My contribution to cybersecurity

Riccardo Gatti

Abstract

Entropy is a powerful concept but its application is limited and context-specific. In particular, understanding security is not only about the properties of the system but also about the nature of information exchange and interpretation. The concept of entropy extends beyond its boundaries. In biology, quantum mechanics, meteorology and influencing various fields through metrics and probabilistic models. Statistical varieties are the best embodiment of a new point of view on entropy. Thus, the question "when can a system be considered secure?" is the IT specialization of a more general question "how, how much and what kind of information can be transferred or is transferred?". The answer, in the IT context, is "when a system generates information that cannot be read outside the agents involved, in case of reading, cannot be understood and in case of understanding cannot be reintroduced into the system and in case of reintroduction cannot change the evolution of the system".

1 Introduction

The research I conducted aims to answer the question, "when can a system be considered secure?" within the field of cybersecurity. This issue was addressed by writing a paper that seeks to be foundational, deriving cybersecurity properties from Shannon original formulation of entropy used to formalize the concept of secrecy in the broadest sense. However, the concept of secrecy (and Shannon purely entropic formulation) is insufficient to determine when a system can be considered confidential, when a message can be deemed intact, or when a process can be regarded as available. Of course, to describe these properties, we have assumed that the security of a system is somehow related to the properties, which for reasons of backward compatibility and ISO standards, are referred to by the acronym CIA (Confidentiality, Integrity, Availability).

In this document, I aim to summarize my reflections, propose new ideas, and keep track of the work done so that I can revisit it in the future, despite the fact that the scope of my research is deeply connected to the studies conducted here.

2 Entropy is boundedly unlimited

The purely entropic description of information is the primary investigative tool used in cybersecurity to address issues related to communication. However,

the fact that entropy is tied to the distribution of symbols within a message does not preclude conceiving a broader definition that allows for the study of word distribution within a sentence, sentences within a text, and texts within knowledge. In this regard, I have encountered significant challenges, not so much in formalizing this extension, but in making it apparent. Perhaps a limitation in my expressive ability has hindered me from effectively communicating a concept that seems so clear and straightforward in my mind.

The problem of information directly stems from the problem of communication, which is formalized based on a language, including its grammar, dictionary, and alphabet. Grammar establishes the rules for composing a message but does not govern its interpretation. The history of language, from Leibniz, Frege, Russell, Hilbert, Gödel, Turing, to Chomsky, has demonstrated that interpretation involves meaning (semantics) rather than grammar (syntax). This issue was brilliantly addressed by the creation of formal languages, which ensure that each proposition has exactly one interpretation: the very essence of the proposition itself. In contrast, natural (non-formal) languages, despite adhering to the rules of syntax, do not allow for a unique interpretation since semantics is not fully captured by syntax as it is in formal languages.

This is where the problem of communication arises: *how, how much and what kind of information can be transferred or are transferring?* The security of a system depends on the intelligibility of the information exchanged within it, which implies that the initial question about security is fundamentally linked to this issue, and it must not remain unanswered. This is the central focus of my research. Cybersecurity is just one of many (possibly infinite, but certainly countable) contexts in which this question can be framed. In cybersecurity, a system is deemed secure when both the information and the system itself exhibit certain characteristics. In biology, the term "transferase" refers to the process of transferring a functional group from one cell to another. Quantum mechanics uses terms like "interaction" or "superposition" to describe the interdependence between observer and observed. Since the observer can only perceive what their senses allow, they are immediately involved in a myriad of interactions that make the system extremely complex to understand. Meteorology employs geostrophic equations, continuity equations, and fluid simulations to study interactions between air masses, clouds, ocean currents, phenomena, and climates.

$$\begin{aligned} &\textit{When can a system be considered secure?} \Rightarrow \\ &\textit{How, how much and what kind of information} \\ &\textit{can be transferred or are transferring?} \end{aligned} \tag{1}$$

Each discipline asks the same fundamental question, merely altering the context without changing the parameters. Cybersecurity, being an information technology field, involves the exchange of information through messages. These messages are exchanged based on the properties of the system, which in turn influences the nature of the messages. The “how” translates to the sys-

tem properties, the "how much" translates to entropy, and the "what kind" translates to the properties of the specific message. Thus, "transfer" equates to communication with the message as its object.

$$\begin{aligned} &\textit{When can a system be considered secure?} \iff \\ \textit{Cybersecurity} &\left(\begin{array}{c}\textit{How, how much and what kind of information}\\ \textit{can be transferred or are transferring?}\end{array}\right) \end{aligned} \tag{2}$$

If you want to know when a system can be considered secure, you must address the underlying question, even if it requires more effort. Computing has significantly contributed to answering this question by introducing the concept of entropy, which was originally used in thermodynamics and, more broadly, in physics. It may seem more or less coincidental that Shannon entropy formula is so structurally similar to Boltzmann entropy formula. Some consider this similarity to be a point of interest, while others view it as a mere curiosity. However, in both cases, we are dealing with a probabilistic phenomenon that needs to be studied, whether it is a string or a gas: the probability of a symbol occurring or the distribution of velocities. Perhaps my approach is too physical, or perhaps it is too abstractly mathematical, but the relationship is profound. Everything I have seen so far feels like dust blown against a speeding train, hoping to slow it down.

The issue concerns the properties of language: syntax and semantics, and it likely pertains more to semantics than syntax. This is because the rules for generating a symbol are known from the language, which includes syntax but not semantics. Essentially, every discipline studies the generation and transfer of information. This is where the problem lies. Hence, we encounter concepts such as graphs, processes, states, and Markov chains (whether discrete or continuous, finite or infinite). Hiding the problem behind functions named *enc*, *dec*, *sec*, or *conc* does not solve it: formalization is not solution.

And so, this is why entropy is a concept of immense power but also profound limitations. Its scope of application is closely tied to the context in which it originated. It is no coincidence that new disciplines have managed to reveal deep connections between seemingly distant phenomena by elevating entropy to something more. Statistical manifolds treat entropy in a somewhat concealed manner but discuss information and provide a metric treatment, modeling manifolds in parametric spaces that represent statistical distances. However, this distance (by the definition of the metric tensor, by the definition of the first fundamental form) is nothing more than the difference in entropy between two infinitesimally close points in space. It is still entropy, but viewed from a new perspective: probability distributions have become generators of manifolds, and differential entropy is the fundamental concept. Perhaps we should have, and maybe we still will, do the same.

3 When can a system be considered secure?

Let's focus on the central problem of this research in the context of cybersecurity: when can a system be considered secure? Communication is an exchange process, so it is clear that this exchange must occur only between the parties involved. Therefore, the system must ensure that communication does not extend beyond the physical, logical, spatial and temporal boundaries of the process, and that no other parties are involved except those specifically engaged in the communication process itself. If errors lead to the involvement of other parties, they must still be unable to understand the information they receive. If the process inadvertently sends information to unauthorized parties, these parties should not be able to comprehend the messages. If they can read and understand the messages, the process must prevent any modifications. More strongly, any reading process should irreversibly alter the message to make it possible to determine if the message has been read. If the process involves external parties, allows them to read and understand the message, the system must ensure that these external parties cannot reinsert the message into the system. In other words, a message modified or read by unauthorized individuals should not be reintegrated into the system (sent to other involved parties). Finally, if the message is sent to other parties, read, understood, modified, and reintegrated into the process, the future evolution of the system must remain unaffected.

Therefore, a system is considered secure if it simultaneously guarantees the following

- Involve: only the specified parties are involved;
- Read: the set of readers must at most be the set of involved;
- Understand: the set of comprehenders must be at most the readers;
- Modify: the set of modifiers must at most be the set of comprehenders;
- Resend: the set of resenders must at most be the set of modifiers;
- Evolve: the system must evolve at most with respect to resenders.

4 What happens now?

In fields closely related to computing and neural networks, objects such as reaction-diffusion equations are studied. These involve a system of four nonlinear differential equations whose solutions require techniques like Hopf bifurcation and Conley index theory from dynamic systems theory. A significant result of solving these equations is the manifestation of the "spike" phenomenon, observed years earlier as a sharp electrical impulse propagating along the neuron's axon. This intersection of neuroscience and biology has led to the development of artificial neural networks, marking a potential revolution in computing.

The concept of self-relation, where a system produces a message that affects and modifies itself, is directly related to differential equations. These

equations provide a mathematical expression of self-relation. Biological neural networks' self-organization presents a fundamental challenge in neuroscience, similar to its importance in computing and other fields of knowledge. Currently, no mathematical model adequately addresses this problem, capturing not only the individual behaviors of system components but also the emergent collective behaviors. In other words, the model must account for individual contributions and phenomena as well as emergent behaviors that are not immediately apparent but manifest as collective actions influencing the system as a whole.In other contexts more closely related to computer science and neural networks, objects like reaction-diffusion equations are studied. These are a system of four nonlinear differential equations whose solution requires techniques from Hopf bifurcation and the use of Conley index theory in dynamical systems theory. The most significant result from solving these equations is the manifestation of the spike phenomenon, observed many years prior as a sharp electrical impulse propagating along a neuron's axon. This intersection of neuroscience with biology has led to what might be one of the great revolutions in computer science after the computer: the development of artificial neural networks.

The aspect most directly related to self-relation (where a system produces a message that feedbacks to modify the system) involves differential equations—mathematical expressions of self-relation. The self-organization of biological neural networks is a major issue in neuroscience, as it is in computer science and likely across all areas of knowledge. Consequently, there is currently no satisfactory mathematical model that adequately addresses the problem, capturing both individual behaviors and collective emergent phenomena. The model must account not only for individual contributions and phenomena expressed by the actors but also for emergent behaviors that manifest as collective actions where individuals influence themselves, others, and the system.

Where should we go from here? Self-organization has found extensive investigation in statistical mechanics and dynamical systems theory. Statistical mechanics, building on Boltzmann's legacy, introduces concepts such as temperature, free energy, and entropy. Meanwhile, dynamical systems theory leads to more rigorous and comprehensive differential equations, which, though more complex to solve and simulate, are potentially more promising. Differential equations, by their nature, can encapsulate the complexity of the systems they describe due to their ability to represent self-relation and the associated quantities qualitatively. Statistical varieties, on the other hand, offer a probabilistic description and provide measures of these quantities quantitatively.