A Survey of Cyber Security Systems: Approaches for Attack Detection, Prediction and Prevention

Jorge Gonçalves
Lusofona University of Porto, Portugal
jorge.oliveira.goncalves@outlook.pt

Hugo Barbosa
Lusofona University of Porto, Portugal
hugo.barbosa@ulp.pt

Abstract. Nowadays the technological advancements that are done daily grow exponentially consequently leading to discoveries being more relevant to improve all the areas that make use of technologies. The purpose of this paper is to centralize some of that progress that can make an improvement in networks through Cyber Security and present the results from an inquiry to a population sample about Cyber Security. It is briefly presented approaches in attack detection, prediction, and prevention. The attack detection topic will be presented on what type of systems exists nowadays, focusing on Intrusion Detection Systems (IDSs). In attack prediction topic will be presented solutions to keep up with the appearance of new forms of attack, allowing to be prepared for then. As for attack prevention, it is done a summary of Social Engineering and Good practice in Cyber Security. It is an essential part of this work an inquire analysis performed in Portugal, prepared in the context of Cyber Security, targeting common users.

Keywords: Cyber Security, Attack Detection, Attack Prediction, Attack Prevention, Good practice, Survey, Inquire.

1 Introduction

Having a network that is able to handle attacks is a must nowadays. To accomplish this, it is crucial to design the network in a way that it is prepared on all fronts. Attack detection, attack prediction, and attack prevention.

This paper's objective it is not to detail every topic under Cyber Security with great detail, but to provide some context information about what exist these days that can enhance security in itself, while also being beginner friendly.

The Cyber Security topic will discuss three sub-topics. In the sub-topic for Attack Detection is presented briefly what systems exist and will be conveyed a summary on Intrusion Detection Systems (IDSs). In the Attack Prediction sub-topic, are going to be reviewed four methods to predict attacks and new forms of attack, those being – Vulnerability Databases, Markov Models, Bayesian Network and Awareness using Twitter. While for Attack Prevention, it will be lightly summarized what Social Engineering is and some good practices to prevent attacks.

10.11228/dpsc.02.01.002

Lastly, is introduced a topic that presents an Inquiry Analysis done in Portugal, which was created with the main focus of gathering data from a population sample regarding their habits on their use of the Internet, the state of cultural knowledge related to cyber security determine how well is the population in general evolving and expose this data to the academic community while serving as a foundation to future related works.

2 Cyber Security Systems

Cyber Security by itself it is the form of protecting computerized network systems against all sorts of attack that gain unauthorized access and may damage the integrity of the system by stealing, blocking access or deleting confidential(private) data or simply by executing dubious tasks. It is ultimately the structuring of systems that are intended to defend a network.

This topic discusses three components of Cyber Security. Attack Detection, Attack Prediction, and Attack Prevention.

2.1 Attack Detection

Even with every prevention and prediction available, it is not always possible to be prepared for all the new techniques of attack, so it is essential to have some sort of defense against it. Having a system or tool that monitories every communication, in and out of the network, and even inside of itself, will lead to an increase in the safety of all the network, ultimately creating a more sustainable system. We highlight the following ones:

- Intrusion Detection Systems.[1]
- Firewalls.
- Anti-virus.

Intrusion Detection Systems

“An intrusion detection system is an application that provides protection from malicious activities or policy violations and generates various rules to defend computer security and this system is relevant for intrusion detection.”[2] Continuous monitorization and application of policies, rules, and verification if and attack signatures if present, require some moderate computational power to maintain this system operational. This leads to the main function of IDSs, which is to warn of suspicious activity is taking place, but it is not its job to prevent it.

Since the use of the Internet of Things (IoT) is growing exponential every day, and systems are more than ever dependent on then, these systems represent a large portion of devices in networks, but with the low computational power that the majority dispose of, they cannot spare processing time. It is crucial to implement new techniques to

protect these devices. Recurring to software as a service seems to be one solution for this challenge[3].

In a more generic view, an IDS monitors and evaluates a suspected intrusion once it has taken place, checking its database to detect attack standards. This monitorization is made from both attacks originated from outside and within the systems. Although an IDS monitors and evaluates intrusions, it should be thought of as a replacement for neither a firewall nor a good antivirus program, but instead, as a complement to increase the security of the system, working in pairs with these resources.

There are several techniques to implement intrusion detection systems (IDS), which leads to the following variants.

Variants. [1][4]

- Statistics-based.
- Pattern-based.
- Rule-based.
- State-based.
- Heuristic-based – Machine Learning, Artificial Intelligence, Data Mining.

Choosing the right variant to implementation for each system depends heavily on what constrains that systems may have, so careful planning is required.

2.2 Attack Prediction

Throughout the year there have been countless researchers aiming to find methods to discover what new tools attackers may have developed so that countermeasures can be taken. This section of the paper is focused on presenting information about systems and methods that can be used as tools so that countermeasures to attacks can be taken. If successfully implemented, these systems can lead to gaining a reasonable amount of time so that defensive actions can be made to prevent malicious effects on the networks.

Beneath, is described briefly some of these methods and systems that can help increase considerably how predictions are made towards new manners of attack.

2.2.1 Existing Methods

Vulnerabilities databases.

One of the tools that can be used to gather data, that is going to be inputted into the next presented methods (e.g. through data-mining), are, National Vulnerabilities Database (NVD) which is a U.S. government repository of standards-based vulnerabilities and Common Vulnerabilities and Exposures (CVE) is a list of entries, from is a U.S.

government division, meant to provide reference from publicly known Cyber Security vulnerabilities.[5]

These databases can provide useful data that can be used to train models, Artificial Intelligence programs and Machin Learning Algorithms.

Markov models.

A Markov model is a stochastic model (mathematical object, in which random variables are correlated with or indexed by a set of numbers), that is used to model arbitrarily shifting systems. One premise is that future states depend only on the current state, not relating to events that occurred before it (this being Markov property), they are memoryless. Usually, these models are represented in graphs, allowing for better visual representation and understanding of the problem.[5] Depending on the type of problem or system we have, different Markov models can be employed, which are categorized into four common uses, differing whether every successive state is observable or not and whether the system is to be corrected based on observations made:

Table 1. Different Markov models

	System state is fully observable	System State is partially observable
System is autonomous	Markov chain	Hidden Markov model
System if controlled	Markov decision process	Partially observable Markov decision process

There have been some successful studies related to Cyber Security, where they used Markov models to predict cyber-attacks, proving that these models, even though they were developed a long time ago, they still are a useful tool.[6][7]

Bayesian Network.

Developed in early 1980, Bayesian Networks consists of models that represent knowledge in a form of graphs allowing to reason some conclusion for both discreet or continuous problems, that are based on uncertainties. They are also known as opinion networks, casual networks, and probabilistic dependency graphs.

Throughout the years they have been used to help develop conclusion in many studies, even in Cyber Security prediction, in topics like:

- Combine different sources of Knowledge.[8]
- Models for motivation and psychology of malicious insiders.[9]
- Calculation of the probability of cyber-attacks towards a specific target.[10]
- Prediction of data breaches.[11]

Awareness using Twitter.

With the rapid pace that developments are made, and new forms of attack are discovered, associated with the fact that databases like NVC and CVE have to take their time to ensure that the data that they introduce into their databases is not a false positive, it was encounter a new form of anticipating the appearance of new entries into those databases throw the use of social media platform, Twitter.[12] [13] By analyzing tweets from reliable accounts, that work and investigate malicious software attacks exploitation, it was developed a system that, through the use of Twitter API Stream, can collect live tweets from Twitter and feed then directly to Synapse[13], an threat intelligence program which uses machine learning techniques, allowing it to track the arrival of new exploits mentions before they are introduced to vulnerabilities databases.

Although there are already some threat intelligence systems that can collect data from a wide variety of sources[14], they simply use a keyword filter that restricts the volume of collected information and does not have any sophisticated procedure to select only the relevant data. Synapse was designed to collect data from various sources on the internet, classify it was Cyber Security related content, and aggregate all relevant tweets though a stream clustering algorithm adapted to the context of Cyber Security.[7]

2.3 Attack Prevention

To complement a good infrastructure, it is required that all possible actions that can be done to avoid major damage are implemented, it is here that comes Attack Prevention. This topic will be present briefly two subjects that can, in fact, increase system security by preventing a certain chain of actions.

Social Engineering.

"In a cyber security context, it is primarily used to induce victims towards disclosing confidential data, or to perform actions that breach security protocols, unknowingly infecting systems or releasing classified information."[15] As mentioned in this quote from Breda F., Barbosa H., Morais T., it is the act of misleading a system user to, unwillingly, compromise the integrity of a system.

Good Practice for Prevention.

From the review and mention, literature was possible to determine some good practice that can significantly increase a system simply by taking preventive measures.[16][17]

- Use of a proper Data Recovery System.
- Keep your systems up to date.
- Instruct your users with basics about Cyber Security.[18]
- Have polices to manage emails, filtering attachments, and potentially misleading sources.

- Have polices to renew passwords, even if they are already considered strong.
- Having Two Factor Authentication for systems that are critical.
- Always report crimes to the cyber fraud complaint center in your counter.
- Among many others.

3 Survey Analysis in Portugal

With the expectation to determine if the common user as evolved alongside the developments in Cyber Security, it was conducted an online survey aiming to get a good population sample from Portugal, with no costs implied, allowing to investigate how well population, in general, is evolving their knowledge regarding Cyber Security and if they are implementing it.[19]

To create and manage the survey it was used Google Forms which allowed the participants to quickly answer and submit their answers anonymously, without the need to use their Google account to authenticate, this way, in a hope, increasing the possible number of participants, since nowadays no one was time. It was also chosen Google Form because of its tools to present and analyze the data extracted from the questions quickly.

On the header of the survey, was given a quick context for why it was created, being that was in the scope of an evaluation for a course unit name Network Complements being lectured in the Lusofona University of Porto. It was also mentioned that the survey was targeted to the public in general that doesn't have responsibilities in managing a computer park, even though it was possible to then to answer, we relied on the professional ethic in a sense that they did not answer. The header also mentioned that the data was going to analyze and was going to be present in a paper about Cyber Security.

In total, there were 258 submissions(participants), being 48.1% (124 of 258) male and 51.9% (134 of 258) female. As for age groups, for ages between 14 and 17 there were 15.1% (39 of 258), from 18 to 25 there were 41.1% (106 of 258), as between 26 and 40 there were 26% (67 of 258) participants, for ages comprehended from 40 to 65 there were 15.1% (39 of 258) and lastly, for the group for participants with 65 and more, it was only possible to get 2.7% (7 of 258).

As for academic abilities, only 0.4% (1 of 258) participants had the fourth grade, for the ninth grade 26.8% (69 of 258), as for the academic level of twelfth grade, there were 36.2% (93 of 258) participants. For higher school education it was possible to get answers from 94 participants (36.6%), from which 24.5% (63 of 258) reported to have Bachelor's degree, 10.9% (28 of 258) participants having Master's degree and 1.2% (3 of 258) Ph.D. Internationally, this academic levels should be equivalent to:

Academic Level in Portugal	Academic Level Internationally
Fourth grade	elementary school
Ninth grade	freshman year
twelfth grade	senior year

This survey was formulated to have two parts not perceptible to the participant. The first part was intended to investigate if the common user, throughout the years with the awareness available online, had begun to implement some of the basics of Cyber Security, like reusing password and not using special characters. For the second part, while still getting some data to evaluate, the main intention was to sub linearly spread some awareness to the participants about some technical terms that sometimes are misunderstood by the common user, in this manner, the options to the answers were quite obvious.

3.1 Population Analysis

Like mentioned above, the survey was intended to be divided into two parts. The first part, focused in this subtopic, was planned to investigate if the population in general has begun to develop any awareness regarding Cyber Security, and if so, if they had started implementing it in their lives, as part of their privacy. Underneath can be observed two tables. One that shows questions, with only yes or no answers, and another table that represents the answers that were provided where the question had more than one choice, with the corresponding percentiles.

From point forward, the questions will be referred to just by their identifier, to increase the comprehension and readability.

Table 2. Question from the survey with yes or no answers related to the first part

Identifier	Question	Yes	No
1	Do you have formation in Cyber Security?	11.2% (29 of 258)	88.8% (229 of 258)
2	Do you know the term Cyber Security?	81.8% (211 of 258)	18.2% (47 of 258)
3	Do you reuse your passwords to do logins?	80.6% (208 of 258)	19.4% (50 of 258)
4	Do you include special characters in your passwords? (Example: ?=”!#)	56.2% (145 of 258)	43.8% (113 of 258)

Table 3. Question from the survey with multiple choices related to the first part

Identifier	Question	Work	Entertainment	Work and entertainment	Don't use
5	In what context do you use computers and/or smartphones?	2.3% (6 of 258)	11.6% (30 of 258)	84.5% (218 of 258)	1.6% (4 of 258)

For starters, questions 1 and 2 were made to understand how familiar participants were regarding some of the basics in Cyber Security and it was positive to see that, even though only 11.2% (29 of 258) had formation in this area, a larger portion of then, 81.8% (211 of 258), affirmed knowing the term Cyber Security. Although knowing the term doesn't mean that they, in reality, know anything about it, analysing the result from question 4, 56.2% (145 of 258) yes and question 3, 19.4% (50 of 258) no, can be deducted that some of the basics principals, the ones asked in those questions, are being applied.

The reuse of passwords, question 3, like mentioned by Don Norman in one of his books [20], Chapter 3 - Memory Is Knowledge in the Head, is something that even security professionals admit to do, but comes with a great risk since having one password compromised can lead to the loss of privacy and integrity in multiple systems. This is even a greater threat when corelated with question 5. Assuming that companies leave to the user the choice of their passwords, this can lead to then reusing one of their passwords which can already be compromised, conducting to introducing a vulnerability in the company network. This is special alarming when in 84.5% (218 of 258) participants affirmed using computers and smartphones for both work and entertainment.

Using special characters can exponentially increase the strength of a password, particular when this one is targeted by dictionary attacks or even brute force[21]. The use of special characters accomplishes this by breaking the sequence of sentences, when the user creates a password with words, introducing a strange element that algorithms will not be able to identify easily. It was optimistic to see that 56.2% (145 of 258) affirmed using special characters, but there is still an alarming large percentile that doesn't do it, 43.8% (113 of 258).

Table 4. Table that relates the educational level with both question 3 and 4.

Identifier	Answer	Fourth Grade	Ninth Grade	Twelfth Grade	Bachelor	Master	Ph.D.
3	Yes	0% (0 of 25)	20.2% (52 of 258)	28.7% (74 of 258)	22.5% (58 of 258)	8.1% (21 of 258)	1.2% (3 of 258)
3	No	0.4% (1 of 258)	7.0% (18 of 258)	7.4% (19 of 258)	1.9% (5 of 258)	2.7% (7 of 258)	0% (0 of 258)
4	Yes	0% (0 of 258)	12.8% (33 of 258)	14.3% (37 of 258)	9.7% (25 of 258)	6.2% (16 of 258)	0.8% (2 of 258)
4	No	0.4% (1 of 258)	14.3% (37 of 258)	21.7% (56 of 258)	14.7% (38 of 258)	4.7% (12 of 258)	0.4% (1 of 258)

Arranging the data retrieved from the survey in academic groups and analyzing question 3, we can see that even with higher education participants still persist in reusing their passwords. In question 4, only above the Master's degree it is determined that a larger percentage of participants do include special characters in their passwords, which may indicate a higher level of awareness.

Analyzing the results from both questions, it is possible to conclude, that having a higher academic level doesn't necessarily mean that participants are more aware of the potential threats.

Table 5. Table that relates the group ages with both question 3 and 4.

Identifier	Answer	14-17	18-25	26-40	40-65	65+
3	Yes	11.2% (29 of 258)	33.3% (86 of 258)	24% (62 of 258)	9.7% (25 of 258)	2.3% (6 of 258)
3	No	3.9% (10 of 258)	7.8% (20 of 258)	1.9% (5 of 258)	5.4% (14 of 258)	0.4% (1 of 258)
4	Yes	6.2% (16 of 258)	17.4% (45 of 258)	11.6% (30 of 258)	8.5% (22 of 258)	0% (0 of 258)
4	No	8.9% (23 of 258)	23.6% (61 of 258)	14.3% (37 of 258)	6.6% (17 of 258)	2.7% (7 of 258)

With the information now organized into the group ages from the survey, it is possible to determine, relatively to question 3, that the answers, in general, are quite negative for the state of Cyber Security nowadays, where only the group from 18-25 seems to have a small indicator, 7.8% (20 of 258), that some participants are becoming more aware of small changes that they can make to increase security. In question 4 we can observe that there are other indicators in both 14-17 and 18-25 that may suggest that some representatives from these groups are more aware, which may lead to a small portion of future generations being more conscious.

It is to notice that participants with 65+ may have their privacy vulnerable since in question 4 none of the participants affirmed including specials characters in their passwords.

The most important objective of this survey was to determine if the common user reuses his passwords (question 3). This was the focus since nowadays, with all the data breaches involving personal data belonging to their user base, from companies that have authentication systems, many times users get their password unveiled and don't even get notified, leaving all their privacy exposed and potentially companies network systems compromised. It was also wanted to discover if they did include any special characters in their passwords (question 4) since this can interfere considerably with dictionary attacks.[22]

3.2 Common Technical Terms

The inquiry second part's main objective was to sub linearly educate the participants about what are some of the technical terms used in Cyber Security and what they mean. With this in mind, the questions were formulated to be multiple choice and with 4 possible answers, one of those being the right one.

The correct answer was formulated based on information acquired from an online curse from the Nacional Centre of Cybersecurity – Portugal.[18]

Bellow, it is presented a resume of all the participant's answers, where all the wrong answers were agglomerated into just one column.

Table 6. Questions from the survey related to the second part

Identifier	Question	Correct	Wrong
6	What is a Malware?	89.1% (229 of 258)	10.9% (29 of 258)
7	What is a Ransomware?	80.6% (208 of 258)	19.4% (50 of 258)
8	At what does Phishing refer to?	83.1% (212 of 258)	16.9% (46 of 258)
9	At what does the concept of Malvertising refer to?	75.2% (194 of 258)	24.8% (64 of 258)

From this data, it is possible to observe that terms more commonly used, like Malware and Phishing, have a higher Correct answer percentage, but when it comes to Ransomware and Malvertising, participants seam to still don't know about then, which suggest that some social awareness could provide some knowledge.

3.3 Acquired Results

Analyzing the answers obtained in this inquiry, it is possible to conclude that the population, in general, is evolving, overall, in a somewhat positive way considering the fast pace from technological developments and how humans must adapt to keep up with it.

Ultimately, there is still a margin for improvement. The fact that only 11.2% (29 of 258) participants have formation in Cyber Security, associated with the facts that 80.6% (208 of 258) reuses their passwords, 56.2% (145 of 258) don't use special characters, that the majority of the next generations doesn't protect their online security properly, suggest that, at an educational level there is still some enhancements that can be done, leading to conclude that it would be a high benefit, for both individuals and companies, to invest in teaching Cyber Security earlier in academic life of children's, this way

making the population more cautious on their daily use of technology, that being in the context of work, entertainment or both.[19]

4 Conclusion and Future Work

Towards accomplishing a sustainable and secure network system and more protected users, it is essential that all intervenient, system and users, are adequately prepared to ensure this. Through the combination of all topics discussed above, seems to be the most advantageous approach when creating a system like so. A combination of, the most adequate Intrusion Detection Systems, updated information about discovered forms of attack and user knowledge will certainly culminate in a sustainable network. Inquires like this yield great information about the state of the general knowledge of the population and as observed by the results, it could be gained a lot by introducing Cyber Security principals earlier in academic carriers.

As for future work, it is intended to continue this study, looking to promote awareness among users to such an important and relevant area, and also seeking to conduct studies with more extensive samples.

References

1. Hung-Jen Liao, Chun-Hung Richard Lin, Ying-Chih Lin, Kuang-Yuan Tung.: Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications (2013)
2. Heena Batra & Gaurav Gautam.: An Improved Intrusion Detection System Using Clustering Technique in Data Mining. https://edupediapublications.org/journals [Las accessed 16-12-2019] (2018)
3. Geethapriya Thamilarasu, Shiven Chawla.: Towards Deep-Learning-Driven Intrusion Detection for the Internet of Things. University of Washington Bothell (2019)
4. Sumeet Dua and Xian Du.: Data Mining and Machine Learning in Cybersecurity. Auerbach Publications (2011)
5. Harold Booth, Doug Rike and Greg Witte.: The National Vulnerability Database (NVD): Overview. ITL Bulletin (2013)
6. Anna Sperotto, Ramin Sadre, Pieter-Tjerk de Boer, and Aiko Pras.: Hidden Markov Model Modeling of SSH Brute-Force Attacks. IFIP International Federation for Information Processing (2009)
7. Subil Abraham, Suku Nair.: A PREDICTIVE FRAMEWORK FOR CYBER SECURITY ANALYTICS USING ATTACK GRAPHS. International Journal of Computer Networks & Communications (IJCNC) Vol.7, No.1. (2015)
8. Sabarathinam Chockalingam, Wolter Pieters, André Teixeira, and Pieter van Gelder.: Bayesian Network Models in Cyber Security: A Systematic Review. Proceedings of the Nordic Conference on Secure IT Systems (2017)
9. Elise T. Axelrad, Paul J. Sticha, Oliver Brdiczka, Jianqiang Shen.:A Bayesian Network Model for Predicting Insider Threats. IEEE Security and Privacy Workshops (2013)
10. Ahmet Okutan, Shanchieh Jay Yang, Katie McConky.:Forecasting Cyber Attacks with Imbalanced Data Sets and Different Time Granularities. Rochester Institute of Technology, Rochester, NY, USA (2018)

11. Lisa de Wilde.:A Bayesian Network Model for Predicting Data Breaches. University of Twente in cooperation with Delft University of Technology (2016)
12. Ba-Dung Le, Guanhua Wang, Mehwish Nasim, M. Ali Babar.:Gathering Cyber Threat Intelligence from Twitter Using Novelty Classification. University of Adelaide (2019)
13. Fernando Alves, Aurélien Bettini, Pedro M. Ferreira, and Alysson Bessani.: Processing Tweets for Cybersecurity Threat Awareness. Faculty of Sciences, University of Lisbon – Portugal (2019)
14. SpiderFoot, Open Source Intelligence Automation. http://spiderfoot.net/. [Last accessed 15-12-2019].
15. Breda F., Barbosa H., Morais T. .:Social Engineering and Cyber Security. INTED 2017 Proceedings (2017)
16. Ms M Lakshmi Prasanthi, Tata A S K Ishwarya.: Cyber Crime: Prevention & Detection. International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 3 (2015)
17. Valdemar Sousa.: A Review on Cyber Attacks and Its Preventive Measures. Proceedings of the Digital Privacy and Security Conference (2019)
18. Cidadão Ciberseguro.: Centro Nacional de Cibersegurança – Portugal (2019)
19. Noam Ben-Asher, Cleotilde Gonzalez.: Effects of cyber security knowledge on attack detection. Computers in Human Behavior 48. (2015).
20. Donald A. Norman.: The Design of Everyday Thing. Basic Books(AZ) (2013)
21. Eugene H. Spafford.: Preventing Weak Password Choices. Computer Science Technical Reports, Purdue University (1991).
22. Kouroush Jenab and Saeid Moslehpour.: Cyber Security Management: A Review. Business Management Dynamics Vol.5, No.11 (2016)
23. Martin Husák, Jana Komárková, Elias Bou-Harb, and Pavel Čeleda.:Survey of Attack Projection, Prediction, and Forecasting in Cyber Security. IEEE Communications Surveys & Tutorials (2018)
24. George Onoh.: Predicting Cyber-Attacks Using Publicly Available Data. Bowie State University (2018)
25. Lisa de Wilde.:A Bayesian Network Model for Predicting Data Breaches. University of Twente in cooperation with Delft University of Technology (2016)
26. Threat Analysis - Intelligence Monitor – Track Cyber Threats. https://www.surfwatchlabs.com/threat-intelligence-products/threat-analyst. [Last accessed 15-12-2019].
27. Geethapriya Thamilarasu, Shiven Chawla.: Towards Deep-Learning-Driven Intrusion Detection for the Internet of Things. University of Washington Bothell (2019)
28. Prabaharan Poornachandran, M. Nithun, Soumajit Pal, Aravind Ashok Nair, Aravind Ajayan.: Password Reuse Behavior: How Massive Online Data Breaches Impacts Personal Data in Web. Innovations in Computer Science and Engineering (2016)
29. Nabie Y. Conteh1, Paul J. Schmick.: Cybersecurity: risks, vulnerabilities and countermeasures to prevent social engineering attacks. International Journal of Advanced Computer Research, Vol 6(23) (2016)