

# Towards Comparative Evaluation of DDoS Defences

Andikan Otung  
Department of Computer Science  
University of Oxford  
Oxford, United Kingdom  
andikan.otung@cs.ox.ac.uk

Andrew Martin  
Department of Computer Science  
University of Oxford  
Oxford, United Kingdom  
andrew.martin@cs.ox.ac.uk

**Abstract**—DDoS defence evaluation provides a way to capture the usefulness of defensive solutions to one of the most notorious Internet attacks of our computing generation. An alternative approach to evaluation offers a valuable mechanism by which different DDoS defences can be commensurably and objectively compared. Such a development would not only enable individual organizations to make better informed decisions on which defences to implement but could also aid collaborations to realize global solutions; and reveal insights into aspects requiring further investigation.

We present CED3 (pronounced “Seed”), a DDoS defence evaluation framework designed to facilitate the commensurable comparison between DDoS defences in a way that captures their strengths and weaknesses. Firstly, CED3 introduces the notion of true effectiveness, which addresses the problem, identified in the literature, of previously validated defences subsequently being shown to be ineffective when evaluated under different attack conditions. CED3 leverages a structured theoretical analysis process to drive empirical data acquisition in order to enhance consistency, transparency and, ultimately, longevity of evaluation conclusions. Lastly, CED3 introduces the concept of defence maps, which applies the idea of true effectiveness to “scopes” in order to communicate the strengths and weaknesses of defences in a way that allows them to be visually compared.

We demonstrate the CED3 framework by applying it to comparatively evaluate three DDoS defences. Using results obtained from extensive simulations in NS-3, we show how the strengths and weaknesses of different defences can be visually compared. We conclude by discussing the merits and limitations of CED3.

**Keywords**—DDoS, Network Security, Defence, Performance, Evaluation, Effectiveness, Comparison

## I. INTRODUCTION

Denial of service (DoS) attacks are attempts by attackers to prevent legitimate users from accessing connected services through the disconnection, corruption or malicious consumption of the resources upon which the victim service depends. Distributed denial of service attacks (DDoS) are a form of DoS attacks that are executed by multiple distributed agents. Since their emergence two decades ago [1], and despite a plethora of ensuing research, DDoS attacks have plagued the Internet, growing consistently in magnitude and prevalence, to become recognized by Internet Service Providers as the top operational threat to their customers [2], [3]; costing victims hundreds of thousands of dollars per attack [4], [5]. This threat is being compounded by growth in the Internet of Things (IoT) [6], [7].

Numerous publications are in agreement that a cooperative, distributed approach is necessary in overcoming

the DDoS threat [8]–[10]. However, due to interdependence between global Internet stakeholders, deployment of such approaches would require acceptance by, and collaboration between many organizations, including government agencies, corporations and Internet organizations.

Underpinning any such collaboration would be agreement on which, of the many possible defences, are deployed. The realization of any such consensus would be greatly aided by a method to systematically assess performance, impact and suitability of possible defensive solutions in a way that allows them to be compared by key decision makers.

Pervasive heterogeneous evaluation, where different methods are used to assess different defences [11]–[15] in the literature, makes it difficult for commensurable comparison between defences to be made. Moreover, the literature contains numerous examples of defences that are shown to be effective under favourable conditions but are either not tested or are shown to be ineffective under unfavourable conditions. An example is TrustGuard [16], which monitors the lack of packet-size variation as a metric to drop malicious flows but is not tested under conditions where an attacker varies the size of its packets.

Nevertheless, notable efforts have been made to improve and unify the testing and evaluation of DDoS defences. One such example is the paper *Benchmarks for DDoS Defense Evaluation* [17], in which the authors propose a benchmark-test-based method for evaluating DDoS defences and identify: the attack, the legitimate traffic and the network topology as key components to constructing suitable benchmark test scenarios. The authors conclude that the quality of user service experienced is the ultimate metric of effectiveness which varies based on the service in question. Their evaluation method tests a defence against popular attack traffic that is gathered by an automated tool, AProf. However, since the tool gathers the latest attack trends at the time of testing, the evaluated strength of a defence changes based on the time at which it is tested, despite the defence itself remaining unchanged.

The paper, *How to Test DoS Defences* [11] also provides valuable insights for improved DDoS defence evaluation. In it, six notable publications that evaluate DDoS defence architectures are selected and their employed evaluation methodologies scrutinized for shortcomings – which are found in all. The paper identifies seven evaluation goals that a DDoS defence evaluation methodology should consider, however, the usefulness of an evaluation method for comparing defences is neither identified as a possible goal of evaluation nor addressed. Similarly, the financial cost of the defence or the cost to an attacker in overcoming the defence are not considered.

This paper builds on the aforementioned prior work to propose CED3 (Comparative Evaluation of DDoS Defences – pronounced “Seed”), a DDoS defence evaluation framework designed to address the complexity and diversity in DDoS defence evaluation by evaluating DDoS defences in a way that allows them to be objectively and commensurably compared. CED3 introduces *true effectiveness*, a new metric of defence performance indicating the lowest-cost effective attack in a given scope of attacker strategies, to address the phenomenon identified in the literature of previously evaluated defences being subsequently shown to be less effective when tested under different attack conditions.

Beginning with an exploration of the fundamental concepts behind the CED3 framework, in section II, this paper proceeds to present the CED3 evaluation methodology and defence map in section III. The paper presents the application of CED3 to the comparative evaluation of three defences (sections IV & V) before concluding with a discussion on the merits and limitations of CED3 including potential avenues for future work (sections VI & VII).

## II. CED3 FOUNDATIONS

### A. Applicability

There are four kinds of defence against DDoS: *alleviation*, *enlargement*, *vulnerability reduction* and *attacker subversion*. *Alleviation* reduces the amount of resource consumption caused by the activities of malicious actors. This involves distinguishing between malicious and legitimate activity, by monitoring packet attributes, sender behavior or sender capabilities, and thwarting such activity through means such as dropping packets and denying requests. *Enlargement* increases the capacity of the victim to service requests. This involves efficiency increases in both hardware and software, and resource enlargement. Using faster disk drives or more powerful servers are examples of these. *Vulnerability reduction* limits exposure to DDoS attacks by minimizing attack-avenues. This involves, securing connected devices to prevent infection and assimilation into botnets, removing bugs in victim software that can be exploited during attack, geographically distributing servers to reduce single points of failure, and applying secure configurations and settings such as blocking unnecessary ports, and specifying hard limits on connection timeouts and the number and type of connections supported by a victim server. *Attacker subversion* involves the targeting of attackers and their assets to prevent or thwart attacks. If one considers attackers as ultimately human agents utilizing processes, systems and devices, the scope of possible attacker subversion is vast. Examples include, thwarting command and control centers used by botnets for communication, introducing deterrence measures and targeting botnets by removing bots from individual machines using antivirus software.

Since CED3 uses quality of service metrics to quantify the effectiveness of a defence during attack, as detailed in section II.C, all defence types, whose effect during attack can be empirically obtained, are able to be examined under the CED3 framework. However, the primary focus of CED3 is on the evaluation of alleviation defences for comparison. At the heart of these defences is the task of *asymmetric facilitation*, which entails observing the properties exhibited by communicating entities to distinguish the malicious from the benign, the normal from the abnormal and the acceptable

from the unacceptable, in order to support legitimate activity and thwart attacks. The features that a defence monitors in order to achieve asymmetric facilitation are known as the *metrics of distinction* of that defence.

Defences that prevent attacks from occurring, such as many attacker subversion strategies, lie outside the scope of CED3 – since their minimum costs of subversion is not easily quantified. Also out of scope are defences against: economic denial of sustainability attacks (EDoS) [18], attacks exploiting specific software or hardware implementation vulnerabilities, and attacks that compromise the integrity of the content sent by legitimate entities.

### B. Inevitable Theoretical Subversion

Alleviation DDoS defences harness observable and interrogatable characteristics of the attack landscape to make distinctions between malicious and legitimate activity. The obvious objective being the thwarting of malicious activity, whilst minimizing false positives and negatives. Prior work has leveraged machine learning for traffic classification [19], [20]. However, since these characteristics are digitally expressed in cyberspace, it is possible for an attacker to manipulate their exhibited characteristics to evade a defence by the emulation of legitimate actors and activity. However, it is important to note that some attributes of legitimate actors are more difficult or costly to emulate than others. For example, an attacker that reduces the individual data rate of its attacking bot agents in order to match that of the victim’s legitimate clients, would require more bots to compensate and maintain its overall attack strength.

We know that even non-malicious activity can cause denial of service (DoS) for users, such flash crowds [21]. This is due to the finite capacity of victim devices to service requests because of limited processing power, memory and bandwidth at both the victim server and the access devices to which it is connected. In such circumstances, any defence relying solely on asymmetric facilitation would prove ineffective as the scale of service demanded by the legitimate actors would mean that even perfect distinction (with no false positives or false negatives) in the defensive measure, would still be unable to prevent DoS to legitimate actors. This leads to the postulate of Inevitable Theoretical Subversion:

*Any DDoS defence technique is theoretically surmountable by an attacker.*

This postulate extends the paradigm of defence effectiveness evaluation from assessment of how well a defence would work, to consideration of what it would take to overcome it. This extension of consideration in evaluation is important as it solves a problem identified in the literature, where a change in attack pattern would prove a defensive technique, that had initially been considered effective, to be ineffective. Therefore, evaluation longevity can be realized by capturing, as part of the evaluation process, the ways in which a defence can be surmounted and the costs to do so, respectively. Therefore:

*any truly representative measure of the effectiveness of a DDoS defence must incorporate the difficulty and or cost to the attacker in overcoming the defence.*

This consideration helps avoid an unstable theoretical measure of effectiveness that is dependent on particular attack patterns or the capability of the attacker. The concept of associating the strength of a security mechanism with what it

would take to subvert it has precedent in information security[22]. Passwords, for example, facilitate privileged access, yet some can be guessed more easily than others.

### C. True Effectiveness

Effectiveness describes how well something achieves its purpose and, in DDoS defence literature, has been portrayed in different ways [11]. In order to clearly define effectiveness concerning DDoS defences, a clear understanding of the purpose of DDoS defence is needed, since effectiveness describes how well a defence achieves its purpose.

The purpose of a DDoS defence is to alleviate the negative effects of a DDoS attack as well as possible, whilst minimizing its own contribution to negative effects. Therefore, the effectiveness of DDoS defences shall describe how well defences alleviate the negative effects of DDoS attacks, where "negative" is defined from the perspective of attack victims and legitimate network users.

In addition to a clear definition of effectiveness, suitable metrics are needed that should sufficiently capture the negative effects of DDoS attacks in a way that reflects the quality of the experience of legitimate users (UEQ). Characteristics of the amount and timeliness of data transfer between server and client can capture the perceived UEQ, which can vary depending on the Internet service between client and server. The metrics: packet loss, delay, jitter and transaction time, capture the connectivity characteristics of legitimate clients and changes in these characteristics may be symptomatic of both network and application layer DDoS attacks. The significance of these metrics is such that Denial of Service cannot be achieved without their degradation. The exceptions to this include integrity compromises in data transmitted, which are not inherent symptoms of flood DDoS attacks and are outside the scope of this evaluation methodology.

From the postulate of inevitable theoretical subversion, derived in section II.B, the true effectiveness of a defence must consider the cost to an attacker to surmount the defence. We define true effectiveness as the lowest cost required for an attacker to surmount a defence within a given *scope* of attacker strategies. The consignment of true effectiveness to a scope of the attack landscape is essential, since niche defences that target particular kinds of attacks would otherwise be considered generally ineffective, despite being particularly effective for a given scope. The way in which true effectiveness incorporates costs for a given scope is described in sections III.C and II.D.

### D. Representative Evaluation

Complexity in comparative DDoS defence evaluation springs from the combinatorial explosion of possible variations in the three core input variables: the attack, the network and the legitimate user activity. The previous section posited principles to assist navigation of the domain of the first core input variable, the attack. This section addresses navigation of the core input variables of network and legitimate user activity. The principle underpinning this navigation is that of representative evaluation. Representative evaluation stipulates that the evaluation context, method and assumptions should be chosen such that evaluation outputs represent the reality. However, since it is often extremely difficult to replicate, or even know, the anticipated deployment environment (ADE), simplifications must be made. For fair evaluations the testing environment must be

able to sufficiently accommodate the features of each defence. For example, an environment for testing a defence that operates at the borders of autonomous systems (ASs), must contain borders between ASs. In addition, for commensurable comparison defences must be evaluated under the same context, which would then be a caveat of any comparative evaluation conclusions.

The papers [11] and [15] discuss DDoS defence testing by theory, simulation, emulation and deployment. From the CED3 standpoint any method is suitable provided it sufficiently captures the mechanics of the ADE. However, the "new dimension" of evaluation put forward in [15], comprising the use of real datasets, is unsuitable for CED3 since it doesn't allow for the realization of true effectiveness due to the use of a single historical static attacker strategy.

## III. CED3 FRAMEWORK

### A. Overview

The primary goal of CED3 is to evaluate DDoS defences in a way that captures their usefulness, is representative of actual deployment, and enables their granular quantitative comparison. Additional goals of CED3 are: to provide a means of visually presenting the performance of a defence in the context of the attack-landscape and other defences, to promote objective consistency of evaluations such that independent evaluations of identical defences yield similar results. CED3 addresses these through its output and processes, which employ both theoretical and empirical means to facilitate robust and meaningful conclusions.

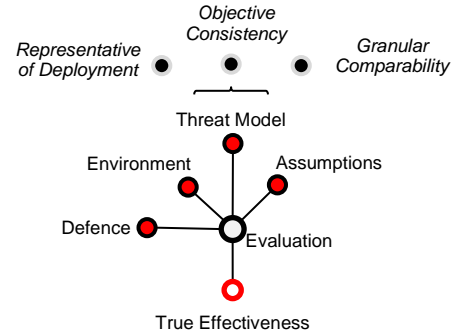


Fig. 1: CED3 Evaluation Overview

Figure 1 illustrates an overview of the CED3 framework. *Environment* refers to the network conditions under which the evaluation will take place and includes aspects such as the topology and legitimate traffic. The *threat model* defines the attacker's capabilities and limitations. *Assumptions* refer to any condition that is considered to be true for the purpose of the evaluation. These inputs (Environment, threat model and assumptions) all form the *evaluation context*, upon which the evaluation is based. The more closely the evaluation context matches the reality of the *anticipated deployment environment* (ADE), the more representative of deployment the evaluation conclusions are likely to be.

The evaluation process begins with a structured theoretical analysis (STA) of each defence to be evaluated. The findings of this allow a suitable test environment to be designed, that can accommodate all the defences to be evaluated. The STA also enables the identification of the weaknesses of each defence to design more potent attacks for each defence. This phase is structured to promote objective consistency of evaluations.

Following the STA, a suitable test environment that can sufficiently support all the defences is designed. Suitable attack scenarios are designed and an experiment plan to identify the lowest-cost effective attack is formulated for each defence. The defences and test environment are then implemented and the experiment plan is executed. Data output from the experiments are then processed and analyzed. If the data is sufficient, as determined from its analysis, then the true effectiveness of each defence is output. However, if analysis indicates that more data is needed then a new experiment plan is made to acquire more data. As an example, if there is a sudden change in UEQ between adjacent experiments, it would be informative to acquire more data points between the differing parameter values. This may be necessary in order to more precisely determine the lowest-cost attack.

### B. Structured Theoretical Analysis

The goal of the *structured theoretical analysis* (STA) is to acquire insights into the defences for evaluation, in order to generate an appropriate test environment and attack plan. Firstly, the defences are reviewed in order to understand test-network requirements and to identify metrics of distinction relevant to each defence, which can be leveraged for more effective testing. Metrics of distinction are parameters that a defence uses to distinguish between malicious and benign activity. Next, a *recursive threat analysis* (RTA) is carried out in order to identify defence weaknesses.

RTA, illustrated in Figure 2, is a structured way we created of discovering the vulnerabilities of DDoS defences. Recursion itself is a widespread algorithmic tool that has been applied to solve many problems, including the task of risk assessment [23]. However, we are not aware of its prior use in this manner for evaluating DDoS defences. RTA begins at a high level and successive iterations end up considering the low-level features of a defence. This enables RTA to be flexible enough to accommodate a broad range of defence types, whilst being able to meaningfully consider security aspects in detail.

RTA begins by identifying the goals of a DDoS defence. These goals are then broken down into the fundamental *mechanisms*, processes or functionality required of the architecture to achieve them. The *security requirements* of these components are then identified and the possible ways in which an attacker may compromise said security requirements are analyzed. These are the *threats*. If a security requirement is identified that the architecture cannot prevent an attacker from compromising, this is noted as a *vulnerability* of the architecture; and is an output of RTA process. However, if further mechanisms address that would-be vulnerability (threat), these mechanisms are input into the next iteration of recursive analysis and the process repeats. The recursive threads of RTA end when no further architectural mechanisms for mitigating discovered vulnerabilities exist and all that remain are a set of vulnerabilities and or assumptions.

In consideration of how an attacker could compromise architectural tenets, a threat model must be applied that defines the attacker's capabilities. It is important, for evaluation consistency, that this same threat model is used for each successive layer of analysis. If the identified threats are addressed by architectural mechanisms of the defence, then these defence mechanisms that address the identified threats are fed back into the RTA process and the cycle of analysis repeats. Further to the threat model, the RTA process is also affected by architectural deployment assumptions, such as

participation rates in various schemes, which may affect the effectiveness of attacks executed by an attacker.

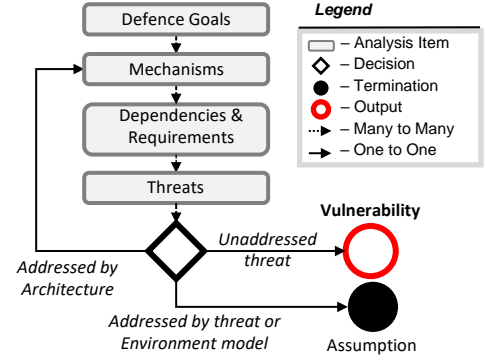


Fig. 2: Recursive threat analysis flow chart

Since it is possible for the vulnerabilities discovered from the RTA process to limit the performance of the defence in question, we consider them in the design of low-cost attack scenarios as part of the discovery of the true effectiveness of the defence. This systematic approach enables a more thorough evaluation that is as replete as the depth of the considerations of each successive level of analysis. Documenting this approach also aids communication and scrutiny, as third parties may gauge for themselves the robustness of the process by assessing the completeness of each recursive step.

### C. Empirical Methods

The effectiveness of a defence and its collateral damage are empirically obtained by measuring the difference between the user-experienced quality of service (UEQ) of the different test scenarios illustrated in Figure 3 (a). This *completeness diagram* depicts four test scenarios constructed from two independent variables: 1) whether an attack is occurring or not and 2) whether the defence-under-test is deployed or not.

Measuring the change in UEQ from during an attack *without* the defence, to during the same attack *with* the defence deployed, provides an indication of the *nominal effectiveness* of the defence. Thus, nominal effectiveness indicates the impact of a defence under a given attack scenario.

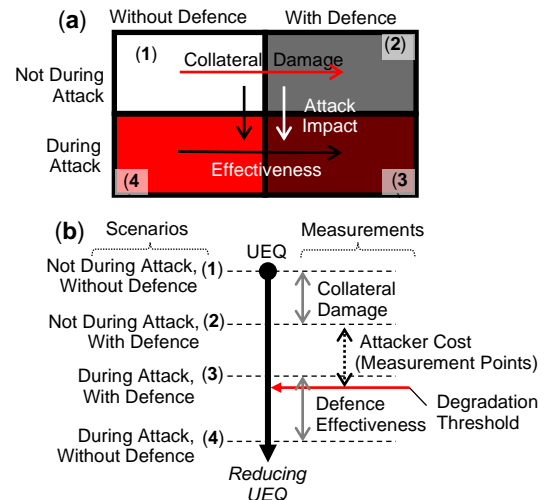


Fig. 3: (a) A completeness diagram of attack and peacetime scenarios for testing (b) Output measurements and scenarios depicted along a UEQ axis

Since nominal effectiveness is derived in this way, by measuring the alleviating impact provided by a defence in the presence of an attack (*ceteris paribus*), the nominal effectiveness of a defence is attack-specific. However, since numerous variations of attacks are possible, some method is needed in order to organize the measurement of effectiveness for a defence in a way that meaningfully characterizes the effectiveness of a defence. CED3 achieves this by obtaining the true effectiveness of defences over a given attack scopes and plotting these values over the areas of a *defence map* (presented in section III.D).

For reasons described in section II.C, degradation in UEQ is captured by measuring changes in delay, jitter, packet loss and transaction time. For a given test scenario, indicated by the  $n$ 'th quadrant in Figure 3 (a), the UEQ of users of the victim server,  $\phi_V^n$ , is given by:

$$\phi_V^n = \frac{K_1}{a_L \bar{\mu}_L^n + a_D \bar{\mu}_D^n + a_J \bar{\mu}_J^n + a_T \bar{\mu}_T^n + K_2} \quad (1)$$

The multiplicand  $\bar{\mu}_x^n$  represents the mean of the measured metric  $x$ , in the  $n$ 'th quadrant, where  $x$  is either L, D, J or T; denoting: loss, delay, jitter or transaction time. The multipliers of the form  $a_x$  are coefficients to balance the contributions of each metric to the UEQ. The dividend  $K_1$  and the summand  $K_2$  are constants to restrict the general magnitude of the UEQ variable and prevent negative values.

The four test scenarios are illustrated in Figure 3 (b) along a UEQ axis, which decreases in the downwards direction. The red horizontal arrow along the vertical UEQ axis, marks the degradation threshold, which is a UEQ value, beyond which, user service is considered to be suitably disrupted for an attack to be deemed successful. The true effectiveness then, is the minimum cost necessary for an attacker, within a given scope<sup>1</sup> of attacker strategies, to degrade the UEQ to below the degradation threshold. This cost is measured from scenario 3, where multiple attacks would be explored to discover the lowest-cost attack that takes the UEQ below the degradation threshold.

For the same scope, we can characterize the *normalized true effectiveness* (NTE) of a defence by dividing the true effectiveness by the lowest attacker cost required to degrade the measured UEQ beyond the degradation threshold when the defence is not deployed by that where there is an attack and the defence is *not* deployed – quadrants 3 ÷ 4, respectively. These costs are that required to acquire, maintain and operate a botnet large enough to generate the attack strength required; and are discussed in [24]. The NTE metric helps highlight the proportional change in resilience offered by the defence, where values greater than one indicate an improvement, values equal to one denote no change, and values less than one indicates that the defence actually makes the victim more vulnerable to denial of service.

Consideration of collateral damage is a crucial part of evaluating the performance of a defence [25]. In CED3, the collateral damage of a defence consists of two parts: that caused by a defence when there is no attack, and that inflicted on servers that are not targeted, during an attack on a victim.

The former, indicated in Figure 3 (a), is obtained from the difference between the UEQ measurements of quadrants 1 and 2. The latter is measured between quadrants 3 and 4 and captures the damage a defence may inflict on legitimate entities that are not under attack, in response to an attack. The peacetime collateral damage  $X_P$  is given by:

$$X_P = \phi_P^1 - \phi_P^2$$

whereas the in-attack collateral damage  $X_A$  is given by:

$$X_A = \phi_A^1 - \phi_A^3$$

If the collateral damage degrades the UEQ of clients of any destination server beyond the degradation threshold, then the attacker cost at which this occurs becomes a candidate for the lowest-cost attack – or true effectiveness of that defence, since denial of service would have been caused.

#### D. Defence Map

The idea of a DDoS defence map provides a way to present the effectiveness of a defence in a common context, highlighting the landscape of possible attacks that said defence can and, crucially, cannot mitigate against. The CED3 defence map achieves this by mapping defence performance to scopes of attacker strategies to allow the distinction between scopes of differing effectiveness within a user-friendly 2D plane.

Legend		Attack Strategy			
		Direct		Indirect	
R	– Reflection	U-R	R	U-R	R
U-R	– Un Reflected				
Attacking Protocol	IPv4	ICMP			
		UDP			
		TCP			
	IPv6	ICMPv6			
		UDP			
		TCP			

Fig. 4: An example CED3 defence map

The defence map is a grid of areas, with each grid area denoting a remit of possible attacks that derive their common properties from the column and row headers shown in black and white rectangles. These headers are independent variables and should be selected according to the defences under evaluation, to highlight differences in their scopes of performance. In the example map on Figure 4, the row labels (to the left of the grid) indicate the type of protocol used in the attack, whereas the column headers at the top represent the attack strategy employed by the attacker. “Direct” in the column header refers to attacks that directly target a victim, whereas “indirect” attacks target the resources upon which a victim depends, such as link-flooding attacks [26], [27]. Reflection attacks are those where the attacking agents recruit unwitting accomplices to attack the target, whereas in un-reflected attacks, the attacking agents attack the target without mobilizing assistance from unwitting accomplices.

The rectangular scope shaded in red consists of attacks using IPv4 UDP packets that directly target the victim in a direct

<sup>1</sup> The scope of an attacker strategy refers to the set of techniques an attacker may use to achieve DoS, which may have explicit exclusions. This is described in section III.D.

non-reflection attack-strategy. The value entered in that scope is the normalized true effectiveness ( $\eta$ ). The greater this value, the greater the true effectiveness of the defence within the given scope.

It is possible for the scope of each major grid area to be further divided, where each sub-scope would represent a distinct remit of possible attacks. This may make sense to highlight nuances in the defence performance. However, in general, no scope or sub-scope should restrict the attacker's capability such that the attacker, within the defined scope, cannot surmount the defence in question. For example, the total attack data rate must not be bounded, the maximum botnet size must not be curtailed, and the per-bot-data-rate should only be restricted by network and device limitations. The idea of a DDoS defence map provides a way to present the effectiveness of a defence in a common context, highlighting the landscape of possible attacks that said defence can and, crucially, cannot mitigate against.

#### IV. CED3 APPLICATION

##### A. Defences

We applied the CED3 framework to comparatively evaluate three types of defences found in the literature: Passport [28], TrustGuard [16], and Increasing the Victim Capacity. These defences were chosen because they each apply different mitigation mechanisms, satisfy all defence phases from detection to mitigation, and could each be implemented in reasonable timeframes.

Increasing the victim capacity, refers to increasing the resources of the victim network thus allowing it to service more clients. We have evaluated an increase by a factor of 2.

Passport is a proposed network scheme that eradicates inter-AS spoofing by identifying spoofed packets through failed verification of message authentication codes (MAC) appended to the packet headers. Identified spoofed packets are then demoted and discarded by intermediate and end routers respectively.

TrustGuard employs both macro and micro level analyses that trigger a probabilistic drop filter to remove attack traffic. Once the *macro-level anomaly indicator* is triggered for a particular destination address, *micro-level anomaly indicators* are used to direct the probabilistic drop filter to drop packets. The equation for the macro-level anomaly indicator is:

$$H(A) = -\sqrt{i_{max}} \cdot \sum_{i=1}^{12} p(x_i) \log_2 p(x_i) \quad (2)$$

Where  $x_i$  refers to the  $i$ 'th packet-size level ( $1 \leq i \leq 12$ ),  $p(x_i)$  refers to the probability that a packet has a size that places it in said packet-size level; and  $i_{max}$  is the packet-size level in which most packets fall for a given destination address. Further explanation of TrustGuard, including equations for the micro-level credit accumulation and the probabilistic drop filter can be found in [16].

A structured theoretical analysis (STA) of these three defences was carried out firstly to determine a suitable topology that accommodates all the defences and, suitable attack scenarios that capture the strengths and weaknesses of each defence.

##### B. Environment

###### 1) Platform & Network

Simulations were coded in C++ and executed using the NS3 simulation software [29]. Simulation was chosen (over theoretical means, emulation or using real systems) due to the lower costs required, experimental repeatability and greater levels of control over the test environment. NS3 was chosen as the simulation software due to its scalability and the benefits offered from it being open-source, such as documentation availability and extensibility to meet experimentation needs.

Figure 5 (a) illustrates the test-environment topology produced which consisted of client-side networks, server-side networks and a mesh core connecting the two. This design accommodated the deployment requirements of three defences: doubling of the victim capacity occurred at the server-side links, TrustGuard was deployed at the client side and Passport was deployed in the core. Full deployment of each defence was implemented in their respective tests.

The test environment topology consisted of approximately 500 nodes with 1 Gbps core links, 100 Mbps server access links and 10 Mbps client access links. Following preliminary baseline simulations, we determined 100 Mbps server bandwidths provided an acceptable balance between server bandwidth realism and overall simulation time. As shown in Figure 5 (b), the pattern of UEQ degradation remained the same as the server bandwidth varied over 3 orders of magnitude.

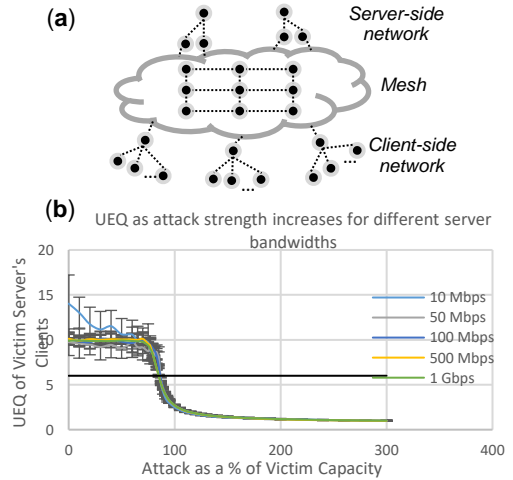


Fig. 5: (a) Test environment topology (b) Preliminary baseline simulations with no defence deployed

However, the simulation time for each scenario with 1 Gbps server access links was up to 4 days for a virtual attack duration of 5 minutes, whereas with 100 Mbps server access links, the simulation time was less than 4 hours. Hence the decision to proceed with 100 Mbps server links since each scenario would need to be repeated (in this case, 5 times) to produce a standard deviation (as plotted).

Since victim and network implementation weaknesses, and software attacks lie outside of scope, it was assumed that each end server node could process all of its received data.

###### 2) Legitimate Traffic

The legitimate traffic is an important part of the evaluation context and, as with topology, should ideally match the

anticipated deployment environment (ADE) as closely as possible. For this evaluation, we derived statistical attributes such as the mean packet size (~600 octets) and data rate (0.3 Mbps) of the legitimate traffic from the literature [30]–[32], since there was no particular ADE. These were randomly generated and normally distributed within and between simulations. Legitimate traffic consisted of both UDP and TCP traffic of equal split. The legitimate clients were distributed across the test environment nodes and supplied data to servers at 15% utilization of the server bandwidth. Echo applications running on each server reflected the received payloads back to the sending clients. A transaction in this case was defined by the receipt of a certain number of packets at the receiving server. The specified number of packets varied between clients but was chosen based on the sending rate of the client to enable the completion of said transaction in 30s. In order to generalize the measurement of transaction times, the transactions were staggered from the start of the simulation.

### C. Threat Model & Attack Scenarios

The threat model employed was an attacker in control of a distributed botnet of attacking agents capable of launching coordinated UDP flood attacks. The attacker was able to increase its attack strength and manipulate its attack characteristics in ways that could undermine the defensive mechanisms of each defence in order to achieve denial of service as cheaply as possible. The attacking agents were implemented as custom NS3 applications running on distributed nodes and the attack strength was increased by increasing the number of attacking agents, which were randomly distributed across the test network. Out of scope of the attacker capability was the execution of intra-AS-domain spoofing.

From the STA carried out on the three defences, two metrics of distinction were identified: packet size, for TrustGuard and packet validity for Passport. Thus, to characterize the performance of these defences, a matrix of attack scenarios were simulated for each. For Passport, the axes of variation were attack strength and packet validity, and for TrustGuard, attack strength and packet size. Each scenario was repeated several times with differing attack and legitimate clients that upheld the same chosen characteristics (of collective bandwidth, mean packet-size and mean sender rate) thus enabling a standard deviation to be obtained for each matrix element. In this way, over 10,000 attack-scenario simulations were executed on the ARC cluster [33], the results of which are presented in the next section.

## V. RESULTS

The results are illustrated in Figure 6, where Figure 6 (a), (b) and (c) are graphs showing how the user-experienced quality of service (UEQ) varied under differing attacks for the three defences. Each plot point represents the average UEQ measured from a set of repeated simulations of a particular attack scenario (which represents one element of the simulation matrix discussed previously). The values of the parameters  $K_1$ ,  $K_2$ ,  $a_L$ ,  $a_D$ ,  $a_I$  and  $a_T$  applied to equation (2) were 10, 1, 5, 2, 1 and 5, respectively. The error bars indicate the standard deviation of the UEQ in the repeated simulations. Figure 6 (d) illustrates a defence map depicting the normalized true effectiveness (NTE) of each defence over an attack-landscape, where the brighter the shade of red, the

higher the value of the NTE depicted and the more effective the defence over the specified scope.

Figure 6 (a) illustrates two curves of UEQ as the attack strength increases. The first curve is the UEQ when the defence of doubling the victim capacity is deployed. Marked on this curve is the point at which the measured UEQ crosses the UEQ threshold, indicating the attack was successful. This point is the lowest strength of attack required to subvert the defence. It directly correlates to the true effectiveness ( $\epsilon$ ) of the defence, in that it is the cost of producing this attack that is the true effectiveness of the defence. We measure this cost in *attack units* ( $\alpha$ ), where one attacking unit corresponds to the financial cost in inflicting an attack of total data rate equal to 1% of the original victim capacity, which in this case is 1% of 100 Mbps – 1 Mbps.

The second curve illustrates, for comparison purposes, the cost of inflicting denial of service (DoS) on the victim server when no defence is deployed. We observe that the cost of effecting DoS increases from 87  $\alpha$ , when no defence is deployed to 190  $\alpha$ , when doubling the victim capacity is deployed. No other curves exist on this graph because, the structured theoretical analysis (STA) identified no metrics of distinction for this defence that could reduce the cost of inflicting denial of service.

Figure 6 (b) illustrates the variation of UEQ as the attack strength increased when Passport was deployed. The five curves plotted represent five mixes of spoofed and non-spoofed attack traffic. We see that when the attack consists of 100% spoofed traffic, Passport is extremely effective, able to withstand attacks greater than three times the victim bandwidth with no degradation of UEQ.

However, as the portion of non-spoofed packets increases, the effectiveness of the defence reduces until DoS is able to be achieved at just 87  $\alpha$ , which is the same cost as that required to cause DoS to the victim server when no defence is deployed. The true effectiveness of Passport is marked on the graph as 87  $\alpha$ , despite the defence being more effective in the presence of greater proportions of spoofed attack traffic. This is because we consider the cost difference, to an attacker, between using spoofed traffic rather than non-spoofed traffic to be negligible.

Figure 6 (c) illustrates the curve when TrustGuard is deployed. We observe that when the mean attack payload size is low, the UEQ drops to a level of approximately 4, and remains steady at that level, despite increasing attack strengths. However, as the mean attack payload size increases, the initial drop in UEQ lessens and increasing attack strengths begin to have an effect on the UEQ. The reason for this phenomenon is that when the mean payload size of the attack packets is small, it activates TrustGuard's probabilistic drop filters by causing the macro-level anomaly detection indicator to cross the activation threshold. The probabilistic drop filter then drops packets from all flows towards the victim, with flows of greater average packet size experiencing less probability of being dropped. As the mean payload size of the attack packets increased, the magnitude of the initial UEQ drop reduced because fewer TrustGuard filters were activated by the macro-level anomaly indicator.

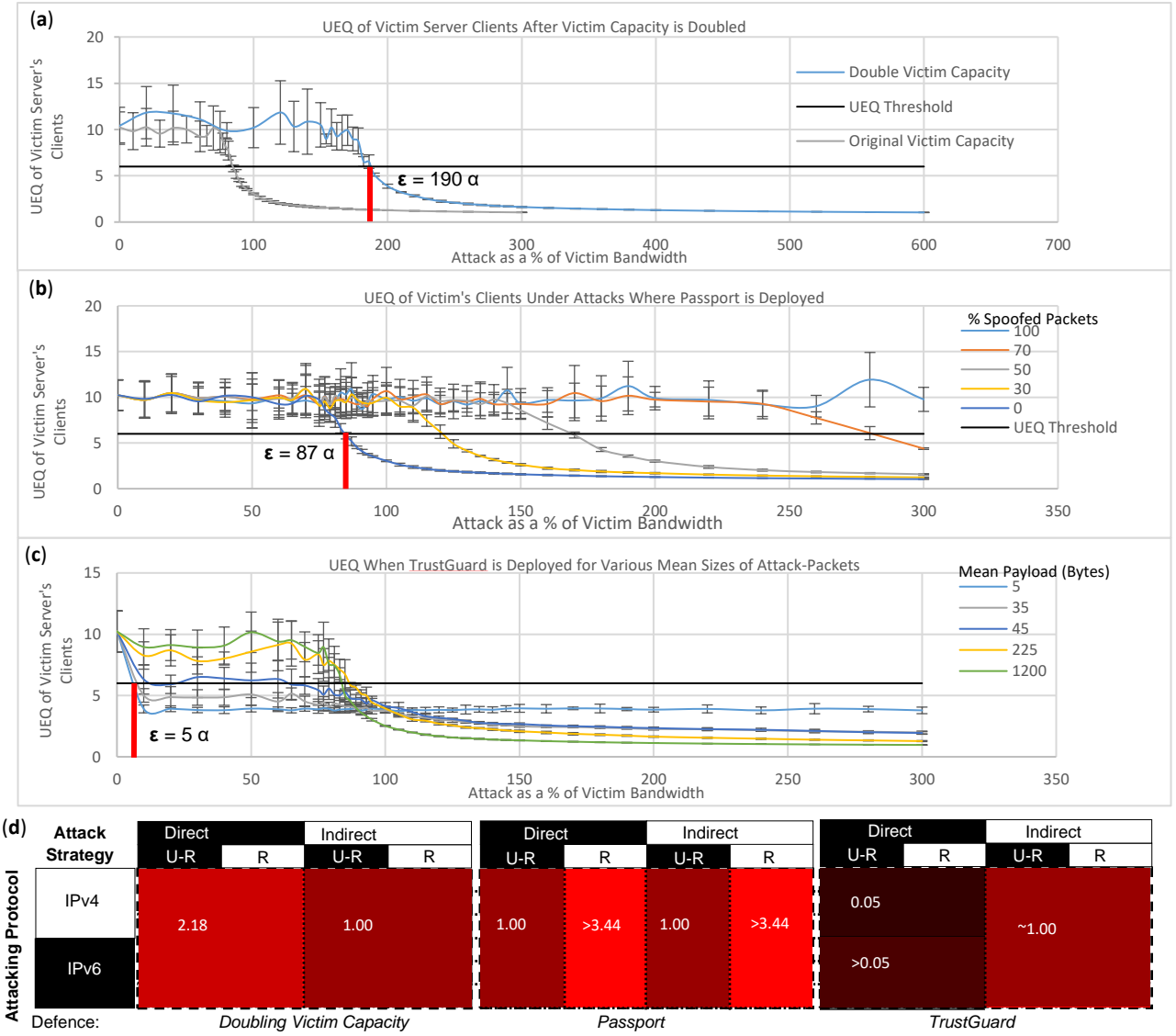


Fig. 6: Graphs of the simulated output showing the UEQ varied under attack for three defences: doubling the victim capacity (a), Passport (b) and TrustGuard (c); and a Defence Map illustrating the normalized true effectiveness of each defence over an attack landscape (d) where the greater the number the greater the true effectiveness within that scope

TrustGuard was the only defence to produce significant levels of collateral damage to the clients of the victim server. However, none of the defences, including TrustGuard produced significant collateral damage to the UEQ of clients communicating with servers-nodes that were *not* under attack.

Figure 6 (d) illustrates the defence map. The areas of the map correspond to the scope of attacks employed by the attacker and the brighter the shade of red of the area, the higher the true effectiveness of the defence within that scope. The values of normalized true effectiveness shown within the defence map were inferred from the effectiveness values obtained by simulation coupled with an understanding of the defences gained from the STA process. Because the constituent values are normalized, inherent amplification effects of reflection attacks and the inherent ineffectiveness of indirect attacks are nullified, allowing the relative differences between the values to come from the defence performance itself.

The true effectiveness of Passport is independent of the attack protocol used and whether the attack is direct or

indirect. However, as shown, since Passport mitigates against spoofed packets, it is most effective against reflection attacks - since reflection attacks require the use of spoofed packets.

For direct un-reflected attacks, TrustGuard has low attack cost regardless of protocol, since the attacker may choose to use low packets in its attack which would activate the TrustGuard filter. We hypothesise that this is also the case for direct reflection attacks assuming that spoofed packets can be crafted to generate reflection packets of low packet size, hence the areas share the same shade. For indirect un-reflected attacks, the attack packets are sent to destinations other than the victim but with the attempt to overload a certain link on the path used by the victim's clients. Therefore, if the attacker were to use small packets in this scenario, the TrustGuard probabilistic drop filters may activate causing both legitimate and attack packets to be dropped. However, since the filtering occurs based on the destination address of the attacking packets, no packets of the clients of the intended victim server would be dropped as a result, since the attack is indirect, yet attack packets would still be being dropped. In this case we can see it would be more effective for an attacker to avoid activating the

probabilistic drop filters by using larger attack packets. Hence for indirect attacks, a NTA of 1 that matches the true effectiveness when no defence is deployed is shown, since the most effective indirect attack is to avoid activating the TrustGuard queues. Since TrustGuard is sensitive to packet-size, we estimate that its true effectiveness for direct attacks will be higher since IPv6 have larger packets, thus requiring a greater flow of smaller packets (greater attack strength) to activate the TrustGuard queues. The greater size of IPv6 packets does not affect indirect attacks on TrustGuard since the lowest-cost attack strategy is to avoid activating the TrustGuard queues, as previously discussed.

## VI. DISCUSSION

The results demonstrate how CED3 can be practically applied to evaluate and compare the *true effectiveness* of multiple defences. Also demonstrated is how the performance of different types of DDoS defences can be compared in a way that highlights both their strengths and weaknesses. This is enabled by the concept of scopes, depicted in the defence maps of Figure 6 (d). These scopes highlight differences in defence effectiveness to different scopes of attack, which could help a practitioner select complementary defences. Because CED3 actively seeks effective lowest-cost attacks, and so is independent of particular attacker capabilities, the comparative conclusions of CED3 evaluations are expected to have longevity.

The process, however, is not without its limitations. Despite the many strengths offered by simulation, including environment control and repeatability, there is always the question of whether a simulation is sufficiently representative of an actual deployment environment. One aspect that deviated from reality was that only the logical functionality of each defence was modeled so that possible transit delays introduced by each defence were not considered. However, we believe the aspects modelled using the NS3 software (including link delays, routing, queueing etc.) were sufficient to approximate the response of each defence to increasing UDP flood attacks and consequently demonstrate the usefulness of true effectiveness as a sustainable evaluation output.

For complex defences, the number of experiments could increase – potentially quadratically – for each additional metric of distinction of the defence. However, this increase can be mitigated by employing more efficient techniques to find the lowest cost attack; such as algorithms that find global minima of multivariate functions, where applicable.

The value of the degradation threshold of the user-experienced quality of service (UEQ), used to determine whether denial of service was achieved, was chosen arbitrarily at a level corresponding to packet loss of approximately 7%. The veracity of this threshold could be augmented by deriving it from experiments where participants are subjected to different degradation levels and are asked to note the point at which they deem the service they are using to be untenable. Likewise, the equation by which the UEQ is calculated, could also be modified so that its degradation matches said empirical data. This is an open research direction. However, we believe that the equation used and the threshold chosen were sufficient to enable the determination of whether an attack succeeded, as part of the process of demonstrating how to find the lowest-cost attack.

The financial cost of executing an attack is expressed in *attacking units*, which are units linked to the total data rate of the attack. We consider this to be appropriate in determining

differences between attack costs, under the assumption that generating attacks with greater total bandwidth generally costs more since more bots would generally be required, which have a financial cost to acquire, operate and maintain [24]. Future work could involve the conversion of attack units into actual dollar values, which would enable more evaluation metrics such as *cost-effectiveness* or *value for money* to be obtained.

**Extending the Framework.** Adoptability is an important consideration for practical deployment decisions since a performant defence may not be suitable for actual deployment for various reasons, such as cost. We define *adoptability* as an attribute of a DDoS defence that describes how conducive a defence is to its own adoption in a given environment. Thus, adoptability may be a helpful addition to the CED3 framework, as an output. In recognizing the distinctions between *implementation*, *deployment* and *adoption*, specified by Eardley et al. [34] we consider it possible for a defence to be implemented and deployed but not used; and therefore not having the intended impact. Therefore, part of adoptability is the consideration of the factors that encourage the use of a defence, such as the incremental incentives a defence provides to key stakeholders. Other relevant considerations of adoptability include feasibility, deployment & operational costs, adverse effects and externalities.

Adoptability-related considerations would enhance the recursive threat analysis (RTA) process by feeding into each level as adoptability goals or security requirements to enable the discovery of pertinent threats that may compromise the adoptability of the defence. For example, privacy requirements to identify whether a defences enables the tracking of user activity.

The challenge of extending the CED3 framework with adoptability outputs that enables granular, objective, commensurable comparison between defences, remains an open research direction.

## VII. CONCLUSION

This paper proposes CED3, a novel framework for evaluating DDoS defences that enables them to be commensurably compared. CED3 introduces a paradigm shift in DDoS defence evaluation from ascertaining “how well” a defence would work to discovering “what it would take” to overcome it. This is achieved through the notion of the true effectiveness of a defence, which promotes evaluation longevity, thus addressing the issue identified in the literature of defences being validated under favourable conditions, only later to be shown to be ineffective under different conditions.

We apply our framework to evaluate three different defences. The results demonstrate how, using defence maps, CED3 is able to highlight defence strengths and weaknesses in a way that allows them to be visually compared.

## ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [EP/P00881X/1] and made use of the University of Oxford Advanced Research Computing facility [<http://dx.doi.org/10.5281/zenodo.22558>].

## REFERENCES

- [1] S. v. Raghavan and E. (Edward) Dawson, An investigation into the detection and mitigation of denial of service (DoS) attacks : critical information infrastructure protection. Springer India Pvt. Ltd, 2011.
- [2] A. Otung and A. Martin, "Distributed Defence of Service (DiDoS): A Network-layer Reputation-based DDoS Mitigation Architecture," 2020.
- [3] "Akamai Blog | 2021: Volumetric DDoS Attacks Rising Fast." <https://www.akamai.com/blog/security/2021-volumetric-ddos-attacks-rising-fast> (accessed Aug. 14, 2022).
- [4] Ponemon Institute, "The Cost of Denial-of-Services Attacks - Sponsored by Akamai Technologies," no. March, p. 10, 2015, [Online]. Available: <https://www.akamai.com/us/en/multimedia/documents/content/the-cost-of-denial-of-services-attacks.pdf>
- [5] D. Kobialka, "Kaspersky Lab Study: Average Cost of Enterprise DDoS Attack Totals \$2M - MSSP Alert," MSSP Alert, 2018. <https://www.msspalert.com/cybersecurity-research/kaspersky-lab-study-average-cost-of-enterprise-ddos-attack-totals-2m/> (accessed Sep. 02, 2018).
- [6] M. Shafiq, Z. Tian, A. Kashif Bashir, X. Du, and M. Guizani, "CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques," IEEE Internet Things J, vol. 8, no. 5, 2021, doi: 10.1109/IJOT.2020.3002255.
- [7] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "IoT malicious traffic identification using wrapper-based feature selection mechanisms," Comput Secur, vol. 94, Jul. 2020, doi: 10.1016/J.COSE.2020.101863.
- [8] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS Problems," ACM Comput Surv, vol. 39, no. 1, pp. 3-es, 2007, doi: 10.1145/1216370.1216373.
- [9] K. Malialis and D. Kudenko, "Distributed response to network intrusions using multiagent reinforcement learning," Eng Appl Artif Intell, vol. 41, pp. 270–284, 2015, doi: 10.1016/j.engappai.2015.01.013.
- [10] T. Peng, L. C., and K. Ramamohanarao, "Detecting Distributed Denial of Service Attacks by Sharing Distributed Beliefs," in ACISP: Information Security and Privacy, 2003, no. June, pp. 214–225. doi: [https://doi.org/10.1007/3-540-45067-X\\_19](https://doi.org/10.1007/3-540-45067-X_19).
- [11] J. Mirkovic, S. Fahmy, P. Reiher, and R. K. Thomas, "How to Test DoS Defenses," in 2009 Cybersecurity Applications & Technology Conference for Homeland Security, Mar. 2009, pp. 103–117. doi: 10.1109/CATCH.2009.23.
- [12] J. Mirkovic and P. Reiher, "D-WARD: A source-end defense against flooding denial-of-service attacks," IEEE Trans Dependable Secure Comput, vol. 2, no. 3, pp. 216–232, 2005, doi: 10.1109/TDSC.2005.35.
- [13] K. Kalkan and F. Alagöz, "A distributed filtering mechanism against DDoS attacks: ScoreForCore," Computer Networks, vol. 108, pp. 199–209, Oct. 2016, doi: 10.1016/j.comnet.2016.08.023.
- [14] H. Beitollahi and G. Deconinck, "Analyzing well-known countermeasures against distributed denial of service attacks," 2012, doi: 10.1016/j.comcom.2012.04.008.
- [15] S. Behal and K. Kumar, "Trends in Validation of DDoS Research," in Procedia - Procedia Computer Science, 2016, vol. 85, pp. 7–15. doi: 10.1016/j.procs.2016.05.170.
- [16] H. Liu, Y. Sun, V. C. Valgenti, and M. S. Kim, "TrustGuard: A flow-level reputation-based DDoS defense system," 2011 IEEE Consumer Communications and Networking Conference, CCNC'2011, no. PerNets, pp. 287–291, 2011, doi: 10.1109/CCNC.2011.5766474.
- [17] P. R. Jelena Mirkovic, Erinc Arikan, Songjie Wei, Roshan Thomas, Sonia Fahmy, "Benchmarks for DDoS Defense Evaluation," in Military Communications Conference MILCOM 2006, 2006, no. 2, pp. 1–10. [Online]. Available: <http://ieeexplore.ieee.org/document/4086729/>
- [18] M. Naresh Kumar, P. Sujatha, V. Kalva, R. Nagori, A. K. Katukojwala, and M. Kumar, "Mitigating Economic Denial of Sustainability (EDoS) in Cloud Computing Using In-cloud Scrubber Service," in 2012 Fourth International Conference on Computational Intelligence and Communication Networks, Nov. 2012, pp. 535–539. doi: 10.1109/CICN.2012.149.
- [19] M. Shafiq, Z. Tian, Y. Sun, X. Du, and M. Guizani, "Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city," Future Generation Computer Systems, vol. 107, pp. 433–442, Jun. 2020, doi: 10.1016/J.FUTURE.2020.02.017.
- [20] M. Shafiq, Z. Tian, A. K. Bashir, A. Jolfaei, and X. Yu, "Data mining and machine learning methods for sustainable smart cities traffic classification: A survey," Sustain Cities Soc, vol. 60, p. 102177, Sep. 2020, doi: 10.1016/J.SCS.2020.102177.
- [21] S. Bhatia, "Detecting Distributed Denial-of-Service Attacks and Flash Events," The LNM Institute of Information Technology, 2013. [Online]. Available: [https://eprints.qut.edu.au/62031/1/Sajal\\_Bhatia\\_Thesis.pdf](https://eprints.qut.edu.au/62031/1/Sajal_Bhatia_Thesis.pdf)
- [22] E. Stinson and J. C. Mitchell, "Towards systematic evaluation of the evadability of bot/botnet detection methods," WOOT'08 Proceedings of the 2nd conference on USENIX Workshop on offensive technologies, 2008.
- [23] W. M. D. Chia, S. L. Keoh, A. L. Michala, and C. Goh, Real-time Recursive Risk Assessment Framework for Autonomous Vehicle Operations. 2021. Accessed: Mar. 29, 2022. [Online]. Available: <http://eprints.gla.ac.uk/234649/http://eprints.gla.ac.uk>
- [24] C. G. J. Putman, A. Abhishta, and L. J. M. Nieuwenhuis, "Business Model of a Botnet," Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018, pp. 441–445, 2018, doi: 10.1109/PDP2018.2018.00077.
- [25] K. Singh, S. Dhindsa, B. Bhushan, K. Singh Dhindsa, and G. N. Khalsa, "PERFORMANCE ANALYSIS OF AGENT BASED DISTRIBUTED DEFENSE MECHANISMS AGAINST DDOS ATTACKS," 2018. Accessed: Aug. 31, 2018. [Online]. Available: [www.computingonline.net](http://www.computingonline.net)
- [26] L. Wang, Q. Li, Y. Jiang, X. Jia, and J. Wu, "Woodpecker: Detecting and mitigating link-flooding attacks via SDN," Computer Networks, 2018, doi: 10.1016/j.comnet.2018.09.021.
- [27] M. Rezazad, M. R. Brust, M. Akbari, P. Bouvry, and N. M. Cheung, "Detecting target-area link-flooding DDoS attacks using traffic analysis and supervised learning," Advances in Intelligent Systems and Computing, vol. 887, pp. 180–202, 2019, doi: 10.1007/978-3-030-03405-4\_12.
- [28] X. Liu, A. Li, and X. Yang, "Passport : Secure and Adoptable Source Authentication," Nsdi, 2008.
- [29] The NS-3 Research Community, "NS-3 Introduction — Tutorial," 2018. <https://www.nsnam.org/docs/tutorial/html/introduction.html#about-ns3> (accessed Jul. 25, 2018).
- [30] P. Velan, J. Medková, T. Jirsík, and P. Čeleda, "Network traffic characterisation using flow-based statistics," in Proceedings of the NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium, Jun. 2016, pp. 907–912. doi: 10.1109/NOMS.2016.7502924.
- [31] J. Barr, "Cloud Computing, Server Utilization, & the Environment," AWS News Blog, 2015. <https://aws.amazon.com/blogs/aws/cloud-computing-server-utilization-the-environment/#:~:text=Servers and related IT resources,levels are often under 20%25> (accessed Mar. 22, 2022).
- [32] P. Jurkiewicz, G. Rzym, and P. Boryło, "Flow length and size distributions in campus Internet traffic," Comput Commun, vol. 167, pp. 15–30, Sep. 2018, doi: 10.1016/j.comcom.2020.12.016.
- [33] A. Richards, "University of Oxford Advanced Research Computing," Aug. 2015, doi: 10.5281/ZENODO.22558.
- [34] P. Eardley, M. Kanakakis, A. Kostopoulos, T. Levä, K. Richardson, and H. Warma, "Deployment and Adoption of Future Internet Protocols," in The Future Internet, 2011, pp. 133–144.