
*Università degli Studi di Salerno*

Dottorato di Ricerca in Informatica e Ingegneria dell'Informazione
Ciclo 30 – a.a 2016/2017

**ABSTRACT**

TESI DI DOTTORATO / PH.D. THESIS

# Statistical Models for the Characterization, Identification, and Mitigation of Distributed Attacks in Data Networks

**MARIO DI MAURO**

SUPERVISOR: **PROF. MAURIZIO LONGO**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Dipartimento di Ingegneria dell'Informazione ed Elettrica
e Matematica Applicata
Dipartimento di Informatica

# Abstract (english)

**PhD Thesis Title**

Statistical Models for the Characterization, Identification, and Mitigation of Distributed Attacks in Data Networks.


**Abstract**

The thesis focuses on statistical approaches to model, mitigate, and prevent distributed network attacks. When dealing with distributed network attacks (and, more in general, with cyber-security problems), three fundamental phases/issues emerge distinctly.

The first issue concerns the threat propagation across the network, which entails an "avalanche" effect, with the number of infected nodes increasing exponentially as time elapses.

The second issue regards the design of proper mitigation strategies (e.g., threat detection, attacker's identification) aimed at containing the propagation phenomenon. Finally (and this is the third issue), it is also desirable to act on the system infrastructure to grant a conservative design by adding some controlled degree of redundancy, in order to face those cases where the attacker has not been yet defeated.

The contributions of the present thesis address the aforementioned relevant issues, namely, propagation, mitigation and prevention of distributed network attacks. A brief summary of the main contributions is reported below.

The first contribution concerns the adoption of Kendall's birth-and-death process as an analytical model for threat propagation. Such a model exhibits two main properties: *i)* it is a stochastic model (a desirable requirement to embody the complexity of real-world networks) whereas many models are purely deterministic; *ii)* it is able to capture the essential features of threat propagation through a few parameters with a clear physical meaning. By exploiting the remarkable properties of Kendall's model, the exact solution for the optimal resource allocation problem (namely, the optimal mitigation policy) has been provided for both conditions of perfectly known parameters, and unknown parameters (with the latter case being solved through a Maximum-Likelihood estimator).

The second contribution pertains to the formalization of a novel kind of randomized Distributed Denial of Service (DDoS) attack. In particular, a *botnet* (a network of malicious entities) is able to emulate some normal traffic, by picking messages from a dictionary of admissible requests. Such a model allows to quantify the botnet "learning ability", and to ascertain the real nature of users (normal or bot) via an indicator referred to as MIR (Message Innovation Rate). Exploiting the considered model, an algorithm that allows to identify a botnet (possibly) hidden in the network has been devised. The results are then extended to the case of a multi-cluster environment, where different botnets are concurrently present in the network, and an algorithm to identify the different clusters is conceived.

The third contribution concerns the formalization of the network resilience problem and the consequent design of a prevention strategy. Two statistical frameworks are proposed to model the high availability requirements of network infrastructures, namely, the Stochastic Reward Network (SRN), and the Universal Generating Function (UGF) frameworks. In particular, since in the network environment dealing with multi-dimensional quantities is crucial, an extension of the classic UGF framework, called Multi-dimensional UGF (MUGF), is devised.

## Abstract (italiano)

**PhD Thesis Title**

Statistical Models for the Characterization, Identification, and Mitigation of Distributed Attacks in Data Networks.


**Abstract**

Il presente lavoro di tesi riguarda la caratterizzazione statistica di modelli per l'identificazione, la mitigazione, e la prevenzione di attacchi distribuiti su reti dati. A tal proposito, è possibile identificare tre problemi rilevanti.

Il primo riguarda il meccanismo di propagazione che caratterizza particolari tipi di attacchi, progettati in maniera tale da espandersi su una rete dati attraverso un procedimento di infezione "a catena": un nodo infettato diventa esso stesso un vettore di infezione dando così origine ad un'espansione a macchia d'olio della minaccia di rete.

Il secondo problema riguarda il fenomeno degli attacchi distribuiti (es. Distributed Denial of Service - DDoS), che prevedono tipicamente l'esistenza di una *botnet* in grado di saturare le risorse di un elemento di rete (es. un router, un web server, etc.). La botnet consiste in un esercito di nodi malevoli (*bots*) che inondano di richieste il "target" in maniera sincronizzata, provocando così un esaurimento delle sue risorse.

Il terzo problema riguarda invece le strategie di prevenzione di rete da mettere in campo per contrastare eventi anomali (attacchi informatici, spegnimenti improvvisi di apparati, etc.).

I contributi forniti in questo lavoro di tesi intendono affrontare le tre problematiche sopra menzionate.

Nel primo contributo, l'autore propone di adottare il modello Birth-Death-Immigration (BDI) ideato da Kendall nel 1948 (per modellare fenomeni di crescita di una popolazione), e che viene qui riadattato per affrontare il problema della propagazione di minacce di rete, unitamente alla progettazione di una soluzione di mitigazione. Il processo di mitigazione viene caratterizzato tramite la risoluzione di un problema di allocazione ottima di risorse considerando due casi: *i)* il vettore di infezione è noto, e la soluzione offerta è quella esatta; *ii)* il vettore di infezione è incognito, e la soluzione offerta è basata sulla stima a massima verosimiglianza del vettore di infezione stesso.

Il secondo contributo riguarda la formalizzazione di un nuovo modello di attacco DDoS a livello 7 della pila TCP/IP, dove una botnet è in grado di utilizzare messaggi "legittimi" collezionati (tramite attività di scanning su rete) da un *botmaster* (il supervisore della botnet) per confezionare un attacco distribuito. Viene quindi introdotto un indicatore chiamato Message Innovation Rate (MIR) utile per quantificare il grado di "innovazione" di una botnet (ovvero quanti messaggi legittimi differenti è in grado di utilizzare per unità di tempo). Dal MIR viene poi ricavata una soglia utilizzata all'interno di un algoritmo chiamato BotBuster che viene messo in campo per rivelare ed identificare i nodi malevoli di una botnet.

Il terzo contributo riguarda la formalizzazione di strategie di resilienza (in particolare alta affidabilità) di una rete dati attraverso la risoluzione di un problema di ridondanza ottima. Sono state utilizzate due tecniche in particolare: le Stochastic Reward Network (SRN) che consentono di descrivere un sistema (in questo caso un'infrastruttura di rete) in termini della sua distribuzione di stati, e la UGF (Universal Generating Function) che è utile per risolvere problemi di affidabilità per sistemi multi-stato. Della UGF ne è stata proposta una estensione utile per affrontare problemi multi-dimensionali, chiamata Multi-dimensional UGF (MUGF).