Scientists recognize intruders in noise

May 25 2021

Credit: Depositphotos

A team of scientists from MIPT and Kazan National Research Technical University is developing a mathematical apparatus that could lead to a breakthrough in network security. The results of the work have been published in the journal *Mathematics*.

Complex systems, such as network traffic or living organisms, do not have deterministic physical laws to accurately describe them and predict future behavior. In this case, an important role is played by correlation analysis, which describes the behavior of the system in terms of sets of statistical parameters.

Such complex systems are described by trendless sequences, often defined as long-term time series or "noise". They are fluctuations produced by a combination of different sources and are among the most difficult data to analyze and extract reliable, stable information.

One of the metrics used in economics and natural sciences in time series analysis is the Hurst exponent. It suggests whether the trend present in the data will persist: for example, whether values will continue to increase, or whether growth will turn to decline. This assumption holds for many natural processes and is explained by the inertia of natural systems. For example, lake level change, which is consistent with predictions derived from analysis of the Hurst exponent value, is determined not only by the current amount of water, but also by evaporation rates, precipitation, snowmelt, etc. All of the above is a time-consuming process.

Catching a cyber attack

The amount of traffic passing through network devices is enormous. This is true for the end devices—home PCs, but especially so for intermediate devices such as routers, as well as high-volume servers. Some of this traffic, such as video conferencing, needs to be sent with the highest priority, while sending files can wait. Or maybe it is torrent traffic that is clogging up a narrow channel. Or at worst, there is a network attack going on and it needs to be blocked.

Traffic analysis requires computational resources, storage space (buffer)

and time—what brings latency in transmission. All of these are in short supply, especially when it comes to low-power intermediate devices. Currently, it is either relatively simple machine learning methods, which suffer from a lack of accuracy, or deep neural network methods, which require quite powerful computing stations with large amounts of memory just to deploy the infrastructure to run, let alone the analysis itself.

The idea behind the work of the team of scientists led by Ravil Nigmatullin is quite simple: generalize the Hearst exponent by adding more coefficients in order to get a more complete description of the changing data. This makes it possible to find patterns in the data that are usually considered noise and were previously impossible to analyze. In this way, it is possible to extract significant features on the fly and apply rudimentary machine learning techniques to search for network attacks. Together, they are more accurate than heavy neural networks, and the approach can be deployed on low-power intermediate devices.

Noise is commonly discarded, but identifying patterns in noise can be very useful. For example, the scientists have analyzed the thermal noise of a transmitter in a communications system. This mathematical apparatus allowed them to isolate from the data a set of parameters characterizing a particular transmitter. This could be a solution to one of the cryptography problems: Alice sends messages to Bob, Chuck is an intruder who tries to impersonate Alice and send Bob a message. Bob needs to distinguish a message from Alice from a message from Chuck.

Data handling is penetrating deeply into all areas of human life, with image and speech recognition algorithms having long since moved from the realm of science fiction to something we encounter on a daily basis. This description method produces signal features that can be used in machine learning, greatly simplifying and speeding up recognition systems and improving the accuracy of decisions.

Alexander Ivchenko, a member of the Multimedia Systems and Technology Laboratory at MIPT, one of the authors of the development, says: "The development of this mathematical apparatus can solve the issue of parameterisation and analysis of processes for which there is no exact mathematical description. This opens up enormous prospects in describing, analyzing and forecasting complex systems."

More information: Raoul Nigmatullin et al, Generalized Hurst Hypothesis: Description of Time-Series in Communication Systems, *Mathematics* (2021). DOI: 10.3390/math9040381

Provided by Moscow Institute of Physics and Technology