Neurocomputing 122 (2013) 13–23

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Network anomaly detection with the restricted Boltzmann machine

Ugo Fiore [a,*], Francesco Palmieri [b], Aniello Castiglione [c], Alfredo De Santis [c]

[a] Centro di Ateneo per i Servizi Informativi, Università di Napoli Federico II, Napoli, Italy
[b] Dipartimento di Ingegneria dell'Informazione, Seconda Università di Napoli, Aversa, Italy
[c] Dipartimento di Informatica, Università di Salerno, Fisciano, Italy

ARTICLE INFO

Article history:
Received 30 September 2012
Received in revised form
23 November 2012
Accepted 26 November 2012
Available online 11 June 2013

Keywords:
Anomaly detection
Restricted Boltzmann machine
Semi-supervised learning
Intrusion detection
Energy-based models

ABSTRACT

With the rapid growth and the increasing complexity of network infrastructures and the evolution of attacks, identifying and preventing network abuses is getting more and more strategic to ensure an adequate degree of protection from both external and internal menaces. In this scenario many techniques are emerging for inspecting network traffic and discriminating between anomalous and normal behaviors to detect undesired or suspicious activities. Unfortunately, the concept of normal or abnormal network behavior depends on several factors and its recognition requires the availability of a model aiming at characterizing current behavior, based on a statistical idealization of past events. There are two main challenges when generating the training data needed for effective modeling. First, network traffic is very complex and unpredictable, and second, the model is subject to changes over time, since anomalies are continuously evolving. As attack techniques and patterns change, previously gained information about how to tell them apart from normal traffic may be no longer valid. Thus, a desirable characteristic of an effective model for network anomaly detection is its ability to adapt to change and to generalize its behavior to multiple different network environments. In other words, a self-learning system is needed. This suggests the adoption of machine learning techniques to implement semi-supervised anomaly detection systems where the classifier is trained with "normal" traffic data only, so that knowledge about anomalous behaviors can be constructed and evolve in a dynamic way. For this purpose we explored the effectiveness of a detection approach based on machine learning, using the Discriminative Restricted Boltzmann Machine to combine the expressive power of generative models with good classification accuracy capabilities to infer part of its knowledge from incomplete training data.

1. Introduction

The main goal of a network anomaly detection system is to discriminate the occurrence of hostile activities from the normal network traffic, and such analysis must be accomplished in a sufficiently flexible and effective way to keep up with the continuously evolving world of cybersecurity where new, previously unknown, anomalies can continuously emerge over time. In doing this, it must either try to model any kind of attack or anomalous event that can affect the network (there are thousands of known ones) or simply construct a sufficiently general model describing the normal traffic.

Such model is usually built on the basis of training data, and used in classifying previously unseen or suspicious events. Classification is the fundamental task in unattended detection, by which the system "learns" to automatically recognize complex traffic patterns, to distinguish between different events based on the corresponding patterns, and to make "intelligent" decisions. Specific machine learning techniques, such as Neural Networks or Support Vector Machines can be used to develop a generalization capability from training data needed to correctly classify future data as normal or abnormal. These resulting approaches can be categorized as generative or discriminative. A generative approach builds a model solely based on normal training examples and evaluates each testing case to see how well it fits the model. A discriminative approach, on the other hand, attempts to learn the distinction between the normal and abnormal classes. Thus, based on the characteristics of training data used to build the model, anomaly detection can be divided into three broad classes [1]:

- *Supervised anomaly detection*: In this class, a training set containing labeled instances for both the normal and anomalous class is available.
- *Semi-supervised anomaly detection*: The training here only contains instances for the normal class. Anything that cannot be characterized as normal is thus marked as anomalous.
- *Unsupervised anomaly detection*: No training set is available nor is it needed.

Obviously, the quality of classification crucially depends on the accurateness and comprehensiveness of the model and hence of

* Corresponding author. Tel.: +39 392 8888562.
E-mail addresses: ufiore@unina.it, ugo.fiore@unina.it (U. Fiore).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.11.050

the training data on which the model is built. If a complete set of pre-classified or "labeled" categories of normal (or anomalous) behavior is included in the training set (which can be difficult to achieve and even harder to maintain), all the corresponding (or "matching") instances will be correctly classified. But if some of such categories are not described in the training data, the corresponding instances may be classified incorrectly. In particular, anomalous events are harder to describe, partly because of their negative definition (*a-nomalous* literally *means* non-normal). Anomalous events both appear less frequently than normal events and embrace a huge variety of aspects. In fact, one technique for producing a labeled training set consists in creating anomalies artificially, injecting anomalous events in a stream containing normal data. But the description of anomalies obtained in such a way is forcibly limited to the scope and characteristics of the artificial anomalies used. Labeled anomalous data gathered from operational, "real-world" networks are not readily available, for a number of reasons, including the sheer amount of effort needed to produce such data, and the reluctance of network administrators to divulge data that could compromise the privacy of their clients or exfiltrate privileged information about the internal structure of their networks. The immense amount of data to be potentially examined, combined with their complexity and with the level of expert knowledge that would be needed for the analysis is a strong motivator for making the training process as independent as possible from the availability of labeled anomalous data. Furthermore, these data are generated mainly through human intervention, as soon as the community becomes aware of a new menace and detailed information about the attack dynamics and behavior becomes available. This may require a not negligible time that is clearly unacceptable when real-time or timely response to anomalies is strictly necessary. Consequently, anomaly detection systems should avoid being limited by the knowledge of any predefined set of anomalies and should be able to flexibly recognize/classify any unknown event affecting the network operations according to a self-learning semi-supervised or better unsupervised detection approach, so that the knowledge of traffic behavior on which the model is based can be progressively constructed and dynamically change/evolve over time. However, building and training *in situ* such a generative self-learning model is a lengthy and costly process. Even when the semi-supervised model is used, the overall result will depend on the completeness of the training data, since it is difficult to represent all the features that normal behavior can have. Thus, in this work we focus on semi-supervised anomaly detection, with a perspective aiming at investigating whether normal traffic behavior (and, conversely, anomalous behavior) shares some inherent similarity that we can use to characterize it. The purpose of this analysis is not to concentrate on a near real-time intrusion detection and reaction system but, in a medium-term perspective, to work towards a better and more adequate description of network traffic, also aiming at being as adaptive as possible. The tool which has been selected for this analysis is the Discriminative Restricted Boltzmann Machine, a network of stochastic neurons behaving according to an energy-based model. These networks couple the ability to express much of the variability of data, given by generative models, with the good classification accuracy derived from discriminative classifiers. The main advantages of this approach are that in line of principle it is not restricted to any specific environment, or *a priori* knowledge base, and that it can enable the detection of any type of unknown anomalous events, being effective in coping with the so-called *zero-day* attacks.

In Section 2, related work is reviewed. An introductory summary of classical neural networks, energy-based models, and Boltzmann Machines is the subject of Section 3. The proposed model and its assumptions are detailed in Section 4. Sections 5 and 6 are dedicated to the description of the experiments and the discussion of their results. Section 7 contains some concluding remarks and directions for future research.

2. Related work

Anomaly detection in computer networks is a long-established area, with more than 40 years of evolution [2] and contributions that have explored many approaches [1,3]. In a wider perspective, other researchers have considered the time-dependent connection between events, focusing on the casual relationship between an event and its consequences. For example, the problem of correlating events from different sources to isolate the root cause has been investigated in [4]. The study of effects produced by hidden dynamics has been the object of works based on nonlinear analysis, first targeted to traffic classification [5] and then focused on the specific issue of network anomaly detection [6]. Feature-modeling anomaly detection techniques such as FRaC [7] focus instead on the linkage between individual features and attempt to build predictive models for each feature, based on the others. Features that deviate from this prediction indicate an anomaly. Semi-supervised learning has received attention by the research community recently. Mao et al. [8] have proposed a co-training framework based on multi-view data, semi-supervised learning and active learning. Their method requires user intervention. Chen et al. [9] evaluated the application of spectral graph transduction and Gauss random fields to the detection of unknown attacks. Besides classification, they also proposed a semi-supervised method for clustering. An unsupervised clustering algorithm based on competitive learning neural network is described in [10], where instability is reduced by means of a reward-punishment update rule.

3. Background

3.1. Anomaly detection

From a theoretical point of view our network anomaly detection problem can be formulated as follows [11].

A collection of traffic data measurements is described by a scalar time series $\{x_t\}_{t=1}^{T}$ governed by a probability distribution $p(\cdot)$. Although all these measurements are associated to the occurrence of specific events within the event space S, the correspondence between them may not be known in advance. We are interested in partitioning the event space S into two subspaces corresponding to the normal and the anomalous network traffic conditions. Also, we need to infer the membership of a particular event in one of the above subspaces starting from the corresponding time series values. To accomplish this task, since the probability distribution p describing the behavior of the time series is unknown, we can use a mechanism enabling the reconstruction of its volumetric representation from the collection $\{x_t\}_{t=1}^{T}$. A general approach to the problem of identifying this representation is based on building a Minimum Volume Set (MVS) characterized by a probability mass $0<\beta<1$ associated to the distribution p for a volume measure μ [12], that is:

$$G_\beta^* = \arg\min\{\mu(G) : p(G)\geq\beta, \quad G \text{ measurable}\} \tag{1}$$

In the most of the common cases μ can be chosen to be the *Lebesgue* measure, although such technique extend easily to other measures. The parameter β can be chosen by the user to reflect a desired false alarm rate of $1-\beta$.

These minimum volumes summarize the regions of greatest probability mass of the distribution p, and are useful for detecting