# Research: Human and Machine Learning

**Human and Machine Learning of Cyber-Strategies**

**Abstract**

As a costless communication channel with global reach, and as a distance-free space where every two nodes are neighbors, cyberspace enabled global distribution of data, content, and ideas. Originally, it gave grounds for hopes for global democracy. Thirty years later, cyberspace is also providing grounds for global threats to democracy, as it has been hijacked for psychological operations and influence campaigns, providing broad reach to narrow interests.

We claim that the novel cyber-attack methodologies can also be used as a grounding for cyber defenses. This claim requires thorough specification and careful validation, because the cyberattack methodologies have been developed through engineering practices, and lack a scientific theory that would display their controls and utility. We contend that the conceptual and technical components of the needed theory are available in different research disciplines, and need to be brought together in a transdisciplinary effort. In this note, we point to some of these components and try to describe the overarching paradigm.

**1 - Background: New directions in cybersecurity**

In the traditional theory of security, formalized in Shannon's "Mathematical theory", the attacker was omnipotent: if an information source contains some information, then Shannon's attacker will surely extract that information. In modern cryptography, initiated by Diffie and Hellman's "New directions", the attacker model has been refined by taking into account computational hardness: an information source may contain the information, but the attacker may run out of computational resources before that information is extracted.

Yet each of modern security definitions is a requirement that no feasible algorithm can solve some problem or another; i.e. that the problem is computationally hard. So although Diffie and Hellman's attackers do not have omnipotent computers anymore (like Shannon's attackers did), modern attackers are still tacitly assumed to have omnipotent programmers: if there is a feasible algorithm that breaks the system, then the attacker's programmers will find and implement that algorithm.

Such conservative assumptions are, of course, prudent, and even necessary if we are unable to assure that finding and implementing an algorithm is beyond the attacker's programmers' capabilities. There may indeed be cryptosystems that are vulnerable to computationally feasible attacks that will never happen because they are **cognitively unfeasible**, i.e. too complex to construct and launch. So we may be missing some systems that would be secure in practice, but look insecure due to the shortcomings of our theory. Modern cryptography is based on proofs of **computational hardness** of attacks, but we lack a corresponding theory of **cognitive hardness**.

**2 - Insight: Cybersecurity already follows a tacit new direction**

On the other hand, modern cryptography and most cybersecurity tools, are based on scores of *unproven* computational hardness assumptions. Some have been specially formulated for particular cryptosystems, like the hardness of the RSA problem, or of the CDH/DDH problem; some are more general, like the integer factoring and the discrete logarithm problems; and some are very general, like the P≠NP hypothesis, or the existence of one-way and trapdoor functions. Since life in cyberspace essentially depends on the validity of these assumptions, researchers and some strategists have been worrying about the possibility that some of these problems might turn out to be computationally easy (e.g. through advances in quantum computation). Users and most practitioners have, however, not been losing sleep. Even if there are **computationally feasible** algorithms for integer factoring or discrete logarithms, the experience of analyzing these problems suggests that such algorithms are **cognitively unfeasible**.

In practice, cybersecurity is thus already following a new direction, where the formal proofs of **computational hardness** of the foundations have been replaced by the empiric evidence of its **cognitive hardness**. Is this a risky strategy, which we use only because we have no other choice? Or is there an underlying practical justification, that theory has so far failed to grasp?

If the latter is the case, then our failure to subsume the engineering practice already deployed in the cyberspace under an adequate scientific theory of cyberspace may be the reason why ad hoc cyber-attacks remain easier to deploy than methodic and reliable cyber-defenses.

**3 - Approach: Modeling cognition and cognitive complexity of cyber-strategies**

Computation complexity measures and the cryptographic gap between computationally hard and easy functions are built upon the universal mathematical models of computation. The main obstacle to defining cognitive complexity measures, and to distinguishing cognitively hard and easy algorithms, has been the lack of a universal mathematical model of cognition, and of algorithm learning. (The mathematical theory of algorithmic complexity, and algorithmic information theory, have deep roots and a long history but have been developed with a theoretic focus on computability, leaving the questions of algorithm learning, and language pragmatics, largely out of scope.)

Recent advances in machine learning led to a leap ahead in natural language processing. The results surpassed the expectations even of the most optimistic experts. It also gave rise to new questions in modeling cognition in general, and in learning algorithms in particular. High-level structures of cognitive processes emerged, underlying both human and artificial cognition. (Cf. Causality and deceit: Do androids watch action movies? -- <https://arxiv.org/abs/1910.04383>)

**4 - Need**

Applications of machine learning to cybersecurity are in the focus of some of the main stakeholders of the cyber-battles. The industry giants, who employ some of the leaders in the field of machine learning, also provide the main avenues for new types of cyber-attacks. This presents a genuine threat to their business models, and a strong incentive for a genuine research push. So far, however, results have been meager, as crucial conceptual components seem to be missing. In any case, it is unlikely that cyberspace will remain insulated from advances in machine learning. It is possible that such advances may decide the outcome of many cyberspace confrontations, and of the new Cold War. Focused government-led research is necessary for a conceptual push on the broad front of cyber-strategy.