Paolo Gasti, Gene Tsudik, Ersin Uzun, and Lixia Zhang. DoS and DDoS in Named Data Networking. In *Proceedings of 22nd International Conference on Computer Communications and Networks (ICCCN)*, July/August 2013.

[85] Dongsu Han, Ashok Anand, Fahad Dogar, Boyan Li, Hyeontaek Lim, Michel Machado, Arvind Mukundan, Wenfei Wu, Aditya Akella, David G Andersen, et al. XIA: Efficient Support for Evolvable Internetworking. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI'12)*, pages 309–322, 2012.

[86] Sajjad Arshad, Maghsoud Abbaspour, Mehdi Kharrazi, and Hooman Sanatkar. An anomaly-based botnet detection approach for identifying stealthy botnets. In *2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, pages 564–569. IEEE, 2011.

[87] Aleksandar Lazarevic, Levent Ertoz, Vipin Kumar, Aysel Ozgur, and Jaideep Srivastava. A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the 2003 SIAM international conference on data mining*, pages 25–36. SIAM, 2003.

[88] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials*, 16(1):303–336, 2013.

[89] Eric Liang, Hang Zhu, Xin Jin, and Ion Stoica. Neural packet classification. In *Proceedings of the ACM Special Interest Group on Data Communication*, pages 256–269. 2019.

[90] Rohan Doshi, Noah Apthorpe, and Nick Feamster. Machine learning ddos detection for consumer internet of things devices. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 29–35. IEEE, 2018.

[91] Sankardas Roy, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, Vivek Shandilya, and Qishi Wu. A survey of game theory as applied to network security. In *2010 43rd Hawaii International Conference on System Sciences*, pages 1–10. IEEE, 2010.

[92] Sivaramakrishnan Ramanathan, Jelena Mirkovic, Minlan Yu, and Ying Zhang. SENSS Against Volumetric DDoS Attacks. In *Proceedings of the 34th Annual Computer Security Applications Conference (ACSAC)*, pages 266–277, 2018.

[93] Benjamin Johnson, Aron Laszka, Jens Grossklags, Marie Vasek, and Tyler Moore. Game-theoretic analysis of ddos attacks against bitcoin mining pools. In *International Conference on Financial Cryptography and Data Security*, pages 72–86. Springer, 2014.

[94] Theodoros Spyridopoulos, G Karanikas, Theodore Tryfonas, and Georgios Oikonomou. A game theoretic defence framework against dos/ddos cyber attacks. *Computers & Security*, 38:39–50, 2013.

[95] Wikipedia. Karush-Kuhn-Tucker (KKT) conditions. https://en.wikipedia.org/wiki/Karush-Kuhn-Tucker_conditions, 2021.

A PROOF OF THEOREM 1

Proof. Note that

$$V_i(p_a, D) = \sum_k \mu_i p_i^k \log D_{1,i}(k) + \mu_a p_a^k \log(1 - D_{1,i}(k))$$

For any $(a, b) \in \mathbb{R}^2 \setminus \{0, 0\}$, the function $y \to a\log(y) + b\log(1 - y)$ achieves it maximum at $\frac{a}{a+b}$. So $D^*_{1,i}(k) = \frac{\mu_i p_i^k}{\mu_i p_i^k + \mu_a p_a^k}$, thus concluding the proof. □

B PROOF OF THEOREM 2

Proof. Under the optimal filters in Equation 3, the DDoS traffic that bypasses these filters to reach the victim is:

$$\mu_a p_a^k D^*_{1,n}(k) = \frac{\mu_a p_a^k \cdot \mu_n p_n^k}{\mu_a p_a^k + \mu_n p_n^k} \geq \frac{1}{2} \min\left(\mu_a p_a^k, \mu_n p_n^k\right), \forall k$$

and "=" holds if and only if $\mu_a p_a^k = \mu_n p_n^k, \forall k$ and yields the mixed-DDoS strategy $\mathbf{p_a} = \mathbf{p_n}$ (i.e., adversary fully mimics the legitimate traffic). Therefore, the total DDoS traffic that bypasses the optimal defenses to reach the victim is at least $\frac{1}{2}\sum_k \min\left(\mu_a p_a^k, \mu_n p_n^k\right)$. In this case, we have $\eta(\mathbf{p_n}, D^*) \leq \max_{\mathbf{p_a}} \eta(\mathbf{p_a}, D^*) = \eta(\mathbf{p_a^*}, \mathbf{D^*_{1,n}})$. In this strategy, as long as legitimate traffic $\mu_n > 0$, we have at least one k with $\mu_a p_a^k D^*_{1,n}(k) > 0$ and thus $\eta(\mathbf{p_n}, D^*) > 0$, which concludes our proof. □

C PROOF OF THEOREM 3

Proof. The symmetry of $d(\mathbf{p}, \mathbf{q})$ is evident by its definition. The concavity of $d(\mathbf{p}, \mathbf{q})$ is also evident since it is a sum of concave functions $f(x, y) = (\frac{1}{x} + \frac{1}{y})^{-1}$ for any $x, y > 0$. To prove $d(\mathbf{p}, \mathbf{q})$'s

bounds and identity, first note $d(\mathbf{p},\mathbf{q}) \geq 0$. Moreover, for any k

$$(\frac{1}{\mu_p p_k} + \frac{1}{\mu_q q_k})^{-1} = \frac{\mu_p p_k \cdot \mu_q q_k}{\mu_p p_k + \mu_q q_k} \leq \frac{1}{4} \cdot (\mu_p p_k + \mu_q q_k)$$

and "=" holds iff. $\mu_p p_k = \mu_q q_k$. Then we have

$$d(\mathbf{p},\mathbf{q}) = \sum_k \frac{\mu_p p_k \cdot \mu_q q_k}{\mu_p p_k + \mu_q q_k} \leq \sum_k \frac{1}{4} \cdot (\mu_p p_k + \mu_q q_k) = \frac{1}{4}(\mu_p + \mu_q)$$

and "=" holds iff. $\mu_p p_k = \mu_q q_k, \forall k$. Since $\sum_k p_k = \sum_k q_k = 1$, we conclude $d(\mathbf{p},\mathbf{q})$ achieves its maximum if and only if $\mu_p = \mu_q$ and $p_k = q_k, \forall k$.

For the monotonicity property, we note

$$\begin{aligned} d_n &= \sum_k \frac{\mu_n p_n^k \cdot d_{n-1} q_{n-1}^k}{\mu_n p_n^k + d_{n-1} q_{n-1}^k} \\ &= d_{n-1} \sum_k q_{n-1}^k \frac{\mu_n p_n^k}{\mu_n p_n^k + d_{n-1} q_{n-1}^k} \\ &\leq d_{n-1} \sum_k q_{n-1}^k = d_{n-1} \end{aligned}$$

where "=" holds iff. $\frac{\mu_n p_n^k}{\mu_n p_n^k + d_{n-1} q_{n-1}^k} = 1, \forall k$, i.e. $d_{n-1} = 0$. Similarly, we can prove $d_n \leq \mu_n$, thus $d_n \leq \min(d_{n-1}, \mu_n)$. By recursion, we conclude $d_n \leq \min_k \mu_k$. □

D PROOF OF THEOREM 4

Proof. We prove Theorem 4 by recursion. When $n = 1$, Theorem 4 naturally holds according to Theorem 1. Assume Theorem 4 holds for hop 1, 2, ..., $n-1$. Consider the n-th hop. Following Theorem 1 and the analysis in Section 5.1, its local optimal DDoS filter as

$$D_{1,n}^*(k) = \frac{\mu_n p_n^k}{\mu_n p_n^k + \mu_a p_a^k D_{n-1}^*(k)}$$

So the accumulative multi-hop optimal DDoS filter is

$$\begin{aligned} D_n^*(k) &= D_{n-1}^*(k) \cdot D_{1,n}^*(k) \\ &= \frac{\mu_n p_n^k D_{n-1}^*(k)}{\mu_n p_n^k + \mu_a p_a^k D_{n-1}^*(k)} \\ &= \frac{\mu_n p_n^k \cdot d_{n-1} q_{n-1}^k}{\mu_n p_n^k \cdot (\mu_a p_a^k + d_{n-1} q_{n-1}^k) + \mu_a p_a^k \cdot d_{n-1} q_{n-1}^k} \\ &= \frac{d_n q_n^k}{d_n q_n^k + \mu_a p_a^k} \end{aligned}$$

where

$$\begin{aligned} q_n^k &= \frac{1}{d_n} \left(\frac{1}{\mu_n p_n^k} + \frac{1}{d_{n-1} q_{n-1}^k} \right)^{-1} \\ d_n &= \sum_k \left(\frac{1}{\mu_n p_n^k} + \frac{1}{d_{n-1} q_{n-1}^k} \right)^{-1} = d(p_n, q_{n-1}) \end{aligned}$$

which retains the same form. So by recursion, we conclude Theorem 4 holds. □

E PROOF OF THEOREM 5

Proof. Equation 7 implies $D_n^*(k)$ decreases monotonically with d_n^k, and $\lim_{d_n^k \to 0} D_n^*(k) = 0, \forall k$. So the DDoS traffic $\mu_a p_a^k D_n^*(k)$ decreases with monotonically d_n^k and $\lim_{d_n \to 0} \mu_a p_a^k D_n^*(k) = 0, \forall k$. This implies $\lim_{d_n \to 0} U(\mathbf{p_a}, \mathbf{D_n^*}) = 0$ based on the threat model in §2. Since $\eta = U(\mathbf{p_a}, \mathbf{D_n^*})/\mu_a$ and the adversary's attack capacity μ_a is independent of all d_n^k, we conclude η also decreases with monotonically d_n^k and $\lim_{d_n \to 0} \eta(\mathbf{p_a}, \mathbf{D_n^*}) = 0$.

Moreover, Theorem 5 implies $d_{n-1} \geq d_n$ and "=" holds iff. $d_{n-1} = 0$, so $d_1 \geq d_2 \geq \ldots \geq d_n$ and $\lim_{n \to \infty} d_n = 0$ since $d_k \geq 0, \forall k$. This implies $\lim_{n \to \infty} U(\mathbf{p_a}, \mathbf{D_n^*}) = 0$ and thus $\lim_{n \to \infty} \eta(\mathbf{p_a}, \mathbf{D_n^*}) = 0$. □

F PROOF OF THEOREM 6

Proof. From Theorem 5, the adversary's optimal DDoS strategy satisfies $\lim_{d_n \to 0} \mu_a = 0$ and $\lim_{n \to \infty} \mu_a = 0$. Then Equation 3 implies optimal local filter $\lim_{d_n \to 0} D_{1,m}^*(k) = 1$ and $\lim_{d_n \to 0} D_{1,m}^*(k) = 1$. Each local filter's precision, recall and accuracy equal to $D_{1,m}^*(k)$ as proved in §3.1, thus concluding our proof. □

G PROOF OF LEMMA 1

Proof. Since $\eta(\mathbf{p_a}, \mathbf{D}) = \frac{U(\mathbf{p_a}, \mathbf{D})}{\mu_a}$ and attack capacity μ_a is given, the adversary's optimal strategy for $U(\mathbf{p_a}, \mathbf{D})$ also optimizes $\eta(\mathbf{p_a}, \mathbf{D})$. So we focus on $U(\mathbf{p_a}, \mathbf{D})$ in the following proof. When all hops will adopt EID's optimal filters, the accumulative optimal filter $D_n^*(k)$ is shown in Theorem 4. Therefore, the adversary's attack gain is

$$U(\mathbf{p_a}, \mathbf{D}) = \sum_k c_k \cdot \mu_a p_a^k D_n(k) = \sum_k c_k \frac{\mu_a p_a^k \cdot d_n q_n^k}{\mu_a p_a^k + d_n q_n^k}$$

So maximizing this $U(\mathbf{p_a}, \mathbf{D})$ implies to solve the following constraint non-linear optimization:

$$\begin{aligned} &\max_{p_a} U(\mathbf{p_a}, \mathbf{D}) \\ &\text{s.t.} \sum_k p_a^k = 1 \\ &0 \leq p_a^k \leq 1, \forall k \end{aligned}$$

This nonlinear optimization can be solved by applying the classical Karush-Kuhn-Tucker (KKT) conditions [95]. Consider the Lagrangian for $\max_{p_a} U(\mathbf{p_a}, \mathbf{D})$ with latents $\lambda, \mathbf{a}, \mathbf{b}$:

$$\begin{aligned} L(p_a, \lambda, \mathbf{a}, \mathbf{b}) = U(\mathbf{p_a}, \mathbf{D}) &- \lambda \left(\sum_k p_a^k - 1 \right) \\ &+ \sum_k a_k (p_a^k - 1) - \sum_k b_k p_a^k \end{aligned}$$

Then the optimal DDoS policy p_a satisfies

$$\frac{\partial L}{\partial p_a^k} = c_k \mu_a \left(\frac{\omega_n q_n^k}{\omega_n q_n^k + (1-\omega_n) p_a^k} \right)^2 + a_k - b_k - \lambda = 0, \forall k \quad (19)$$

$$a_k (p_a^k - 1) = 0, b_k p_a^k = 0, \forall k \quad (20)$$

$$\sum_k p_a^k = 1 \quad (21)$$

From Equation 20, there are 4 cases for each k:

- Case 1: $a_k \neq 0, b_k \neq 0$. This is impossible for any p_a^k in Equation 20.
- Case 2: $a_k = 0, b_k \neq 0$. Equation 20 implies $p_a^k = 0$.
- Case 3: $a_k \neq 0, b_k = 0$. Equation 20 implies $p_a^k = 1$, which immediately implies $p_a^j = 0, \forall j \neq k$. Then $U(q_a, D^*) = \frac{d_n q_n^k \cdot \mu_a c_k}{d_n q_n^k + \mu_a}$. To maximize $U(q_a, D^*)$, the adversary should choose k that maximizes $\frac{d_n q_n^k \cdot \mu_a c_k}{d_n q_n^k + \mu_a}$, which results in the **simple DDoS** policy in Theorem 1.
- Case 4: $a_k = 0, b_k = 0$. Then Equation 20 holds for any p_a^k settings. Meanwhile, Equation 19 implies

$$\left(\frac{\omega_n q_n^k}{\omega_n q_n^k + (1-\omega_n) p_a^k}\right)^2 = \frac{\lambda}{c_k \mu_a}, \forall k$$

Then $p_a^k = \frac{\omega_n q_n^k}{1-\omega_n}(\sqrt{c_k \mu_a / \lambda} - 1)$. Since $\sum_k p_a^k = 1$, we have

$$\sum_k p_a^k = \frac{\omega_n}{(1-\omega_n)} \sqrt{\frac{\mu_a}{\lambda}} \sum_k q_n^k \sqrt{c_k} - \frac{\omega_n}{(1-\omega_n)} = 1$$

So $\sqrt{\lambda} = \sqrt{\mu_a}\omega_n \sum_k q_n^k \sqrt{c_k} = \sqrt{\mu_a}\omega_n \mathbb{E}_{q_n}\sqrt{c}$ and therefore

$$p_a^k = \frac{\omega_n q_n^k}{1-\omega_n}\left(\frac{\sqrt{c_k}}{\omega_n \mathbb{E}_{q_n}\sqrt{c}} - 1\right)$$

which implies **mixed DDoS** attack in Theorem 1. Note that as a probability distribution, we mandate $0 \leq p_a^k \leq 1, \forall k$. This constraint implies $\omega_n \mathbb{E}_{q_n}\sqrt{c} \leq \sqrt{c_k} \leq \omega_n(1 + \frac{1-\omega_n}{\omega_n q_n^k})\mathbb{E}_{q_n}\sqrt{c}$ which results in the condition in Equation 12. Once this condition holds, we apply this policy to $U(q_a, D^*)$ and get

$$\begin{aligned}\max_{q_a} U(q_a, D^*) &= d_n \sum_k q_n^k c_k - d_n \omega_n \left(\sum_k q_n^k \sqrt{c_k}\right)^2 \\ &= d_n \mathbb{E}_{q_n}[c] - d_n \omega_n (\mathbb{E}_{q_n}[\sqrt{c}])^2\end{aligned}$$

thus concluding the proof. □

H PROOF OF THEOREM 7

Proof. We show that even the optimal reflective DDoS policy in Theorem 1 holds these bounds, thus demotivating *all* reflective DDoS attacks. For the simple DDoS policy in Theorem 1, we have

$$\max_{p_a} U(\mathbf{p_a^*}, \mathbf{D}^*) = \frac{d_n q_n^k \cdot \mu_a c_k}{d_n q_n^k + \mu_a} \leq \frac{d_n \mu_a}{d_n + \mu_a} c_k \leq d_n c_k$$

since $\max_{p_a} U(\mathbf{p_a^*}, \mathbf{D}^*)$ increases monotonically with $q_n^k \in [0,1]$ and $\frac{\mu_a}{d_n + \mu_a} \leq 1$. Note $\lim_{d_n \to 0} \max_{q_a} U(\mathbf{p_a^*}, \mathbf{D}^*) = 0$, which is consistent with Theorem 5.

For the mixed DDoS policy in Theorem 1, we have

$$\max_{p_a} U(\mathbf{p_a^*}, \mathbf{D}^*) = d_n \mathbb{E}_{q_n}[c] - d_n \omega_n (\mathbb{E}_{q_n}[\sqrt{c}])^2 \leq d_n \mathbb{E}_{q_n}[c]$$

Similar to simple DDoS attack, $\lim_{d_n \to 0} \max_{q_a} U(\mathbf{p_a^*}, \mathbf{D}^*) = 0$, which is consistent with Theorem 5. □