# PART 1

# AI and Network Security

# 1

# Intelligent Security of Computer Networks

**Abderrazaq SEMMOUD and Badr BENMAMMAR**
*Abou Bekr Belkaid University, Tlemcen, Algeria*

## 1.1. Introduction

Artificial intelligence (AI) and machine learning have rapidly progressed in recent years, facilitating the development of a broad range of applications. For example, AI is an essential component of widely used technologies such as automatic speech recognition, machine translation, spam filters and facial recognition. Promising technologies are currently the object of research or small-scale pilot projects, among which it is worth mentioning self-driving cars, digital assistants and drones activated by AI. Looking further into the future, advanced AI may reduce the need for human labor and improve governance quality.

A wide variety of tasks are automated using AI. Games, car driving and image classification are some of the tasks commonly studied by AI researchers. A broad set of tasks can be transformed by AI. At the very least, every task requiring human intelligence is a potential target for AI innovation. While the field of AI dates back to 1950, several years of rapid progress and growth have recently led to higher reliability. Sudden performance gains have been accomplished by researchers in a number of fields. Figure 1.1 illustrates this trend in the case of image recognition, where over the past few years AI systems have increased their performance in terms of classification accuracy from about 70% to nearly perfect classification accuracy (98%), which surpasses the human reference (95%) (Brundage *et al*. 2018).

*Intelligent Network Management and Control*,
coordinated by Badr BENMAMMAR. 

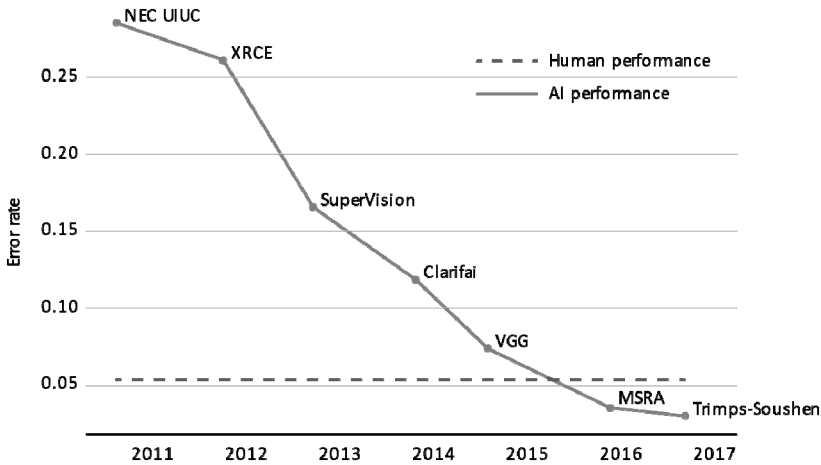


**Figure 1.1.** *Progress in image recognition (benchmark ImageNet), "Electronic Frontier Foundation's AI Progress Measurement" (August 2017)*

From a security perspective, a number of AI developments are worth mentioning. For example, target-face recognition and space navigation capacities are applicable to autonomous weapons systems. Similarly, image, text and voice generation possibilities could be used online to imitate other persons or influence public opinion by disseminating AI-generated content via social networks. These technical developments can also be considered early indicators of the potential of AI. Unsurprisingly, AI systems may soon qualify for an even wider range of security related tasks.

Information security is defined as the protection of computer systems against any unauthorized access, use, disruption, modification or destruction in order to provide confidentiality, integrity and availability (Peltier 2010). Information security does not refer to any particular security technology, but rather to a strategy involving persons, processes, rules and tools required in order to detect, prevent, document and mitigate current threats. With increasingly interconnected networks, security services are becoming ever more important. Connectivity is no longer an option in the commercial world, and its potential risks do not outweigh its advantages. Consequently, cybersecurity services should offer adequate protection to companies operating in a relatively open environment. Compared to classical approaches to computer security, several new hypotheses related to current computer networks should be formulated:

– modern networks are very large and further interconnected, and they are more accessible; consequently, potential attackers can easily connect and access these networks remotely;

– network interconnection increases the probability of attacks directed at large size networks such as the Internet by means of a set of widely known and open protocols.

The complexity of computer systems and applications is steadily growing. Consequently, it has become increasingly difficult to correctly analyze, secure and test computer system security. When these systems and their applications are connected to large networks, the risk of threats significantly increases. In view of adequate protection of computer networks, the deployed procedures and technologies must ensure (Khidzir *et al*. 2018):

– *confidentiality*: due to data confidentiality, only authorized users have access to sensitive information;

– *integrity*: due to data integrity, only authorized users can modify sensitive information; integrity could also ensure data authenticity;

– *availability*: due to system and data availability, authorized users have uninterrupted access to resources and important data.

The confidentiality, integrity and availability triad is a fundamental concept of information security. Each organization strives to ensure these three elements of the information system. Confidentiality prevents unauthorized disclosure of sensitive information (Kumar *et al*. 2018). Integrity prevents any unauthorized modification of information, thus ensuring information accuracy. Cryptographic hashing functions (such as SHA-1 or SHA-2) can be used to ensure data integrity. Availability prevents loss of access to resources and information (Kumar *et al*. 2018).

## 1.2. AI in the service of cybersecurity

AI systems are generally efficient, being less time and money-consuming than a human being when fulfilling a given task. AI systems are also evolutionary, as their computation power enables the completion of far more tasks in the same amount of time. For example, a typical facial recognition system is both efficient and evolutionary; once developed, it can be applied to numerous camera flows with a significantly lower cost than that of human analysts employed to perform a similar job. This explains why cybersecurity experts are seriously looking into AI and its potential contribution to mitigating certain problems. As an example, machine learning used by many AI algorithms can help detect malware, which are

increasingly difficult to identify and isolate due to their growing capacity to adapt to traditional security solutions (Veiga 2018).

Capgemini Research Institute has conducted a survey of 850 managers of seven large industrial companies: among the top management members included in this survey, 20% are information systems managers and 10% are responsible for information systems security. Companies headquartered in France, Germany, United Kingdom, the United States, Australia, India and Italy are mentioned in the report (Capgemini Research Institute 2019). Capgemini noted that, as digital companies develop, their cyberattack risk increases exponentially. It has been noted that 21% of companies declared one cybersecurity breach experience leading to unauthorized access in 2018. The price paid by companies for cybersecurity breaches is heavy (20% declared losses of over 50 million dollars). According to this survey, 69% of the companies estimate a need for AI to counteract cyberattacks. The majority of telecommunications companies (80%) declared that they relied on AI to identify the threats and counteract the attacks. According to the Capgemini report, the telecommunications sector declared the highest losses of over 50 million dollars, which led to AI being considered a priority in counteracting the costly breaches in this sector. Understandably, consumer goods sellers (78%) and banks (75%) came second and third, respectively, in this ranking, as these sectors increasingly rely on digital models. Companies based in the United States have as their top priority AI-based cybersecurity applications and platforms.

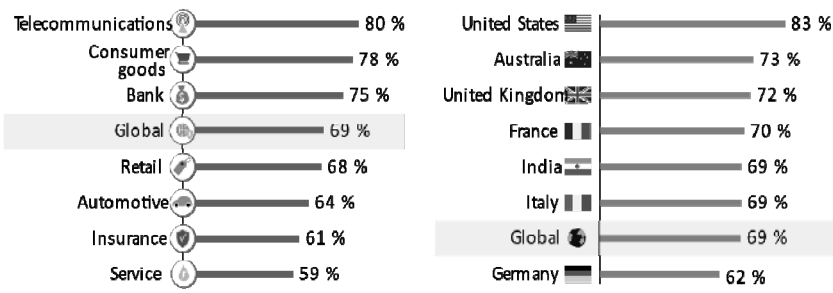


**Figure 1.2.** *Organizations and countries relying on artificial intelligence to identify threats and counteract attacks*

New vulnerabilities are discovered every day in the current programs, and these may infect and take control of a company's entire network. In contrast to traditional software vulnerabilities (for example, buffer memory overflow), the current intelligent systems have a certain number of vulnerabilities. This involves in particular data input causing errors in learning systems (Biggio *et al*. 2012), taking

advantage of the flaws in the design of autonomous systems' objectives (Amodei *et al*. 2016) and the use of inputs designed to falsify the classification of machine learning systems (Szegedy *et al*. 2013). As these vulnerabilities show, intelligent systems may outperform humans, but their potential failures are also unrivaled.

An ideal cyberdefense would offer full protection to users, while preserving system performances. Although this ideal cyberdefense may currently seem very distant, steps could be taken toward it by rendering cyberdefense more intelligent. The idea of using AI techniques in cybersecurity is not new. Landwehr (2008) states that, at their start, computer security and AI did not seem to have much in common. Researchers in the field of AI wanted computers to do by themselves what humans were able to do, whereas the researchers in the security field tried to solve the leakages in the computer systems, which they considered vulnerable. According to Schneier (2008), "The Internet is the most complex machine ever built. We barely understand how it works, not to mention how to secure it". Given the rapid multiplication of new web applications and the increasing use of wireless networks (Barth and Mitchell 2008) and the Internet of Things, cybersecurity has become the most complex threat to society.

The need for securing web applications against attacks (such as Cross Site Scripting [XSS], Cross Site Request Forgery [CSRF] and code injection) is increasingly obvious and pressing. Over time, XSS and CSRF scripts have been used to conduct various attacks. Some of them can be interpreted as direct bypasses of the original security policy. The same security policy was similar to a simple and efficient protection, but it turned out it could be easily bypassed and certain functionalities of modern websites could be blocked. According to Crockford (2015), the security policies adopted by most browsers "block useful contents and authorize dangerous contents". These policies are currently being reviewed. However, the detection of attacks such as XSS, CSRF or code injection requires more than a simple rule, namely a context-dependent reasoning capacity.

The use of AI in cybersecurity generally involves certain smart tools and their application to intrusion detection (Ahmad *et al*. 2016; Kalaivani *et al*. 2019) or other aspects of cybersecurity (Ahlan *et al*. 2015). This approach involves the use of other AI techniques developed for problems that are entirely different from cybersecurity; this may work in certain cases, but it has inherent and strict limitations. Cybersecurity has specific needs, and meeting them requires new specifically developed AI techniques. Obviously, AI has substantially evolved in certain fields, but there is still a need for learning and developing new intelligent techniques adapted to cybersecurity. In this context, according to Landwehr (2008) one "AI

branch related to computer security from its earliest age is automated reasoning, particularly when applied to programs and systems. Though the SATAN program of Dan Farmer and Wietse Venema, launched in 1995, has not yet been identified as AI, it has automated a process searching for vulnerabilities in system configurations that would require much more human efforts". Ingham *et al*. (2007) have proposed an inductive reasoning system for the protection of web applications. The works of Vigna and co-workers (Mutz *et al*. 2007; Cova *et al*. 2007, 2010; Kirdaa *et al*. 2009; Robertson *et al*. 2010) have also dealt with the protection of web applications against cyberattacks. Firewalls using deep packet inspections can be considered a sort of AI instantiation in cybersecurity. Firewalls have been part of the cyberdefense arsenal for many years. Although in most cases more sophisticated techniques (Mishra *et al*. 2011; Valentín and Malý 2014; Tekerek and Bay 2019) are also used, filtering relies on the port number. Firewalls cannot rely on the port number, as most web applications use the same port as the rest of the web traffic. Deep packet inspection is the only option enabling the identification of malware code in a legitimate application. The idea of application layer filtering of the Transmission Control Protocol/Internet Protocol (TCP/IP) model was introduced in the third generation of firewall in the 1990s. The modest success of these technologies is an indication that much more is still to be done in AI, so that it can make a significant difference in terms of cybersecurity. Nevertheless, it is worth noting that using AI in cybersecurity is not necessarily a miracle solution. For example, attacks without malware, which require no software download and dissimulate malware activities inside legitimate cloud computing services, are on the increase, and AI is not yet able to counteract these types of network breach.

### 1.3. AI applied to intrusion detection

Intrusion detection is defined as the process of intelligent monitoring of events occurring in a computer system or network and their analysis in search for signs of security policy breach (Bace 2000). The main objective of intrusion detection systems is to protect network availability, confidentiality and integrity. Intrusion detection systems are defined both by the method used to detect the attacks and by their location in the network. The intrusion detection system can be deployed as a network- or host-based system in order to detect the anomalies. Abusive use is detected based on the correspondence between known models of hostile activities and the database of previous attacks. These models are very effective for identifying known attacks and vulnerabilities, but less relevant in identifying new security threats. Anomaly detection looks for something rare or uncommon, applying statistical or intelligent measurements to compare the current activity to previous

knowledge. Intrusion detection systems rely on the fact that they often need many data for the artificial learning algorithms. They generally require more computer resources, as several metrics are often preserved and must be updated for each system activity (Ahmad *et al*. 2016). The intrusion detection expert system (IDES) (Lunt 1993) developed by Stanford Research Institute (SRI) formulates expert knowledge on the known models of attack and vulnerabilities of the system in the form of if–then rules. The time-based inductive machine (Teng and Chen 1990) learns several sequential models to ensure the detection of anomalies in a network. Several approaches using the artificial neural networks for intrusion detection systems have been proposed (Kang and Kang 2016; Kim *et al*. 2016; Vinayakumar *et al*. 2017; Hajimirzaei and Navimipour 2019). AI-based techniques are categorized in various classes (Mukkamala and Sung 2003a; Novikov *et al*. 2006).

### 1.3.1. *Techniques based on decision trees*

Decision trees are powerful and widespread nonparametric learning tools used for classification and prediction problems. Their purpose is to create a model that predicts the values of the target variable, relying on a set of sequences of decision rules deduced from learning data. Rai *et al*. (2016) have developed an algorithm based on the C4.5 decision tree approach. The most relevant characteristics are selected by means of information gain and the fractional value is selected so that it renders the classifier unbiased with respect to the most frequent values. In the work of Sahu and Babu (2015), a database referred to as ”Kyoto 2006+” is used for the experiments. In Kyoto 2006+, each instance is labeled as “normal” (no attack), “attack” (known attack) and “unknown attack”. The Decision Tree algorithm (J48) is used to classify the packets. Experiments confirm that the generated rules operate with 97.2% accuracy. Moon *et al.* (2017) proposed an intrusion detection system based on decision trees using packet behavior analysis to detect the attacks. Peng *et al.* (2018) proposed a technique that involves a preprocessing for data digitization, followed by their normalization, in order to improve detection efficiency. Then a method based on decision trees is used.

### 1.3.2. *Techniques based on data exploration*

Data exploration aims to eliminate the manual elements used for the design of intrusion detection systems. Various data exploration techniques have been developed and widely used. The main data exploration techniques are presented in the following sections.

#### 1.3.2.1. *Fuzzy logic*

Fuzzy logic has been used in the field of computer networks security, particularly for intrusion detection (Idris and Shanmugam 2005; Shanmugavadivu and Nagarajan 2011; Balan *et al*. 2015; Kudłacik *et al*. 2016; Sai Satyanarayana Reddy *et al*. 2019), for two main reasons. First, several quantitative parameters used in the context of intrusion detection, for example processor use time and connection interval, can be potentially considered as fuzzy variables. Second, the security concept is itself fuzzy. To put it differently, the fuzzy concept helps in preventing a sharp distinction between normal and abnormal behaviors. Kudłacik *et al.* (2016) have applied fuzzy logic for intrusion detection. The proposed solution analyzes the user activity over a relatively short period of time, creating a local user profile. A more in-depth analysis involves the creation of a more general structure based on a defined number of local user profiles, known as a "fuzzy profile". The fuzzy profile represents the behavior of the computer system user. Fuzzy profiles are directly used in order to detect user behavior anomalies, and therefore potential intrusions. Idris and Shanmugam (2005) proposed a modified FIRE system. It is a mechanism for the automation of the fuzzy rule generation process and the reduction of human intervention making use of AI techniques.

#### 1.3.2.2. *Genetic algorithms*

Genetic algorithms are techniques derived from genetics and natural evolution, which have been used to find approximate solutions to optimization and search problems. The main advantages of genetic algorithms are their flexibility and robustness as global search method. As for drawbacks, they are computationally time-consuming, as they handle several solutions simultaneously. Genetic algorithms have been used in various manners in the field of intrusion detection (Hoque *et al*. 2012; Aslahi-Shahri *et al*. 2016; Hamamoto *et al*. 2018). Hoque *et al.* (2012) presented an intrusion detection system using a genetic algorithm to effectively detect anomalies in the network. Aslahi-Shahri *et al.* (2016) proposed a hybrid method that uses support vector machines and genetic algorithms for intrusion detection. The results indicate that this algorithm can reach a 97.3% true positive rate and a 1.7% false positive rate.

### 1.3.3. *Rule-based techniques*

Rule-based techniques (Li *et al*. 2010; Yang *et al*. 2013) generally involve the application of a set of association rules for data classification. In this context, if a rule stipulates that *if event X occurs, then event Y is likely to occur*, events X and Y can be described as sets of pairs (*variable, value*). The advantage of using rules is that they tend to be simple and intuitive, unstructured and less rigid. Nevertheless, a

drawback is that rules are difficult to preserve and, in certain cases, inadequate for the representation of various types of information.

Turner *et al.* (2016) developed an algorithm for monitoring the enabled/disabled state of the rules of an intrusion detection system based on signatures. The algorithm is implemented in Python and runs on Snort (Roesch 1999). Agarwal and Joshi (2000) proposed a general framework in two stages for learning a rule-based model (PNrule) in order to learn classifier models on a set of data. They extensively used various distributions of classes in the learning data. The KDD Cups database was used for learning and testing their system.

### 1.3.4. *Machine learning-based techniques*

Machine learning can be defined as the capacity of a program to learn and improve the performances of a series of tasks in time. Machine learning techniques focus on the creation of a system model that improves its performances relying on the previous results. Furthermore, it can be said that machine learning–based systems have the capacity to handle the execution strategy depending on the new inputs. The main machine learning techniques are presented in the following sections.

#### 1.3.4.1. *Artificial neural networks*

Artificial neural networks learn to predict the behavior of various system users. If correctly designed and implemented, neural networks can potentially solve several problems encountered by rule-based approaches. The main advantage of neural networks is their tolerance to inaccurate data and uncertain information and their capacity to deduce solutions without previous knowledge on data regularities. Cunningham and Lippmann (2000) of MIT Lincoln Laboratory conducted a number of tests using neural networks. The system searched for attack-specific key words specific in the network traffic. In Ponkarthika and Saraswathy (2018), a model of intrusion detection system is explored as a function of deep learning. Long–short term memory (LSTM) architecture was applied to a recurrent neural network for the learning of an intrusion detection system using the KDD Cup 1999 dataset.

#### 1.3.4.2. *Bayesian networks*

A Bayesian network is a probabilistic graphical model representing a set of random variables in the form of an acyclic oriented graph. This technique is generally used for intrusion detection in combination with statistical diagrams. It has several advantages, notably the capacity to code the interdependences between variables and to predict events, as well as the possibility of integrating both previous knowledge and previous data (Heckerman 2008). Its major drawback is that results

are comparable to statistical techniques, but this requires additional computation efforts. Kruegel *et al.* (2003) proposed a multisensor fusion approach using a Bayesian network–based classifier for the classification and cancellation of false alarms, according to which the outputs of various sensors of the intrusion detection system are aggregated to generate a single alarm. Han *et al.* (2015) proposed an intrusion detection algorithm based on Bayesian networks relying on the analysis into main components. The authors calculate the characteristic data value of the attack on the original network, and then extract the main properties by analysis into main components.

#### 1.3.4.3. *Markov chains*

A Markov chain is a random process related to a finite number of states, with memoryless transition probabilities. During the learning phase, probabilities associated with transitions are estimated from the normal behavior of the target system. Detection of anomalies is then achieved by comparing the anomaly score obtained for the sequences observed at a fixed threshold. In the case of a hidden Markov model (Hu *et al*. 2009; Zegeye *et al*. 2018; Liang *et al*. 2019), the system we are interested in is assumed to be a Markov process in which states and transitions are masked. In the literature, several methods have been presented for solving the intrusion detection problem by inspecting the packet headers. Mahoney and Chan (2001) experimented with anomaly detection on DARPA network data by comparing the header fields of the network packet. Several systems use the Markov model for intrusion detection: PHAD (Packet Header Anomaly Detector) (Mahoney and Chan 2001), LERAD (Learning Rules for Anomaly Detection) (Mahoney and Chan 2002a) and ALAD (Application Layer Anomaly Detector) (Mahoney and Chan 2002b). In the book of Zegeye *et al.* (2018), an intrusion detection system using the hidden Markov model is proposed. The phase of network traffic analysis involves characteristic extraction techniques, reduction of dimensions and vector quantization, which plays an important role in large sets of data, as the amount of data transmitted increases every day. Model performances with respect to the KDD 99 dataset indicate an accuracy above 99%.

#### 1.3.4.4. *Support-vector machines*

The support-vector machine is a technique used for solving various learning, classification and prediction problems. The support-vector machine was employed in an implementation of the structural risk minimization (SRM) principle of Vapnik (1998), which minimizes the generalization error, in the sense of true error on unseen examples. The basic support-vector machine addresses problems with two classes, in which data are separated by a hyperplane defined by a certain number of support vectors. Support vectors are a subset of learning data serving to define the

limit between the two classes. When the support-vector machine cannot separate two classes, it solves this problem by mapping the input data in spaces of high-dimensional functions by means of a kernel function. In a high-dimensional space, it is possible to create a hyperplane enabling a linear separation (which corresponds to a curved surface in the lower input space). Consequently, the kernel function plays an important role in the support-vector machine. In practice, various kernel functions can be used, such as linear, polynomial, or Gaussian. A remarkable property of the support-vector machine is its learning capacity, which does not depend on the dimensionality of the characteristic space. This means that the support-vector machine can generalize when given numerous functionalities. Mukkamala and Sung (2003b) showed the many advantages of the support-vector machine compared to other techniques. Support-vector machines surpass neural networks in terms of upgradability, learning time, runtime and prediction accuracy. Mukkamala and Sung (2003a) also applied support-vector machines for the extraction of intrusion detection characteristics of KDD files. They empirically proved that the functionalities selected using the support-vector machine yielded similar results as the use of a full set of functionalities. This decrease in the number of functionalities reduces the computation efforts. Chen *et al.* (2005) also proved that support-vector machines surpassed neural networks.

### 1.3.5. *Clustering techniques*

Clustering techniques operate by organizing observed data in groups, depending on a given similarity or a distance measurement. Similarity can be measured by using the cosine formula, the binary weighted cosine formula proposed by Rawat (2005) or other formulas. The most commonly used procedure for clustering involves the selection of a representative point for each cluster. Then each new data point is classified as belonging to a given group depending on the proximity to the corresponding representative point. There are at least two approaches for the classification-based detection of anomalies. In the first approach, the anomaly detection model is formed using unlabeled data including both normal and attack traffic. In the second approach, the model is formed using only normal data and a normal activity profile is created. The idea underlying the first approach is that abnormal or attack data represent a small percentage of the total data. If this hypothesis is verified, anomalies and attacks can be detected depending on cluster size: large clusters correspond to normal data and the other data points to attacks. Liao and Vemuri (2002) used the *K-nearest neighbor* (K-nn) approach, based on the Euclidian distance, to define the belonging of data points to a given cluster. The Minnesota intrusion detection system is a network-based anomaly detection approach that uses data exploration and clustering techniques (Levent *et al*. 2004).

Leung and Leckie (2005) proposed an unsupervised anomaly detection approach for intrusion detection on a network. The proposed algorithm, known as "fpMAFIA", is a clustering algorithm based on density and on grid for large data sets. The major advantage of this algorithm is that it can produce arbitrary forms and cover over 95% of the set of data with appropriate values of parameters. The authors proved that the algorithm evolves linearly with respect to the number of registrations in the set of data. They evaluated the accuracy of the newly proposed algorithm and proved that it enables reaching a reasonable detection rate.

### 1.3.6. *Hybrid techniques*

Many researchers suggested that the monitoring capacity of current IDS systems could be improved by adopting a hybrid approach including detection techniques of both anomalies and signatures (Lunt *et al*. 1992; Anderson *et al*. 1995; Fortuna *et al*. 2002; Hwang *et al*. 2007). Sabhnani and Serpen (2003) proved that no single classification technique enables the detection of all the attack classes at an acceptable false alarm rate and with a good detection accuracy. The authors used various techniques to classify the intrusions by means of a KDD 1998 dataset. Many researchers proved that the hybrid or set-based classification technique can improve detection accuracy (Mukkamala *et al*. 2005; Chen *et al*. 2005; Aslahi-Shahri *et al*. 2016; Hamamoto *et al*. 2018; Hajimirzaei and Navimipour 2019; Sai Satyanarayana Reddy *et al*. 2019). A hybrid approach involves the integration of various learning or decision-making models. Each learning model operates differently and uses a different set of functionalities. The integration of various learning models yields better results than the individual learning or decision-making models and reduces their individual limitations. A significant advantage of the combination of redundant and complementary classification techniques is that it increases robustness and accuracy in most applications.

Various methods combining various classification techniques were proposed in the literature (Menahem *et al*. 2009; Witten *et al*. 2016). Ensemble methods have a common objective: to build a combination of certain models, instead of using a single model to improve the results. Mukkamala and its collaborators (2005) proved that the use of ensemble classifiers led to the best possible accuracy for each category of attack models. Chebrolu *et al.* (2005) used the Classification And Regression Trees-Bayesian network (CART-BN) approach for intrusion detection. Zainal *et al.* (2009) proposed the hybridization of linear genetic programming of the adaptive neural fuzzy inference system and of random forests for intrusion detection. They proved empirically that by assigning appropriate weights to the classifiers in a hybrid approach, the accuracy of detection of all the classes of network traffic is

improved compared to an individual classifier. Menahem *et al.* (2009) used various classifiers and tried to take advantage of their strengths. Hwang *et al*. (2007) proposed a three-level hybrid approach to detect intrusions. The first level of the system is a signature-based approach in order to filter the known attacks using the black list concept. The second level of the system is an anomaly detector that uses the white list concept to distinguish between the normal traffic and the attack traffic surpassed by the first level. The third level of the system uses support vectors machines in order to classify the unknown attack traffic. The success of a hybrid method depends on many factors, notably the size of the learning sample, the choice of a basic classifier, the exact manner in which the forming set is modified, the choice of combination method and finally the data distribution and the potential capacity of the basic classifier chosen for solving the problem (Rokach 2010).

## 1.4. AI misuse

AI is a double use domain. AI systems and the manner in which they are designed can serve both civilian and military purposes, and in a broader sense, beneficial or harmful purposes. Given that certain tasks requiring intelligence are benign while others are not, AI is double edged in the same way that human intelligence is. Researchers in the field of AI cannot avoid producing systems that can serve harmful purposes. For example, the difference between the capacities of an autonomous drone used for delivering parcels and the capacities of an autonomous drone used for delivering explosives is not necessarily too wide. Moreover, fundamental research aiming to improve our comprehension of AI, its capacities and its control seem to be inherently double edged.

AI and machine learning have an increasingly important impact on the security of citizens, organizations and states. Misuse of AI will impact the way in which we build and manage our digital infrastructure, as well as the design and distribution of AI systems, therefore it will probably require an institutional policy. It is worth noting here that the threats caused by AI misuse have been highlighted in heavily publicized contexts (for example, during a Congress hearing (Moore and Anderson 2012), a workshop organized by the White House and a report of the US Department for Homeland Security).

The increasing use of AI for the development of cyberattack techniques and the absence of development of adequate defenses has three major consequences.

### 1.4.1. *Extension of existing threats*

For many known attacks, the progress of AI is expected to enlarge the set of players capable of conducting the attack, their attack speed and the set of possible targets. This is a consequence of the efficiency, upgradability and ease of dissemination of AI systems. In particular, the dissemination of intelligent and efficient systems can increase the number of players who can afford specific attacks. If the reliable intelligent systems are also evolutionary (upgradable), then even the players who already have the required resources to conduct these attacks may acquire the capacity to execute them at a much faster pace.

An example of a threat that is susceptible to develop in this manner is the phishing attack threat. These attacks use personalized messages to obtain sensitive information or money from their victims. The attacker often introduces himself as one of the friends, colleagues or professional contacts of the target. The most advanced phishing attacks require significant qualified manpower, as the attacker must identify the high value targets, research their social and professional networks, and then generate messages that are acceptable to the target.

### 1.4.2. *Introduction of new threats*

AI progress will enable new varieties of attacks. These attacks may use AI systems to conduct certain tasks more successfully than any human being.

Due to their unlimited capacities, in contrast with those of humans, intelligent systems could enable players to conduct attacks that would otherwise be impossible. For example, most persons are not able to efficiently imitate the voice of other persons. Consequently, the creation of audio files resembling recordings of human speech becomes essential in these cases. Nevertheless, significant progress has been recently achieved in the development of speech synthesis systems, which learn to imitate human voice. Such systems would in turn enable new methods for spreading disinformation and imitating others.

Moreover, AI systems could be used to control certain aspects of malware behavior that would be impossible to control manually. For example, a virus designed to modify the behavior of ventilated computers, as in the case of the Stuxnet program, used to disrupt the Iranian nuclear program, cannot receive commands once these computers are infected. Limited communication problems also occur under water and in the presence of signal jammers.

### 1.4.3. *Modification of the typical threat character*

Properties of AI such as efficiency, upgradability and capacities surpassing those of humans may enable very relevant attacks. Attackers are often faced with a compromise between the frequency, the extent of their attacks and their efficiency. For example, spear phishing is more effective than classical phishing, which does not involve adapting messages to individuals, but it is relatively costly and cannot be conducted en mass. More generic phishing attacks are profitable despite their very low success rates, simply because of their extent. If the frequency and upgradability of certain attacks, including spear phishing, are improved, AI systems can mitigate these compromises. Moreover, properties such as efficiency and upgradability, particularly in the context of target identification and analysis, lead also to finely targeted attacks. The attackers are often interested in adapting their attacks to the characteristics of their targets, aiming at targets with certain properties, such as significant assets or an association with certain political groups. Nevertheless, the attackers must often find a balance between efficiency, the upgradability of their attacks and target precision. A further example could be the use of drone swarms that deploy facial recognition technology to kill specific individuals in a crowd, instead of less targeted forms of violence.

Cyberattacks are increasingly alarming in terms of complexity and quantity, a consequence of the lack of awareness and understanding of the actual needs. This lack of support explains the insufficient dynamism, attention and willingness to commit funds and resources for cybersecurity in many organizations. In order to limit the impact of cyberattacks, the following recommendations are suggested (Brundage *et al*. 2018):

– decision-makers should closely cooperate with technical researchers to study, prevent and limit the potential misuse of AI;

– researchers and engineers in the AI field should seriously consider the double-edged nature of their work, by allowing considerations linked to abusive use to influence the research priorities and norms and by proactively addressing concerned players when harmful applications are predictable;

– public authorities should actively try to broaden the range of stakeholders and experts in the field that are involved in the discussions related to these challenges.

## 1.5. Conclusion

AI is a broad domain to be explored by cybersecurity researchers and experts. As the capacity of intelligent systems increases, they will first reach and then surpass

human capacities in many fields. In cybersecurity, AI can be used to strengthen the defenses of computer infrastructure. It is worth noting that, as AI covers fields considered reserved to humans, the security threats will increase in variety, difference and intelligence compared to actually existing techniques. Defense against these threats is very difficult, as cybersecurity experts themselves can be targeted by spear phishing attacks. Consequently, preparing for potential misuses of AI associated with this transition is an important task. The use of intelligent techniques aims to identify real-time attacks, with little or no human interaction, and to stop them before they cause damages. In conclusion, AI can be considered as a powerful tool in solving cybersecurity problems.

## 1.6. References

Agarwal, R. and Joshi, M.V. (2000). A new framework for learning classifier models in data mining [Online]. Available at: https://pdfs.semanticscholar.org/db6e/1d67f7912efa65f94807dc81b24dea2de158.pdf [Accessed January 2019].

Ahlan, A.R., Lubis, M., and Lubis, A.R. (2015). Information security awareness at the knowledge-based institution: Its antecedents and measures. *Procedia Computer Science (PCS)*. 72(2015), 361–373.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané. (2016). Concrete problems in AI safety [Online]. Cornell University. Available at: https://arxiv.org/abs/1606.06565.

Anderson, D., Frivold, T., and Valdes, A. (1995). Next-generation intrusion detection expert system (NIDES). Report, US Department of the Navy, Space and Naval Warfare Systems Command, San Diego.

Aslahi-Shahri, B.M., Rahmani, R., Chizari, M., Maralani, A., Eslami, M., Golkar, M.J., and Ebrahimi, A. (2016). A hybrid method consisting of GA and SVM for intrusion detection system. *Neural Computing and Applications*, 27(6), 1669–1676.

Bace, R.G. (2000). *Intrusion Detection*. Sams Publishing, Indianapolis.

Balan, E.V., Priyan, M.K., Gokulnath, C., and Devi, G.U. (2015). Fuzzy based intrusion detection systems in MANET. *Procedia Computer Science*, 50, 109–114.

Barth, C.J. and Mitchell, J.C. (2008). Robust defenses for cross-site request forgery. *Proceedings of 15th ACM Conference*. CCS, Alexandria.

Biggio, B., Nelson, B., and Laskov, P. (2012). Poisoning attacks against support vector machines. *29th International Conference on Machine Learning*. ICML, Edinburgh, 1467–1474.

Capgemini Research Institute (2019). Reinventing cybersecurity with artificial intelligence: The new frontier in digital security [Online]. Available at: https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_2019 0711_V06.pdf.

Chebrolu, S., Abraham, A., and Thomas. (2005). Feature deduction and ensemble design of intrusion detection systems. *Computers & Security*, 24(4), 295–307.

Chen, W.-H., Hsu, S.-H., and Shen, H.-P. (2005). Application of SVM and ANN for intrusion detection. *Computers & Operations Research*, 32(10), 2617–2634.

Cova, M., Balzarotti, D., Felmetsger, V., and Vigna, G. (2007). Swaddler: An approach for the anomaly-based detection of state violations in web applications. *Proceedings of the 10th International Symposium on Recent Advances in Intrusion Detection*. RAID, Gold Coast.

Cova, M., Kruegel, C., and Vigna, G. (2010). Detection and analysis of drive-by-download attacks and malicious JavaScript code. *Proceedings of the 19th International Conference on the World Wide Web*. WWW, Raleigh.

Crockford, D. (2015). Json [Online]. Available at: https://github.com/douglascrockford/JSON-js/blob/master/README [Accessed March 2018].

Cunningham, R. and Lippmann, R. (2000). Detecting computer attackers: Recognizing patterns of malicious stealthy behavior. Presentation, CERIAS, Anderlecht.

Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P.N., Kumar, V., Srivastava, J., & Dokas, P. (2004). Minds-Minnesota intrusion detection system. *Next Generation Data Mining*, August, 199–218.

Fortuna, C., Fortuna, B., and Mohorčič, M. (2002). Anomaly detection in computer networks using linear SVMs [Online]. Available at: http://ailab.ijs.si/dunja/SiKDD2007/Papers/Fortuna_Anomaly.pdf.

Hajimirzaei, B. and Navimipour, N.J. (2019). Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. *ICT Express*, 5(1), 56–59.

Hamamoto, A.H., Carvalho, L.F., Sampaio, L.D.H., Abrão, T., & Proença Jr, M.L. (2018). Network anomaly detection system using genetic algorithm and fuzzy logic. *Expert Systems with Applications*, 92, 390–402.

Han, X., Xu, L., Ren, M., and Gu, W. (2015). A Naive Bayesian network intrusion detection algorithm based on principal component analysis. *7th International Conference on Information Technology in Medicine and Education*. IEEE, Huangshan.

Heckerman, D. (2008). A tutorial on learning with Bayesian networks. *Innovations in Bayesian Networks*, Holmes, D.E. and Jain, L.C. (eds). Springer, Berlin, 33–82.

Hoque, M.S. *et al.* (2012). An implementation of intrusion detection system using genetic algorithm. *International Journal of Network Security & Its Applications (IJNSA)*. AIRCC publisher, 4(2), 109–120.

Hu, J., Yu, X., Qiu, D., and Chen, H.H. (2009). A simple and efficient hidden Markov model scheme for host-based anomaly intrusion detection. *IEEE Network*, 23(1), 42–47.

Hwang, T.S., Lee, T.-J., and Lee, Y.-J. (2007). A three-tier IDS via data mining approach. *Proceedings of the 3rd Annual ACM Workshop on Mining Network Data*. ACM, San Diego.

Idris, N.B. and Shanmugam, B. (2005). Artificial intelligence techniques applied to intrusion detection. *Annual IEEE India Conference (Indicon)*. IEEE, Chennai.

Ingham, K., Somayaji, A., Burge, J., and Forrest, S. (2007). Learning DFA representations of HTTP for protecting web applications. *Journal of Computer Networks*, 51(5), 1239–1255.

Kalaivani, S., Vikram, A., and Gopinath, G. (2019). An effective swarm optimization based intrusion detection classifier system for cloud computing. *5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, Coimbatore.

Kang, M.-J. and Kang, J.-W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. *PLOS ONE*, 11(6), 1–17.

Khidzir, N.Z., Daud, K.A.M., Ismail, A.R., Ghani, M.S.A.A., and Ibrahim, M.A.H. (2018). Information Security Requirement: The Relationship Between Cybersecurity Risk Confidentiality, Integrity and Availability in Digital Social Media. *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*. 4–6 December 2016, Penang, Malaysia.

Kim, J., Kim, J., Thu, H.L.T., and Kim, H. (2016). Long short term memory recurrent neural network classifier for intrusion detection. *International Conference on Platform Technology and Service (PlatCon)*. IEEE, Jeju.

Kirdaa, E., Jovanovicb, N., Kruegelc, C., and Vigna, G. (2009). Client-side cross-site scripting protection. *Computers & Security*, 28(7), 592–604.

Kruegel, C., Mutz, D., Robertson, W., and Valeur, F. (2003). Bayesian event classification for intrusion detection. *Proceedings of the 19th Annual Computer Security Applications Conference*. IEEE, Las Vegas.

Kudłacik, P., Porwik, P., and Wesołowski, T. (2016). Fuzzy approach for intrusion detection based on user's commands. *Soft Computing*, 20(7), 2705–2719.

Kumar, S., Krishna, C.R., and Solanki, A.K. (2018). A technique to resolve data integrity and confidentiality issues in a wireless sensor network. *8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, Noida.

Landwehr, C. (2008). Cybersecurity and artificial intelligence: From fixing the plumbing to smart water. *IEEE, Security and Privacy*, 6(5), 3–4.

Leung, K. and Leckie, C. (2005). Unsupervised anomaly detection in network intrusion detection using clusters. *Proceedings of the 28th Australasian Conference on Computer Science*. Australian Computer Society Inc., Darlinghurst, 333–342.

Li, L., De-Zhang, Y. and Chen, F.-S. (2010). A novel rule-based Intrusion Detection System using data mining. *3rd International Conference on Computer Science and Information Technology*. IEEE, Chengdu.

Liang, J. *et al*. (2019). A filter model for intrusion detection system in vehicle ad hoc networks: A hidden Markov methodology. *Knowledge-Based Systems*, 163, 611–623.

Liao, Y. and Vemuri, V.R. (2002). Use of k-nearest neighbor classifier for intrusion detection. *Computers & Security*, 21(5), 439–448.

Lippmann, R.P. and Cunningham, R.K. (2000). Improving intrusion detection performance using keyword selection and neural networks. *Computer Networks*, 34(4), 597–603.

Lunt, T. (1993). Detecting intruders in computer systems. *Proceedings of the 1993 Conference on Auditing and Computer Technology*. Baltimore Convention Center, Baltimore.

Lunt, T.F. (1990). Real-time intrusion detection expert system. Computer Science Lab., SRI International, Technical Report.

Mahoney, M.V. and Chan, P.K. (2001). PHAD: Packet header anomaly detection for identifying hostile network traffic [Online]. Available at: https://pdfs.semanticscholar.org/1505/f3658f5af7dff88e88d6a2b381de12e03036.pdf.

Mahoney, M.V. and Chan, P.K. (2002a). Learning models of network traffic for detecting novel attacks. Technical Report, Florida Institute of Technology, Melbourne.

Mahoney, M.V. and Chan, P.K. (2002b). Learning nonstationary models of normal network traffic for detecting novel attacks. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Edmonton.

Menahem, E., Shabtai, A., Rokach, L. and Elovici, Y. (2009). Improving malware detection by applying multi-inducer ensemble. *Computational Statistics & Data Analysis*, 53(4), 1483–1494.

Miles, B., Shahar, A., Jack, C., Helen, T., Peter, E., Ben, G., Allan, D., Paul, S., Thomas, Z., Bobby, F., Hyrum, A., Heather, R., Gregory, C.A., Jacob, S., Carrick, F., Seán, Ó. h., Simon, B., Haydn, B., Sebastian, F., Clare, L., Rebecca, C., Owain, E., Michael, P., Joanna, B., Roman, Y. and Dario, A. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation [Online]. Available at: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.

Mishra, A., Agrawal, A. and Ranjan, R. (2011). Artificial intelligent firewall. *Proceedings of the International Conference on Advances in Computing and Artificial Intelligence*. ACM, Rajpura/Punjab.

Moon, D., Im, H., Kim, I. and Park, J. H. (2017). DTB-IDS: An intrusion detection system based on decision tree using behavior analysis for preventing APT attacks. *The Journal of Supercomputing*, 73(7), 2881–2895.

Moore, T. and Anderson, R. (2012). *Internet Security. The Oxford Handbook of the Digital Economy*. Oxford University Press, Oxford.

Mukkamala, S. and Sung, A.H. (2003a). Artificial intelligent techniques for intrusion detection. *International Conference on Systems, Man and Cybernetics*. IEEE, Washington.

Mukkamala, S. and Sung, A.H. (2003b). A comparative study of techniques for intrusion detection. *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03)*. IEEE, Washington.

Mukkamala, S., Sung, A.H., and Abraham, A. (2005). Intrusion detection using an ensemble of intelligent paradigms. *Journal of Network and Computer Applications*, 28, 167–182.

Mutz, D., Robertson, W., Vigna, G., and Kemmerer, R. (2007). Exploiting execution context for the detection of anomalous system calls. *Proceedings of the International Symposium on Recent Advances in Intrusion Detection*. RAID, Gold Coast.

Novikov, D., Yampolskiy, R.V., and Reznik, L. (2006). Artificial intelligence approaches for intrusion detection. *IEEE Long Island Systems, Applications and Technology Conference*. IEEE, Long Island.

Peltier, T.R. (2010). *Information Security Risk Analysis*. CRC Press, Boca Raton.

Peng, K., Leung, V., Zheng, L., Wang, S., Huang, C., and Lin, T. (2018). Intrusion detection system based on decision tree over big data in fog environment [Online]. Available at: https://www.hindawi.com/journals/wcmc/2018/ 4680867/.

Ponkarthika, M. and Saraswathy, V.R. (2018). Network intrusion detection using deep neural networks. *Asian Journal of Applied Sciences*, 2(2), 665–673.

Quamar, N., Weiqing, S., Ahmad, Y.J., and Mansoor, A. (2016). A deep learning approach for network intrusion detection system. *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*. ICST publisher, December 3–5, 2015, New York, USA, 21–26.

Rai, K., Devi, M.S., and Guleria, A. (2016). Decision tree based algorithm for intrusion detection. *International Journal of Advanced Networking and Applications*, 7(4), 2828.

Rawat, S. (2005). Efficient data mining algorithms for intrusion detection. *Proceedings of the 4th Conference on Engineering of Intelligent Systems (EIS 2004)*. EIS, Madeira.

Robertson, W., Maggi, F., Kruegel, C., and Vigna, G. (2010). Effective anomaly detection with scarce training data. *Proceedings of the Network and Distributed System Security Symposium*, NDSS, San Diego.

Roesch, M. (1999). Snort: Lightweight intrusion detection for networks. *Lisa*, 99(1), 229–238.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1/2), 1–39.

Sabhnani, M. and Serpen, G. (2003). Application of machine learning algorithms to KDD intrusion detection dataset within misuse detection context. *International Conference on Machine Learning; Models, Technologies and Applications*. MLMTA, Las Vegas.

Sahu, S. and Mehtre, B.M. (2015). Network intrusion detection system using J48 Decision Tree. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, Kochi.

Sai Satyanarayana Reddy, S., Chatterjee, P., and Mamatha, C. (2019). Intrusion detection in wireless network using fuzzy logic implemented with genetic algorithm. In *Computing and Network Sustainability*, Peng, S.-L, Dey, N., and Bundele, M. (eds). Springer, Berlin, 425–432.

Scharre, P. (2015). *Counter-swarm: A guide to defeating robotic swarms* [Online]. Available at: https://warontherocks.com/2015/03/counter-swarm-a-guide-todefeating-robotic-swarms/.

Schneier, B. (2008). The psychology of security. *International Conference on Cryptology in Africa.* AFRICACRYPT, Casablanca.

Shanmugavadivu, R. and Nagarajan, N. (2011). Network intrusion detection system using fuzzy logic. *Indian Journal of Computer Science and Engineering*, 2(1), 101–111.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. and Fergus, R. (2013). Intriguing properties of neural networks [Online]. Available at: https://arxiv.org/abs/1312.6199.

Tekerek, A. and Bay, O.F. (2019). Design and implementation of an artificial intelligence-based web application firewall model. *Neural Network World*, 189, 206.

Teng, H.S. and Chen, K. (1990). Adaptive real-time anomaly detection using inductively generated sequential patterns. *Proceedings of the 1990 IEEE Computer Society Symposium on Research in Security and Privacy*. IEEE, Oakland.

Turner, C., Jeremiah, R., Richards, D., and Joseph, A. (2016). A rule status monitoring algorithm for rule-based intrusion detection and prevention systems. *Procedia Computer Science*, 95, 361–368.

Valentín, K. and Malý, M. (2014). Network firewall using artificial neural networks. *Computing and Informatics*, 32(6), 1312–1327.

Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley and Sons, Hoboken.

Veiga, A.P. (2018). Applications of artificial intelligence to network security [Online]. Available at: https://arxiv.org/abs/1803.09992.

Vinayakumar, R., Soman, K.P. and Poornachandran, P. (2017). Applying convolutional neural network for network intrusion detection. *6th International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Manipal University, Karnataka.

Witten, I.H. and Frank, E. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Burlington.

Yang, Y., McLaughlin, K., Littler, T., Sezer, S. and Wang, H.F. (2013). Rule-based intrusion detection system for SCADA networks. *2nd IET Renewable Power Generation Conference (RPG 2013)*. RPG, Beijing.

Zainal, A., Maarof, M.A. and Shamsuddin, S.M. (2009). Ensemble classifiers for network intrusion detection system. *Journal of Information Assurance and Security*, 4(3), 217–225.

Zegeye, W.K., Moazzami, F. and Dean, R. (2018). Hidden Markov Model (HMM) based Intrusion Detection System (IDS). *International Telemetering Conference Proceedings*, 5–8 November 2018, Glendale, Arizona.