Abstract—The Internet is on the way of becoming the universal communication network, and then needs to provide various services with guaranteed quality for all kinds of applications. Denial of Service (DoS) attacks are then more efficient in a guaranteed multi-services network than in the "old" best effort Internet. Indeed, with best effort services, a DoS attack has to forbid the target of the attack to communicate. With a multi-services network, it is sufficient to make the network not respect the SLA (Service Level Agreement) committed with clients, what is easier and can be performed using simple flooding attacks. Then, the question is: how does a DoS attack impact the quality of service (QoS) of a network given that networks are hugely over-provisioned, and that DoS attacks never succeed to completely overflow these high speed networks? This paper aims at answering this question as we do believe that it can help for defending the network against such attacks. The analysis of DoS attacks has been performed using traffic monitoring tools on the Internet. In particular, the analysis of attacks shows that they are increasing long range dependence (LRD) in the traffic, breaking the invariant power laws of normal Internet traffic. It is also explained in the paper, based on some normal traffic traces characterization and analysis why LRD is such a bad parameter for having good QoS.

Index Terms—Internet monitoring, traffic characterization, DoS attacks, QoS

I. INTRODUCTION

The Internet is on the way of becoming the universal communication network for all kinds of information, from the simple transfer of binary computer data to the transmission of voice, video, or interactive information in real time. The Internet is evolving from a single best effort service to a multi-services network. As a consequence, the Internet is very sensitive to attacks and especially DoS and distributed DoS attacks. Indeed, if the network is the target of the attack (as with DNS attacks for instance) or if it is not the target (as with a "TCP flooding" attacks whose target is one of the network users), large changes in traffic characteristics – which we call disruptions – due to these attacks can provoke some changes in the QoS perceived by all users of the network, and then break the service level agreement (SLA) at the Internet service provider (ISP) fault. DoS attacks can then provoke financial losses for ISPs.

Fighting DoS is very hard. Even knowing what constitutes DoS is a difficult problem. Current intrusion detection systems (IDS), and especially the ones based on anomaly detection are not very efficient at detecting DoS. Indeed, it is hard to distinguish DoS from traffic that presents marked legitimate variations. These are conclusions that have emerged from the METROPOLIS and METROSEC projects\textsuperscript{1} as well as from many other recent research projects across the globe, which have shown that Internet traffic is very far from being regular, and presents large variations in its throughput at all scales\textsuperscript{14}. These projects have shown that Internet traffic exhibits characteristics such as self-similarity\textsuperscript{16}, (multi-)fractality\textsuperscript{6}, and long-range dependence (LRD)\textsuperscript{5}, which is to say in all cases that traffic can vary significantly. In addition, given the highly variable nature of Internet traffic, anomaly based IDS are raising alarms for many disruptions that are not attacks. The high rate of false positives is one of the major shortcomings of current IDS and the current evolution of Internet traffic with larger and larger variations among time continues to limit the efficiency of anomaly based IDS.

In order to detect and fight DoS attacks, it is important to study and analyze the way a DoS attack works. The question is: how does a DoS attack impact the QoS of a network? In particular, how can network QoS be reduced during a DoS attack, even if the network has enough capacity to handle it? It is important to note that generally high speed networks, especially in the core backbone, are not overflowed by DoS attacks, even by the ones based on flooding. But these attacks nevertheless impact the network QoS, thus often breaking the SLA between client and carrier. The main goal of the work described in this paper is to study the impact of attacks on the network, its traffic, and its QoS.

This work proposes to use network monitoring and analysis tools which aim at finding out the characteristics of current Internet traffic. The objective deals with comparatively analyzing the characteristics of normal traffic, and traffic containing DoS attacks, trying to isolate the pa-

\textsuperscript{1}METROPOLIS and METROSEC are French projects granted and funded by RNRT and ACI S&I respectively. This work also got the support of the European E-NEXT network of excellence.
rameters responsible of the QoS degradation. However, because of existing results (quoted just above) on traffic characterization, modeling and analysis, this work focuses mainly on traffic dynamics, and, therefore, on high order statistical moments to characterize them. Indeed, it has been shown in previous projects that traffic dynamics exhibit invariant laws. Our work aims at studying whether these invariant laws could be impacted by DoS attacks. If any, these changes could represent signatures for attacks that might help for detecting and fighting them.

This paper is constructed as follows: section II deals with characterizing and analyzing normal Internet traffic, i.e. traffic without attack. The results presented have been obtained on RENATER traffic. RENATER traffic analysis helped us to understand the causes of current Internet traffic characteristics, especially its huge variability. In particular, section II recalls that traffic variability is characterized by Long Range Dependence (LRD) which is very damaging for the network QoS as they make it very unstable. Section II also recalls that the LRD function is “invariant” for long periods of time (several hours) on all tested Internet links (as long as attacks are not arising), and exhibited two invariant power laws that are due to transmission protocols, and in particular TCP. Section III proposes the same analysis, but this time for traffic containing attacks. In particular, it is shown in section A that attacks break the “invariant” power laws of the LRD function, and that attacks introduce more LRD in the global traffic, and then more disturbances for QoS. Section III.B then discusses on the robustness of this LRD based method for detecting the presence of attacks in the traffic. Section IV discusses about some few other studies that also start to use high order statistics for detecting DoS attacks. This section also discusses some ideas on how these results can be used for improving network security against DoS attacks in the future.

II. Traffic characterization

This section presents the characterization and analysis results of RENATER traffic. These results have been produced by computing several traffic traces captured on the RENATER network. For this purpose, we deployed 3 years ago, 3 DAG systems [4] that are capturing traffic traces on demand. DAG systems are located in Toulouse (1) and in Paris (2) which are two main places in the RENATER topology. Each trace consists of at least 4 hours of traffic (depending on the average throughput on the monitored link). The analysis of all the traces we have been capturing since three years shows very similar results. Note however that the results presented in this section are not surprising as the same results have been already described in the existing literature on traffic characterization, analysis and modeling (some of the publications exhibiting similar results are referenced in the text). This section then describes our characterization and analysis work on RENATER traffic and justifies why we decided to focus on LRD for characterizing attacks. In particular, we have shown that LRD is a parameter which characterizes and quantifies the QoS provided by a network [13].

However, to start and to well understand the new traffic characteristics, it is first required to analyze the evolution of the Internet in terms of usages. The evolution of Internet traffic these last years has been marked by the increase of P2P traffic (Kaaza, eDonkey, ...), and now, on some links of the RENATER Network, it can represent the same proportion than HTTP traffic [13]. In fact, the amount of P2P traffic in RENATER is pretty low compared to the results observed on the commercial network of France Télécom3, especially on the ADSL POP were P2P traffic can grow up to 70% - and sometimes more!

Such an increase of P2P has necessarily an impact on traffic characteristics. In particular, because of the nature of file exchanged - mostly music and movies - (0.01 s, 0.1 s and 1s), it appears that Internet traffic does not smooth as fast as Poisson traffic when increasing observation granularity. The analysis demonstrated that this result is completely due to elephants. In fact, the transmission of elephants creates in the traffic the arrival of a large wave of data that has the particularity of lasting for a long time - more than 1 second - while web flows are generally transmitted in less than one second on the current Internet. That is why we have this difference between Poisson and real traffic: the nature of oscillations between the two traffics changes, with oscillations in actual current traffic that are more persistent.

An in deep analysis of LRD and its causes presented in [13] demonstrated that it is due to TCP and its congestion control mechanisms – slow start and congestion avoidance – which are not suited for transmitting long flows on high speed networks. While transmitting elephants on such high speed networks, TCP generates traffic with strong variations lasting long, then explaining the characteristics depicted on figure 1. Then, with the current Internet us-

3 France Télécom R&D is part of the METROPOLIS project, but the results got on the France Télécom network are not public and will not be discussed more in this paper.

4 Poisson traffic is taken as the reference as it was a good model to represent traffic several years ago. And even nowadays, Poisson is most of the time the model considered by engineers and researchers in simulations, or for evaluating network performances, ..., even if Internet traffic is nowadays completely not respecting a Poisson model, and if using it leads to dramatic mistakes.
ages, the increase of large files as music or video exchanged
thanks to P2P applications favors high amplitude oscillations,
dependent on very long ranges. Of course, oscillations are very
damaging for the global utilization of network resources. It
introduces a decrease of the global QoS of the network. In fact, the more the traffic oscillates, the
lower the performances [15].

But what has been shown in what precedes just presents
some qualitative results on the nature of Internet traffic.
It is however also needed to quantify its LRD characteristics.
For that, it is most of the time accepted to use the
auto-correlation function, that can show on a graph the
correlation (and then the dependence) that exists between
the transmission of two packets separated by k packets.
The auto-correlation function can show for large values
of k the trends of LRD in the traffic. But, for accuracy
reasons, and to limit bias in the computing of LRD, it is
recommended to use a wavelet based analysis which allows
a multiscale analysis of the traffic [1]. The principle deals
with extracting from the traffic several wavelet functions
with different frequencies to get the density of the differ-
ent ranges of variability of the analyzed traffic. The waves
with the largest periods represent the very long waves, i.e.
the ones generated by elephant flows. Interested readers
can refer to [2].

The curve on Figure 3.(a) has been obtained using the
LD Estimate tool [1] that estimates the LRD that appears
in Internet traffic at all scales. The output of this tool is a
graphical representation of the dependence laws at all time
scales. Also note that the Hurst factor H that is the fac-
tor fully characterizing a self-similar process - and Internet
traffic is often said to be self-similar [10] [16] - can be ob-
tained directly depending on the slope of the LRD curve.
The curve on figure 3.(a) shows a bi-scaling phenomenon
(two lines in a log-log scale)[6], with an elbow around octave

5 The auto-correlation function is equivalent to the auto-covariance
function. The two names designate the same function. This is a very
basic function in statistics.

6 It is important to note that the same qualitative result as been
6, which shows a difference in the LRD level between short
and long time scales for the traffic exchanged, and meaning
that there are two different invariant power laws. For short
scales (octave < 6, left part of the curve), representing
the dependence between close packets (i.e. packets whose
sending time are not very far from each other), the depen-
dence is limited. Such dependence is the one that can exist
for packets belonging to the same congestion window and
that are then very close from each other. On the other side,
for long time scales (octave > 6, right part of the curve),
LRD can be very high. For octaves greater than 6 that cor-
respond for instance to the dependence between packets of
consecutive congestion windows, the dependence is higher.
This is explained by the closed control loop of TCP conges-
tion control mechanisms in which the sending of one packet
of a congestion control window depends on the receiving of
the acknowledgement of one packet of the previous conges-
tion control window. Of course, this phenomenon exists
for consecutive congestion window, but also for all congestion
windows of the same flow. This means that the presence in
the traffic of very long flows introduces very long scale de-
pendence phenomenon, as depicted on figure 3.(a) for very
large octaves. Note however that some additional experi-
ments also showed that the elbow in the curve corresponds
to the mean size of flows, meaning that the right part of
the curve corresponds to the impact of elephant flows.

What comes out from this LRD analysis, given the fact
that traffic LRD has a very bad impact on network QoS
[13], is that the use of network resources is far from being
optimal (especially because, as it has been demonstrated
before, TCP is not suited for the transmission of long flows
on high speed links). This implies that LRD forces Internet
carriers to hugely over-provision link capacities compared
to the amount of effective traffic to transmit. It comes
out that LRD (that is a parameter helping to characterize
the variability of Internet traffic) is a good parameter for
quantifying the level of QoS a network can provide in the
transmission of the considered traffic. The higher the LRD,
the worse the QoS [13] [15].

III. IMPACT OF DoS ATTACKS ON TRAFFIC CHARACTERISTICS

This section then proposes the same analysis as the one
presented in section II, but this time the traces that are
analyzed contain DoS attacks.

A. DoS attacks and LRD

In fact, while analyzing RENATER traffic, it appears
that this traffic does not contain any DoS attacks that
are removed at the access points of the network by the
CERT-RENATER. Looking for some illegitimate traffic,
we have just been able to observe some port scanning ac-
tivity. Therefore, we have been obliged to generate DoS
attacks in the traffic to measure and analyze its impact.
Generating ourselves DoS attacks is a good point for the
accuracy of our following analysis, as we can completely
exhibited for all links that have been monitored all over the world by
researchers working on Internet links monitoring and using this tool

Fig. 1. Comparison between Internet and Poisson traffics
control the profile of this attack and then perfectly analyze the impact of each parameter of the attack on traffic.

Several DoS attacks have been generated and their impact on the traffic analyzed: TCP Syn flooding, UDP flooding, Smurf, etc. All these attacks have also been generated with different “intensity” parameters (Syn packet rate for Syn flooding, packet rate and size for UDP flooding, number of reflectors and packet rate for Smurf attacks, etc.), and on 10 minutes periods. Practically speaking, they have been generated between LAAS (in Toulouse) and LIP6 (in Paris). Results got with the different DoS attacks tested (that were all causing some flooding of their targets) are very similar. Because of space available, we chose to only present the results of the UDP flooding attack which is the most representative flooding attack.

Figure 2.(a) shows the auto-correlation function of the normal traffic. As it has been said before, auto-correlation is generally admitted as a good way for analyzing dependences properties in the traffic. Figure 2.(b) shows the auto-correlation function of the traffic containing an attack. It clearly appears on figure 2 that the traffic containing an attack is more correlated. As correlation implies dependence, and we showed that dependence has a bad impact on QoS, it is easy to understand how the DoS attack manages to reduce the network QoS. In addition, such increase of the auto-correlation level in a traffic trace is a signature of the presence of an attack. How such signature can be used will be discussed in section IV.

Finally, for having a better quantitative evaluation of the traffic dependence introduced by an attack in the network, Figure 3.(a) shows the LRD function for normal traffic and figure 3.(b) traffic with an attack. It appears that the two invariant power laws disappear during the attack, and that the new LRD function presents now 3 power laws, that are invariant for the whole duration of the attack. In addition, by changing the frequency of attacking UDP packets, we have been able to observe that the peak between power laws identified as 2 and 3 on figure 3.(b) moves accordingly. As for auto-correlation, it is easy, given the LRD analysis, to measure the impact of an attack on traffic QoS. As well, the new LRD function is also a signature of the presence of a DoS attack in the traffic.

B. Robustness of LRD based analysis of DoS attacks

Given the interesting results presented in section III.A that shows how a DoS attack manages to reduce network QoS by introducing in the traffic some extra LRD, it is necessary to evaluate the robustness of this approach, i.e. to evaluate if it is possible for a hacker to generate DoS attack that would not change traffic auto-correlation and LRD functions?

Of course, it is very difficult to give a formal evaluation of this complexity for a hacker. Besides, in this section we are only discussing this question according to theoretical mathematical aspects. Nevertheless, we will also give an informal illustration based on our experience in re-simulating traffic and some experiments. Intuitively, as the detection method relies on computing second order statistical moments - auto-correlation and LRD - that are related to very dynamic characteristics of Internet traffic, it is mathematically evident that it would be very difficult for a hacker to generate attacks having exactly the same dynamics. We can argue on this point by quoting our experience for playing traffic having all characteristics of original traffic [12]. In fact, the goal of the work described in [12] was to be able to play in simulators or emulators traffic having all the real characteristics of real traffic, especially the same correlation and LRD functions that are the most difficult characteristics to reproduce. [12] shows that we got good results by replaying traffic traces captured on the network we want to simulate. But, for a hacker, being able to generate traffic having the right correlation and LRD functions means having access to traffic traces captured on the network he wants to attack. Indeed, the presence of the two invariant power laws is a qualitative result. But it is to note that the quantitative values for these power laws are different from one link to the other. In addition, these power laws are quantitatively stable on periods lasting several hours, but their values change according to the day periods, and for instance they are different during night and day, different in the morning, at lunch time, in the evening, etc. It is then improbable that a hacker can have the monitoring information needed for all the links that will be crossed by his attack from the attack source(s) to the attack target, at the exact moment he wants to generate his attack. In addition, even if the hacker succeed to generate an attack that will not change the traffic correlation and LRD functions, and given the analysis we performed on the impact of LRD on QoS decrease, it is then not sure that the QoS of the network on which the attack is transmitted will be impacted?

The second aspect of LRD based detection method we want to discuss for evaluating its robustness deals with studying if a DoS attack impact on LRD can be hidden in some cases, especially if the attack is transmitted on a link propagating a very huge amount of traffic. The question then is: is the traffic LRD function impacted enough by a DoS attack to exhibit the three power laws even if the attack is hidden by a very big traffic? The theoretical answer based on mathematical theory is yes. We also checked this conclusion on 1 Gbps links. But we were unable to raise this experiment on a faster link, as we are not monitoring links faster than 1 Gbps. However, we are confident with our positive answer because of the theory of statistics.

IV. RELATED WORK AND FUTURE RESEARCH DIRECTIONS FOR FIGHTING DOSS

This kind of method for analyzing the impact of attacks is quite original, and has not been addressed much for the moment in research. There have been several attempts based on the single auto-correlation function ([8] for example), but this function does not provide enough information and, as it is illustrated in this paper, it is necessary to use a complementary function to go into more details. [11] chose to use the LRD function and reaches the same results as us, but with the final objective to propose new
techniques for anomaly based IDS (what is not our goal: see further). On another hand, the work by [7] relies on analyzing the impact of DoS attacks on traffic spectral density. This work and ours that relies on auto-correlation and wavelet analysis (that are in fact two very close approaches for evaluating almost the same properties) reach the same conclusion. This approach then appears as very promising.

In fact, the main contribution of both papers concerns the presence of invariant power laws in the traffic that can be impacted by DoS attacks. These changes, in both papers are therefore considered as signatures for attacks, and these signatures can help to detect their presence in the traffic. In particular, these changes can help to detect attacks that are usually transparent to classical signature or anomaly based IDS. More generally, this work can restart the interest for profile based IDS whose detection principle can be based on more significant parameters for characterizing traffic properties, and not only the first statistic moments as it is still the case in many IDS. Anomaly based IDS are generally based on analyzing evolutions of first order statistics moments as mean. But these moments are also impacted by all kinds of traffic disruptions, and in particular legitimate ones, thus leading either to false negative, either to false positive, depending on how the detection system thresholds are tuned. As a consequence, these IDS were not very efficient and popular.

However, at the opposite of all quoted work as [7] [11], we do not propose to design a new kind of IDS based on the LRD detection principle. In fact, even if such approach appears as very powerful for detecting the presence of DoS attacks in the traffic, it still presents some important lacks: indeed, detecting a change in traffic power law does not indicate what attack constituting packets have to be discarded. Instead of putting our research effort in the design and development of a new IDS, and because we are considering the point of view of a carrier or ISP which has to enforce a stable and guaranteed QoS, we do believe that the analysis of attacks can be of great interest for helping to make the network more robust and not sensitive to attacks. In particular, the analysis we performed on the source of LRD, and that points out TCP, helped to design a new transport protocol - for instance MBCC [9] - that avoids
propagating LRD (and in particular the extra one created by DoS attacks) in the network. As well, by studying the topology of the network and its properties (the Internet is often said to have small world properties [3]) it should be possible to improve it for limiting the propagation of LRD, i.e. limit the propagation of the impact of attacks. Making the network able to continue providing a good QoS even in the presence of attacks is the direction we are following. In particular, it is more interesting than trying to design a new kind of IDS, as it will also be able to handle in the same way all kinds of disruptions in the traffic. Indeed, disruptions that make the network QoS change are not only due to attacks. They can also be due to faulty equipments (crashes that can change traffic matrices, byzanthin behaviors, etc.) but also to some legitimate increases of the traffic, for instance related to some very popular events broadcasted on the Internet (flash crowds). Disruptions in the traffic being quite frequent, our approach aiming at improving network robustness and making it able to continue providing good QoS in all cases should make it more efficient, compared to classical IDS, which are not able to handle all kinds of disruptions. In addition, an IDS has to be updated very often to be able to recognize new attacks discovered recently, with the huge risk that a new attack can pass the IDS.

With our approach, attacks will not impact network QoS. But even if attacks have no effect on the network QoS, it is not recommended to let the attacks continue transiting in the network. Of course, our approach can be combined with more classical security solutions. Indeed, a non-sensitive network will give some time for monitoring tools to detect the presence of an attack, to isolate attack constituting packets, and then to cut its source(s), for instance using traceback mechanisms as [17]. And thanks to network insensitivity such work has not to be performed in real time, what can also help in reducing the number of false positive.

VI. ACKNOWLEDGEMENTS

I want to thank Kaouther Blibech who started this work during her master internship at LAAS-CNRS under my advising.

REFERENCES