# Estimates of success rates of Denial-of-Service attacks

Teodor Sommestad, Hannes Holm, Mathias Ekstedt
Industrial information & control systems
Royal Institute of Technology (KTH)
Stockholm, Sweden
{teodors, hannesh, mathiase}@ics.kth.se

*Abstract*— **Denial-of-service (DoS) attacks are an imminent and real threat to many enterprises. Decision makers in these enterprises need be able to assess the risk associated with such attacks and to make decisions regarding measures to put in place to increase the security posture of their systems. Experiments, simulations and analytical research have produced data related to DoS attacks. However, these results have been produced for different environments and are difficult to interpret, compare, and aggregate for the purpose of decision making. This paper aims to summarize knowledge available in the field by synthesizing the judgment of 23 domain experts using an establishing method for expert judgment analysis. Different system architecture's vulnerability to DoS attacks are assessed together with the impact of a number of countermeasures against DoS attacks.**

***Keywords – denial of service; DoS; distributed denial of service; flooding attack; semantic attack; expert judgment; Cooke's classical method***

## I. INTRODUCTION

Denial-of-service (DoS) attacks on information technology based services are a relatively common type of security incident and produce a substantial share of the losses incurred from attacks on information technology.

To manage the risk related to DoS attacks in practice, decision makers need to be able to understand and estimate the probability that their information technology based services can be disturbed by this type of attack. Hence, data on the probability of attack success given different conditions in the information technology infrastructure would contribute to more informed decision making when it comes to risks associated with DoS attacks.

There are literature that summarizes this problem domain and the potential of different countermeasures, for example, the review made by Peng et al. [1]. In this review, four categories of defense against DoS attacks are identified: attack prevention, attack detection, attack source identification, and attack reaction. All of these are relevant, however, this study only focus on the first type of defense – attack prevention.

There is plenty of research on techniques for attack prevention in terms of simulations, experiments, and analytical calculations. However, this research is difficult to use in a decision making situation. The simulations, experiments, and calculations are made for a specific configuration and aims to be representative for a specific context [2]. Therefore, unless the decision maker has this specific situation at hand, these results must first be interpreted and somehow synthesized before they can be used to answer questions related to the decision making situation at hand.

This paper aims to summarize knowledge that exists in the research community on how difficult it is to succeed with DoS attacks in general, and how effective different preventive countermeasures are against these attacks. This is done through a survey distributed to experts on DoS attacks. The experts were asked to estimate success probabilities in different scenarios. Since the scenarios were defined on a high level of abstraction, the answers from any expert would be inherently uncertain. In order to take this fact into account the answers were given as probability distributions of attack success. In order to arrive at as credible results as possible estimates of the experts were weighted using an established method for expert judgment analysis. Thus, in summary, the estimates are made for a number of selected system scenarios and show both expected effectiveness of countermeasures and the uncertainty of these estimates.

The rest of this paper is structured as follows. Section II presents related work and the scenarios for which success probabilities were assessed. Section III presents the method for expert judgment analysis, known as Cooke's classical method. Section IV presents the data collection method. Section V shows the result. Section VI discusses these results and their implications. Section VII draws conclusions.

## II. STUDIED DENIAL-OF-SERVICE ATTACK SCENARIOS

Denial-of-service (DoS) attacks can be divided in two types [3]. The first type, *semantic attacks*, causes DoS by sending carefully crafted packets to the targeted system (also known as software exploits [4]). These packets exploit vulnerabilities in the target system and make it unresponsive, e.g. by crashing the system. The second type, *brute force attacks*, occupies the target service with massive amounts of traffic that impairs it so that it cannot serve legitimate users (also known as flooding attacks [4]). This study covers both these classes of attacks. Previous work in both types of attack is presented below together with the variables included in this study. The selected variables have been chosen based on (1) relevance to practical applications and their usage in practice today, (2) their expected impact in the possibility to succeed with the attack and (3) their relevance to decision makers of software based services. Relevant

variables have been selected based on a literature review. This selected variables relevance and the prioritizations made were validated by two external security professionals.

### A. Semantic attacks

Software vulnerabilities are common in software products and many of these can be used to influence the availability of the vulnerable system. More than two thirds of known software vulnerabilities have an impact on availability [5], i.e., they can be used to cause DoS. There are several aspects that influence if an attacker can exploit the software vulnerability. The Common Vulnerability Scoring System [6] includes: the access vector that is possible to use (i.e. remotely exploitable or only locally exploitable), if the attacker must be able to bypass authentication before exploitation, and the ease of exploitation (e.g. if it is easy to construct the exploit code).

The most obvious countermeasure for this type of attack is to remove the software vulnerability, e.g., by updating the software to a version without the vulnerability. However, this is not always possible to do and is in the typical case associated with an effort and cost. Also, exploitation might be possible even if the software is without what would be regarded as software vulnerabilities per se. For example, by exploiting the intended functionality in an abusive way as when recursive payloads are sent to a web service [7].

There are also measures that influence the exposure that is experienced when software exhibits a software vulnerability. There are a number preventive measures for semantic attacks, for instance, in [8] a toolkit for defensive programming is presented. However, preventive measures these are seldom used in practice.

This study only investigates remote attacks. Hence, the investigated attack vector is remote exploitation. The model used to assess DoS attacks success rate includes three variables for semantic attacks (cf. TABLE I). These are: (1) if the attacker can provide access credentials to the targeted system (*AC, Access Credentials)*, (2) the presence of a software vulnerability (SV, *Software Vulnerability*) in the target, and (3) the target of the DoS attack. For (3), the goal is to cause DoS for an entire machine or the target is to cause DoS on a specific service.

**TABLE I. VARIABLES STUDIED FOR SEMANTIC DOS ATTACKS**

| Variable | Description |
|---|---|
| AC | Access credentials: if the attacker can authenticate itself as a legitimate user of the service. |
| SV | Software vulnerability: if the software has an implementation vulnerability. |
| Machine | If the DoS attack targets a machine (e.g. a CPU), or a specific service running on the machine. |

### B. Flooding attacks

A substantial amount of research has been spent on brute force attacks, in particular distributed DoS attacks. Excellent compilations of attack form within this category of attacks can be found in [3], [9], [1]. The taxonomy in [1] focus on preventive measures on the network level, e.g., ingress and egress filtering at internet service providers. While this certainly has an influence on the possibility to perform certain attacks, it is difficult to influence as a decision maker of software based services at their enterprise. In the taxonomy of [3] preventive measures against flooding attacks include: system security (e.g. to reduce botnets on the internet), protocol design, resource accounting, and resource multiplication. The first two of these are again difficult to influence as an enterprise decision maker and; the third can be seen as a reactive measure [1]. In addition to the abovementioned measures, the taxonomy given in [9] includes: changing IP address, honeypots, disabling unused services, and secure overlay services.

Based on the criteria given above the following variables were selected for this study: changing IP address through proactive server roaming [10], [11] and resource multiplication (i.e. redundancy) with load balancing [3].

**TABLE II. VARIABLES STUDIED FOR BRUTE FORCE DOS ATTACKS.**

| Variable | Description |
|---|---|
| Roaming | The service uses proactive server roaming. |
| Load balancing | There is a load balancer in between the attacker and the target. |

### C. Assumptions

In addition to the variables given above a number of conditions were kept constant in the scenarios. The attacker is an outsider with the competence of a professional penetration tester who has access to tools that are free or commercially available. The attacker has spent one week preparing for the attack and the attack is performed from an external network. Also, in the case of brute force attacks it should be assumed that there is an enterprise firewall between the attackers host(s) and the targeted service. However, in all cases the attacker can reach the targets IP address and port.

Even with these assumptions the scenario definitions only covers a subset of the variables of relevance. They are also given on a coarse detail level. For instance, the details associated with the software vulnerability are not specified and the amount of redundancy implemented behind the load balancer. To avoid unnecessary ambiguity the respondents were asked to consider unspecified variables to be in the state they typically are in an enterprise environment. For instance, if enterprises often are protected by ingress and egress filtering this should be accounted for and considered in the estimates given. Any uncertainty caused by this should be reflected in the estimates.

## III. SYNTHESIZING EXPERT JUDGMENTS

There is much research on how to combine, or synthesize, the judgment of multiple experts to increase the calibration of the estimate used. Research has shown that group of individuals assess an uncertain quantity better than the average expert, but the best individuals in the group are often better calibrated than the group as a whole [12]. The combination scheme used in this research is the classical model of Cooke [13]. Experience shows that Cooke's classical method outperforms both the best expert and the "equal weight" combination estimates. In an evaluation involving 45 studies it performs significantly better than

both in 27 studies and performs equally as well as the best expert in 15 of them [14].

In Cooke's classical method *calibration* and *information* scores are calculated for the experts based on their answers on a set of seed questions, i.e,. questions for which the true answer is known at the time of analysis. The calibration score shows how correct the respondent's answers match the true value; the information score shows how precise the respondent's answer are. These two scores are used to define a *decision maker* which assigns weights to the experts based on their performance. The weights defined by this decision maker are used to weight the respondents answer's to the questions of interest – in this case the operational scenarios described in section II. In sections III.A, III.B and in III.C Cooke's classical method is explained. For a more detailed explanation the reader is referred to [13].

### A. Calibration score

In the elicitation phase the experts provide individual answers to the seed questions. The seed questions request the respondents to specify a probability distribution for an uncertain continuous variable. This distribution is typically specified by stating its 5th, 50th, and 95th percentile values. This yields four intervals over the percentiles *[0-5, 5-50, 50-95, 95-100]* with probabilities of $p$= *[0.05, 0.45, 0.45, 0.05]*. As the seeds are realizations of these variables the well calibrated expert will have approximately 5% of the realizations in the first interval, 45 % of the realizations in the second interval, 45 % of the realizations in the third interval and 5% of the realizations in the fourth interval. If $s$ is the distribution of the seed over the intervals the relative information of $s$ with respect to $p$ is:

$$I(s, p) = \sum_{i=1}^{4} \ln (s_i/p_i) \quad (1)$$

This value indicates how surprised someone would be if one believed that the distribution was p and then learnt that it was $s$.

If N is the number of samples/seeds the statistic of $2NI(s, p)$ is asymptotically Chi-square distributed with three degrees of freedom. This is asymptotic behavior is used to calculate the calibration *Cal* of expert $e$ as:

$$Cal(e) = 1 - \chi^2_3(2N\, I(s, p)) \quad (2)$$

Calibration measures the statistical likelihood of a hypothesis. The hypothesis tested is that realizations of the seeds ($s$) are sampled independently from distributions agreeing with the expert's assessments ($p$).

### B. Information score

The second score used to weight experts is the information score, i.e., how precise and informative the expert's distributions are. This score is calculated as the deviation of the expert's distribution to some meaningful background measure. In this study the background measure is a uniform distribution over the interval zero to one.

If $b_i$ is the background density for seed $i \in \{1,\ldots,N\}$ and $d_{e,i}$ is the density of expert $e$ on seed $i$ the information score for expert $e$ is calculated as:

$$inf(e) = \frac{1}{N} \sum_{i=1}^{N} I(d_{e,i}, b_i) \quad (3)$$

In other words, the information score is the relative information of the expert's distribution with respect to the background measure. It should be noted that the information score does not reflect calibration and does not depend on the realization of the seed questions. So, regardless of what the correct answer is to a seed question a respondent will receive a low information score for an answer that is similar to the background measure, i.e., the answer is distributed evenly over the variable's range. Conversely, an answer that is more certain and focused the probability density over few values will yield high information scores.

### C. Constructing a decision maker

The classical method rewards experts who produce answers with high calibration (high statistical likelihood) and high information value (low entropy). A strictly proper scoring rule is used to calculate the weights the decision maker should use. If the calibration score of the expert $e$ is at least as high as a threshold value the expert's weight is obtained as:

$$w(e) = Cal(e) * Inf(e) \quad (4)$$

if the expert's calibration score is less than the threshold value α. If the experts calibration is less than α, the expert's weight is set to zero, a situation which is common in practical applications.

The threshold value α corresponds to the significance level for rejection of the hypothesis that the expert is well calibrated. The value of α is identified by resolving the value that would optimize a virtual decision maker. This virtual decision maker combines the experts' answers (probability distributions) based on the weights obtained at the chosen threshold value (α). The optimal level for α is where this virtual expert would receive the highest possible weight if it was added to the expert pool and had its calibration and information scored as the actual experts.

When α has been resolved the normalized value of the experts weights *w(e)* are used to combine their estimates of the uncertain quantities of interest.

## IV. Data Collection Method

This section presents how the survey data was collected by explaining: how seed questions for Cooke's classical method were assessed; which population and sample of experts that was chosen; how the measurement instrument was developed and tested.

### A. Seed questions

As the experts performance on answering the seed questions are used to weight them, it is critical that the seeds are highly validated and also that they lie in the same domain as the studied variables. Thus, the seeds should represent the truth and it should be difficult to tell them apart from the questions of the study. They need to be drawn from the respondents' domain of expertise, but need not necessarily be directly related to questions of the study [13].

Naturally, the robustness of the weights attributed to individual experts depends on the number of seeds used. Experience shows that eleven seed questions are more than enough to see substantial difference in calibration [13]. This study used eleven seed questions to weight the respondents.

These eleven seed questions were of two types. The first type asked the respondents to estimate characteristics of known vulnerabilities related to DoS attacks. The correct answer was drawn from US Department of Commerce National Vulnerability Database [5]. The second type of question related to actual distributed DoS attacks of activity and how it influenced enterprises. The data for these questions came from the survey result presented in [15]. Summaries of the actual questions are presented in TABLE III.

**TABLE III. SEED QUESTIONS.**

| # | Question | Value (%) |
|---|---|---|
| 1 | What is the share of known vulnerabilities with some impact on availability? | 71 |
| 2 | Of the known vulnerabilities with some impact on availability, how large portion can be exploited from external networks? | 85 |
| 3 | Of the known vulnerabilities with some impact on availability, how large portion requires that the attacker can bypass authentication? | 5 |
| 4 | What is the share of known vulnerabilities with some impact on availability that affect Windows 7? | 85 |
| 5 | What is the share of known vulnerabilities with complete impact on availability? | 23 |
| 6 | What portion of organizations in EMEA and US that operate their business online has an important online reputation use some on-premise/in-house DDoS protection technology? | 65 |
| 7 | What portion of organizations in EMEA and US that operate their business online or have an important online reputation over provision their bandwidth to protect against potential DDoS threats? | 28 |
| 8 | What portion of organizations in EMEA and US that operate their business online, have an important online reputation or operate financial services are primarily suffering from target DDoS attacks and aware of whom the attackers are? | 30 |
| 9 | What portion of organizations in EMEA and US that operate their business online or have an important online reputation or operate online financial services is primarily suffering from random DDoS? | 52 |
| 10 | What portion of organizations in EMEA and US that operate their business online or have an important online have experienced a DDoS attacks during a year that did disrupt services? | 31 |
| 11 | What portion of organizations in EMEA and US that operate their business online, has an important online have experienced and has experienced DDoS attacks needed more than 5 hours to recover from the most severe attack? | 41 |

### *B. The domain experts*

Studies of expert's calibration have concluded that experts are well calibrated in situations where with learnability and with ecological validity [16]. Learnability comes with models over the domain, the possibility to express judgment in a coherent quantifiable manner that could be verified, and the opportunity to learn to from historic predictions and outcomes. Ecological validity is present if the expert is used to making judgments of the type they are asked for.

This study asks questions on the success of attempted DoS attacks, given different conditions. These judgments can be expressed in a quantifiable coherent and quantifiable manner. Persons with experience in DoS attacks (directly or indirectly) will also have access historic outcomes to learn from. Good candidates for this are researchers and penetration testers in the security field. These can be expected to both reason in terms of success or failure of an attacks in different condition. They also make such judgments in their line of work and evaluate different options (i.e., there is ecological validity). DoS attack researchers were therefore chosen as the population to survey.

To identify suitable security researchers articles published in the SCOPUS [17], INSPEC or Compendex [18] databases between January 2005 and September 2010 were reviewed. Authors who had written articles in the information technology field with any of the words "denial of service attack" or "denial-of-service attack" in the title, abstract, or keywords were identified. If their contact information could be found they were added to the list of potential respondents, resulting in a sample of 1378 respondents. After reviewing and screening respondents and their contact information a sample of 1065 individuals was assessed. Of these the used contact information to at approximately 180 turned out to be incorrect or outdated.

Out of approximately 885 researchers invited to the survey 296 opened the survey and 65 submitted answers to questions in the survey. A response rate of this magnitude is reasonable to expect from a slightly more advanced survey as this. Consistency checks and completeness checks were used to ensure the quality of answers used in the analysis. After these controls 23 respondents' answers remained and these 23 were used in the final analysis.

As recommended by [19], motivators were presented to the respondents invited to the survey: i) helping the research community as whole, ii) the possibility to win a gift certificate on literature, and iii) being able to compare their answers to other experts after the survey was completed.

### *C. Elicitation instrument*

A web survey was used to collect the probability distributions from the invited respondents. The survey was structured into four parts, each beginning with a short introduction to the section. First, the respondents were given an introduction to the survey as such that explained the purpose of the survey and its outline. In this introduction they also confirmed that they were the person who had been invited and provided information about themselves, e.g., years of experience in the field of research. Second, the respondents received training regarding the answering format used in the survey. After confirming that this format was understood the respondents proceeded to its third part. In the third part both the seed questions and the questions of the study were presented to the respondents. Finally, the respondents were asked to provide qualitative feedback on the survey and the variables covered by it.

Questions in section three were each described through a scenario entailing a number of conditions. Scenarios and

conditions for the seed questions can be found in Table III; scenarios and conditions for the questions of interest in this study is described in section V.

In the seed questions and the questions on semantic attacks the respondent was asked to provide a probability distribution that expressed the respondent's belief. As is custom in applications of Cooke's classical method this probability distribution was specified by setting the 5th percentile, the 50th percentile (the median), and the 95th percentile for the probability distribution. In the survey the respondents specified their distribution by adjusting sliders or entering values to draw a dynamically updated graph over their probability distribution. The three points specified by the respondents defines four intervals over the range [0, 100]. The graphs displayed the probability density as a histogram, instantly updated upon change of the input values.

In the questions concerning brute force attacks, the respondent also specified a probability distribution through the 5th, 50th and 95th percentile. However, they now specified the number of hosts the attacker would need to control to make 5, 50 or 95 percent of the legitimate requests being dropped. As before the estimates dynamically updated a graph representing the answer.

Use of graphical formats is known to improve the accuracy of elicitation [20]. Figures and colors were also used to complement the textual formulations and make the content easier to understand. In Figure 1 the format presented to respondents is exemplified.

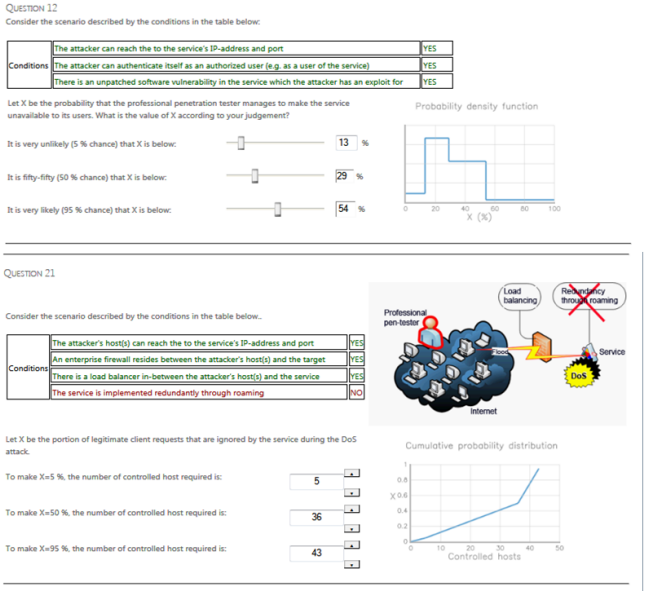


Figure 1. Example of questions and answering formats used in the survey.

Elicitation of probability distributions is associated with a number of issues [20]. Effort was therefore spent on ensuring that the measurement instrument held sufficient quality. Before distribution of the survey the used question format as such had been tested in a pilot study on other security parameters. In that pilot study a randomized sample of 500 respondents was invited; 34 of these completed the pilot during the week it was open. The questions in this pilot survey were presented in the same way as in the present survey. A reliability test using Cronbach's alpha [21], [22] was carried out using four different ways to phrase questions for one variable. Results from this test showed a reliability value of 0.817, which indicates good internal consistency of the instrument.

## V. RESULTS

This section presents the result of the analysis performed on the judgment of the 23 experts. In section V.A the overall performance of the respondents on the seed questions is presented. In section V.B the synthesized estimates of those respondents who were assigned weight are presented.

### A. *Respondents' performance*

As in many other studies involving expert judgment many of the experts were poorly calibrated on the seed questions. Their calibration score varied between $3.853*10^{-11}$ and 0.3697 with a mean of 0.0375; their information score varied between 0.222 and 1.974 with a mean of 1.00.

Cooke's classical method aims is to identify those respondents whose judgment is well calibrated and informative. The virtual decision maker was optimized at a significance level (α) of 0.1317. This meant that two experts were assigned a weight. They received weights 0.5288 and 0.4712 after normalization. As noted above it is not uncommon that a substantial number of respondents receive the weight zero with this method. The aim is to identify those respondents that are likely to be well calibrated on the questions at issue.

### B. *Success rate in the scenarios*

The respondents' weights were used to construct the estimates on denial of service attacks' success rate given different conditions, i.e., the weighted mean of their distributions was calculated. The estimated distributions were assumed to be distributed in the same way as they were presented to the respondents, i.e., as depicted in Figure. Note that certain variables are kept constant over the scenarios, c.f. section II.C.

#### 1) *Semantic attacks*

As depicted in Table IV the synthesized estimates show clear differences among the scenarios. The median for the scenarios varies between 16 and 76 percent; the value at the 5th percentile varies between 2 and 32 percent; the value at the 95th percentile varies between 56 and 95 percent.

**Table IV. Attack scenarios for semantic attacks.**

| # | Target | SV | AC | 5% value | 50% value | 95% value | Expected value |
|---|---|---|---|---|---|---|---|
| 1 | Machine | Yes | Yes | 0.32 | 0.76 | 0.95 | 0.72 |
| 2 | Machine | Yes | No | 0.14 | 0.56 | 0.80 | 0.53 |
| 3 | Machine | No | Yes | 0.22 | 0.62 | 0.94 | 0.60 |
| 4 | Machine | No | No | 0.05 | 0.37 | 0.69 | 0.38 |
| 5 | Service | Yes | Yes | 0.10 | 0.48 | 0.93 | 0.50 |
| 6 | Service | Yes | No | 0.08 | 0.25 | 0.67 | 0.30 |
| 7 | Service | No | Yes | 0.11 | 0.42 | 0.86 | 0.46 |
| 8 | Service | No | No | 0.02 | 0.16 | 0.56 | 0.21 |

In general it is more difficult to cause DoS for a single service than it is to cause DoS for an entire machine. As expected it is also more difficult to cause DoS in scenarios

where there is access controls restricting access and where there is no software vulnerabilities.

The estimates in Table IV are on the same format as results from a factorial experiment investigating all possible combinations. The influence strength of variables and their interactions can be calculated by comparing the scenarios with each other. For instance, the mean influence a software vulnerability (*SV*) has can be assessed as the mean of pairwise difference between scenarios #1 and #3, #2 and #4, #5 and #7, and #6 and #8.

The variable weights are depicted in TABLE V. The values show the influence this variable, or variable combination, have on the success probability. The target and presence of a software vulnerability are most important. If a machine is targeted (and not a specific service alone) the probability of success increase by 19 percent on average; the increase that comes from a software vulnerability is 21 percent. If the attacker has access credentials it increases the success rate with about 10 percent on average. The variables are more or less independent. This can be seen from the low values associated with variable combinations. These show the impact these particular combinations have on the success probability. For instance, the combination of software vulnerability and access credentials has the joint effect on the expected value of minus two percent units. The joint effect in addition to their individual influence of 21 and 10 percent units is thus comparably small.

**TABLE V. SEMANTIC ATTACKS – INFLUENCE OF VARIABLES ON THE SUCCESS RATE.**

| Variable or variable combination | 5% value | 50% value | 95% value | Expected value |
|---|---|---|---|---|
| Machine | +0.11 | +0.25 | +0.09 | +0.19 |
| SV | +0.12 | +0.24 | +0.24 | +0.21 |
| AC | +0.06 | +0.12 | +0.08 | +0.10 |
| Machine & SV | +0.06 | -0.01 | -0.04 | -0.01 |
| Machine & AC | +0.04 | +0.05 | -0.02 | +0.03 |
| SV & AC | -0.02 | -0.02 | -0.04 | -0.02 |
| Machine & SV & AC | +0.02 | -0.01 | -0.02 | 0.00 |

*2) Brute force attacks*

Table VI list the estimates for brute force attacks in terms of the number of hosts required to attain a certain level of unavailability for users, i.e., 5, 50 and 95 percent ignored legitimate traffic. An intrinsic interval [13] of 10 percent was used to estimate the expected number of host required to denial a legitimate user access.

**TABLE VI. ATTACK SCENARIOS FOR BRUTE FORCE ATTACKS – HOSTS REQUIRED TO CAUSE UNAVAILABILITY.**

| # | Roaming | Load balancing | 5% unav. | 50% unav. | 95% unav. | Expected value |
|---|---|---|---|---|---|---|
| 1 | Yes | Yes | 15 | 30 | 58 | 33 |
| 2 | Yes | No | 15 | 21 | 47 | 25 |
| 3 | No | Yes | 15 | 26 | 43 | 26 |
| 4 | No | No | 10 | 18 | 38 | 20 |

The variable weights derived from these scenarios are shown in Table VII. Both load balancing and roaming has an effect on the number of host required. The joint effect is marginal also here. To have both at the same time only increase the expected number of hosts required with one, in addition to their individual effects.

**TABLE VII. BRUTE FORCE ATTACKS – INFLUENCE OF VARIABLES ON HOSTS REQUIRED TO CAUSE UNAVAILABILITY.**

| Variable or variable combination | 5% value | 50% value | 95% value | Expected value |
|---|---|---|---|---|
| Load balancing | +2.5 | +3.5 | +12 | +6 |
| Roaming | +2.5 | +8.5 | +8 | +7 |
| Load balancing & Roaming | -2.5 | +0.5 | +3 | +1 |

## VI. DISCUSSION

The method used to analyze the experts' judgments and combine these is discussed in section VI.A below. The elicitation instrument used is discussed in section VI.B. The result as such and the importance variables included in the study are discussed in in section VI.C.

### *A. The expert judgment analysis*

In this study Cooke's classical method [13] was used to synthesize expert judgments. This performance based method aims to select the experts that are well calibrated and combine their judgments in an optimal way. The track record of this method [14] positions it a best-practice when it comes to combining eliciting expert judgment of uncertain quantities.

Eleven seed questions were used to evaluate calibration and information scores. These seed questions are of two types. The first type of seed questions is drawn from a vulnerability database [5]. The second type is drawn from a survey on brute force attacks [15]. They have an obvious relation to the questions of interest and are therefore suitable for rating the respondents.

A concern to the validity is that these sources are available to the respondents who could have used them to identify the answers to the seed questions. If they would do so these seeds would not work well as a gauge for how well calibrated and informative the expert's own judgment is. However, it is unlikely that anyone did so. None of the respondents answering the survey has given comments that indicate that they have realized that the correct answer can be found in online databases or in publications. Also, the uncertainty expressed in their answers suggests that they did not base them directly on these sources.

The answers on the seed questions show that many experts in the field are poorly calibrated, i.e., their estimates do not match empirical observations well. Two respondents were assigned weight when the virtual decision maker was optimized. It is appropriate to perform robustness test of the solution when applying Cooke's classical method [13]. These are made with respect to both seed variables experts by removing one at a time and investigating the impact of this removal [13]. Such tests were performed and no undue influence was identified.

Experts are better at estimating quantities in domains where they are possible to learn from observations, e.g. from experiments or simulations [16]. In the survey the respondents were asked to state from where they had obtained the knowledge used to answer the survey's questions. Of the 22 respondents whose assessment was analyzed 10 had defended systems in practice, 20 had learnt

from simulations, 22 had learnt from literature and 9 had learnt it from experiments. The two respondents receiving weight from Cooke's classical method had defended systems, learnt from simulations, and learnt from literature.

### B. Validity and reliability of the elicitation instrument

Cooke [13] suggests that seven guidelines used when data is elicited from experts: (1) formulate clear questions, (2) use an attractive format for the questions and a graphical format for the answers, (3) preform a dry run, (4) have an analyst present during the elicitation, (4) prepare an explanation of the elicitation format and how answers will be processed, (6) avoid coaching and (7) keep elicitation sessions to less than one hour long.

This study follows with all these guidelines except (4) – to have an analyst present during elicitation. The invited researchers were given contact information to the research group when invited to the survey which they were encouraged to use any if questions arose. However, it is possible that analysts' physical absence suppressed some potential issues from being brought up during elicitation. The respondents were asked to comment the clarity of the questions and the question format used in the survey. Two respondents indicated that they had difficulties with answering in the format used while several others stated that the format was clear and understandable. The two respondents who had difficulties would have preferred a format without probability distributions instead; an ordinal rating was suggested instead. While this probably would make the questions easier to answer it would also be less expressive and more difficult to interpret.

### C. Variables importance to the success rate

This study investigated three variables related to semantic attacks and two variables related to brute force attacks.

With respect to sematic DoS attacks the result indicates that it is easier to cause DoS for an entire machine than it is to cause DoS in a specific service. The increase on the success rate is on average 20 percentiles, which increase on the success rate with about 50 percent on average. The same magnitude of influence comes from to the existence of software vulnerabilities. If the attacker can authenticate itself to the target this increase the success probability with approximately 10 percent units. Removing software vulnerabilities and implementing access control that protects service's functionality against illegitimate users are two measures that can be implemented by decision makers. Together they would decrease the success probability with about 30 percent units and thereby reduce the probability of success to about half of what it would be without these measures.

With respect to brute force attacks, e.g., distributed DoS attacks, load balancing and roaming both increase the requirements placed on the attacker. Together they increase the number of hosts required to succeed with DoS by about 50 percent. Looking at the confidence intervals in TABLE VII it also appears as if load balancing primarily help to protect against a complete DoS (c.f. the 95 percent value in TABLE VII), but it has less impact on the number of hosts required to make some users experience unavailability.

The scenarios estimated in this study do not detail all variables of relevance. As this was the case the respondents were asked to provide probability distributions representing the values for typical enterprises. If variations exists between enterprises (e.g. in terms of other protection mechanisms, hardware capacity, etc.) this should be accounted for in the estimates and thereby spread the estimated distributions over larger intervals. Judging from the span of the intervals on semantic attacks there are possibilities to increase (or decrease) the defense with other variables than the one included here. For instance, for the five percent of best defended systems the success probability of semantic attacks is below two percent, given that software vulnerabilities are removed, access controls are between the attacker and the target is a specific service (see #8 in TABLE VII). Conversely, the success probability for the same scenario is above 56 percent for the least defended five percent. How much of this uncertainty that arise from epistemic uncertainty and how much that arise from variations between enterprises is difficult to know. But it appears likely that both contribute to the uncertainty reflected in the estimates.

The variables included in this study were selected based on literature with the assistance of domain experts. To narrow the intervals and allow more precision, further variables need to be included in the scenarios' definitions. The respondents of the survey were asked to suggest other variables that they would like to replace the selected variables with. The suggestions were diverse, which suggests that the most significant factors were included. The full list of suggestions of variables included: defining if it is forced or strict load balancing, the amount of redundancy used by the load balancer, adjustments of the load balancer, routing schemes, the number of requests the target is designed for, and bandwidths of connections. Further work could explore these variables impact and produce narrower probability distributions. Based on the result presented here it appears as if the influences of the studied variables are independent. This could be valuable input to further work on this field.

## VII. CONCLUSION

This research generalizes quantities related to DoS attacks using expert judgment available in the research community and present approximate estimates on attackers ability cause DoS. The result shows the weight of key factors in semantic attacks and brute force attacks. Applying measures that are included in this research does have a significant impact on the success rate for semantic attacks and the number of controlled host required for a brute force attack. However, the result also shows the variation that is expected to be found between enterprises solutions through the probability intervals produced. The cause of these intervals is likely to arise because from a number of factors. The impact of other factors and their influence on the success of DoS attacks could be investigated in further work. This could include investigations of how large the epistemic uncertainty is about the actual values, i.e., how precise the research community's knowledge is on DoS attacks and factors that influence their success.

## VIII. REFERENCES


[1] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of network-based defense mechanisms countering the DoS and DDoS problems," *ACM Computing Surveys*, vol. 39, no. 1, p. 3-es, 2007.

[2] R. Chertov, S. Fahmy, and N. B. Shroff, "Emulation versus simulation: A case study of TCP-targeted denial of service attacks," in *Testbeds and Research Infrastructures for the Development of Networks and Communities, 2006. TRIDENTCOM 2006. 2nd International Conference on*, 2006, p. 10–pp.

[3] J. Mirkovic and P. Reiher, "A taxonomy of DDoS attack and DDoS defense mechanisms," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, p. 39, Apr. 2004.

[4] A. Hussain, J. Heidemann, and C. Papadopoulos, "A framework for classifying denial of service attacks," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, 2003, p. 99110.

[5] NIST, "National Vulnerability Database Home Page," 2010. [Online]. Available: http://nvd.nist.gov/. [Accessed: 16-Jun-2010].

[6] P. Mell, K. Scarfone, and S. Romanosky, "A complete guide to the common vulnerability scoring system version 2.0," in *Published by FIRST-Forum of Incident Response and Security Teams*, 2007, pp. 1-23.

[7] A. Vorobiev and J. Han, "Security Attack Ontology for Web Services," *2006 Second International Conference on Semantics, Knowledge, and Grid*, pp. 42-42, Nov. 2006.

[8] X. Qie, R. Pang, and L. Peterson, "Defensive programming: Using an annotation toolkit to build DoS-resistant software," *ACM SIGOPS Operating Systems Review*, vol. 36, no. SI, pp. 45–60, 2002.

[9] C. Douligeris, "DDoS attacks and defense mechanisms: classification and state-of-the-art," *Computer Networks*, vol. 44, no. 5, pp. 643-666, Apr. 2004.

[10] S. M. Khattab, C. Sangpachatanaruk, R. Melhem, and T. Znati, "Proactive server roaming for mitigating denial-of-service attacks," in *Information Technology: Research and Education, 2003. Proceedings. ITRE2003. International Conference on*, 2003, pp. 286–290.

[11] C. Sangpachatanaruk, S. M. Khattab, T. Znati, R. Melhem, and D. Mossé, "A simulation study of the proactive server roaming for mitigating denial of service attacks," in *Proceedings of the 36th annual symposium on Simulation*, 2003, p. 7.

[12] R. T. Clemen and R. L. Winkler, "Combining probability distributions from experts in risk analysis," *Risk Analysis*, vol. 19, no. 187, pp. 187-204, 1999.

[13] R. Cooke, *Experts in uncertainty: opinion and subjective probability in science*. 1991.

[14] R. Cooke, "TU Delft expert judgment data base," *Reliability Engineering & System Safety*, vol. 93, no. 5, pp. 657-674, May. 2008.

[15] Forrester Consulting, "The trends and Changing Landscape of DDoS Threats and Protection," *Study on behalf of VeriSign, Inc*, 2009.

[16] F. Bolger and G. Wright, "Assessing the quality of expert judgment: Issues and analysis," *Decision Support Systems*, vol. 11, no. 1, pp. 1-24, Jan. 1994.

[17] Elsevier B.V., "Scopus," 2011. [Online]. Available: http://www.scopus.com/.

[18] Elsevier Inc, "Engineering Village," 2011. [Online]. Available: http://www.engineeringvillage.com. [Accessed: 24-Feb-2011].

[19] S. T. Cavusgil and L. A. Elvey-Kirk, "Mail survey response behavior: A conceptualization of motivating factors and an empirical study," *European Journal of Marketing*, vol. 32, no. 11/12, pp. 1165–1192, 1998.

[20] P. H. Garthwaite, J. B. Kadane, and A. O'Hagan, "Statistical methods for eliciting probability distributions," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 680-701, 2005.

[21] L. J. Cronbach and R. J. Shavelson, "My Current Thoughts on Coefficient Alpha and Successor Procedures," *Educational and Psychological Measurement*, vol. 64, no. 3, pp. 391-418, Jun. 2004.

[22] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.