European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 06, 2020

Cyber Security Trend Analysis using Web of Science: A Bibliometric Analysis

Gargi Shukla and Dr. Saikat Gochhait
Symbiosis Institute of Digital and Telecom Management
Constituent of Symbiosis International (Deemed University)

Abstract: Post COVID-19, it is almost certain that most IT/ ITES enabled industries will harness the capabilities of the cloud to promote remote working culture. With organizations going online, it is evident that probability of cyber-attacks will spike exponentially. Therefore it is very crucial for all digitally enabled industries to keep abreast with the current trends, concerns, and on-going research in the field of cyber-security (Gochhait, Shou, & Fazalbhoy, 2020). To enhance this study, the research has been conducted from the year when the first paper in the field of cyber security was published to date,which is from 1998-2020. The extensive trend analysis has been conducted using bibliometric analysis taking into consideration various parameters for analysis. The research uses Web of Science directory for data analysis, studying approximately 2184 records to enlighten the scholars around the world. This research will help academicians, students, and experts to get a complete idea of the development of cyber-security as a research field.The analysishas revealed apositive growth in the literature. The sudden growth of publications was found after 2010 with about 300 published yearly in recent past.To name a few, IEEE, Elsevier and Springer were amongst the most popular publications for quality papers on cyber security. Countries like the US, UK, Netherlands, Switzerland, Germany, and Canada have contributed significantly to the research related to cyber security.

Keywords – Cyber security, Bibliometric analysis, Scientometrics, Trend analysis

1. Introduction

In the beginning stage of researches, researchers used information security, computer security, and web security for conducting researches under this field. It was in 1998 when the first paper under the term Cyber was published which talked about the adoption of Internet-Based Securities Trading Systems to prevent securities from potential manipulation by offline brokers(Gallagher, 1998). Since then cybersecurity hasflourished as a separate research field focusing particularly on network, hardware, and software securities whereas information security formed a larger domain cutting across various other domains like law, healthcare, and finance etc.

The exponential increase in the number of portable devices and the need for mobility has pushed organizations to opt for online platforms for commercial activities.Hence cybersecurity has become the topmost concern for businesses and governments across the globe. With evolving technologies, even the hackers are getting smarter. It is because of this, IT/ITES enabled industries have dedicated auditing at regular intervals.Not only this but even healthcare systems are now digitally equipped, adopting HIT(Healthcare Information technology). But to this, is a potential risk of losing information through cyber-attacks like Man in middle, Denial of service and phishing, etc. (Bhuyan, 2020).

Numerous experiments have been going on in this field with the latest being the incorporation of machine learning in detecting possible attacks. As a result in recent years, many researches have been concentrating on proposing new ways of mitigating risk by integrating machine learning algorithms with cyber security monitoring systems.

Bibliometric analysis is a structured way of studying the already present publicationsto understand the maturity of literature. With the help of certain metrics and visualizations, a researcher can get to know the current trends, good sources of publications, and any new findings in the field.

There has been little research on the cyber security literature using bibliometrics. A trend analysis using bibliometrics is an attempt to understand the recent developments in this vast field using already present literature. The objective of this study is to uncover insights using Web of Science data points about cybersecurity as a research field.

2. Literature review

Numerous researches done in past will help an expert understand the birth of cybersecurity. The first formal paper published by Lucent Technologies exposed possible vulnerabilitiesUNIX and Windows NT systems had as they were going online with the Internet. The paper also emphasized on securing computers as they are the source of intellectual property and proposed a framework to do so(Chang , 1999).

Viruses like Melissa infected the systems in 1999, leading to the government to intervene in the digital space.Papers published between 2000 and 2001 shows the US and UK were the first to start activities for securing information. The research paper by Stanford scholars reflects how hackers wereharming commerce, government, and public informationand suggest the implementation of standards and policies through international cooperation (Sofaer & Goodman, 2001).The highly cited paper of the decade 2001-2010 proposes vulnerability assessment framework and impact of cybersecurity in SCADA systems which is quite relevant to industries in analyzing real time data(Ten, Liu & Govindarasu, 2008).

Otlet, father of Bibliometrics is known to inventmetrics and methods to assessaspects of publications and documents to discover the patterns and trends hidden in them(Rousseau, R., 2014).There have been very fewpapers measuring the research outputs on cyber security using bibliometric studies. Few of them have concentrated on adopting cybersecurity measures in specific areas of activity, such as healthcare (Jalali, 2019). Other studies have focused on the bibliometric analysis of various aspects and factors of cybersecurity, including Big Data, Industry 4.0 (Cobo, 2018), and Internet of things (IoT) (Sakhnini, 201s).Cloud forensicsand Mobile Forensicsis a reaction to new security threat classes (Gill, 2018). Other investigations were concerned with bibliometric analyses of current research being conducted on implementing machine learning in cybersecurity or bibliometric analysis of cyberbehavior(Makawana & Jhaveri, 2018). Some works aimed to provide a systematic literature review focusing on cybersecurity management, intellectual capital, and trust.

Now there are softwares like VOS viewer in the market that aid in text analysis of huge bodies of text enabling us to visualize it with the help of maps. For creating visualisations, the tool uses amapping technique known as Visualisation of similarities, hence the name VOSViewer(Van Eck and Waltman, 2007). Older tools like SPSS or Pajek, used by researchers for visualisation of co-occurrence terms had many issues related to labels overlapping and in-depth analysis of smaller parts ofmap (Van Eck and Waltman, 2010). But this tool got rid of all older vulnerabilities. This tool can directly extract datafrom various databases like Web of Science, helping researchers to create maps using varied options such as analysing the abstract, or reading the exported files directly. Users can directly use the VOS mapping method or can use multidimensional scaling methods of other tools like SPSS for creating visualisations (Van Eck and Waltman, 2010). After these maps are generated, the tool offers two ways to visualise the data. They are network and density visualisation view. Network mapping creates maps in which terms are presented by labels on top of circles. The size of circle depicts the frequency of term in the dataset and the colour of the circle relates to the cluster to which the keyword belongs. In density mapping, the terms corresponds to frequency of dataset. The colour in the density view ranges from blue to red colour scheme, expressing it from lowest density to highest density respectively. The value of a colour differs with the factors like frequency of occurrence of nearest keywords around a point or weight, and in case of co-occurrence maps, the relative frequency of terms in the data points(Van Eck and Waltman, 2015). Each view differs in a way of expressing a unique pattern hidden in the data.

3. Methodology

3.1. Study methodology

A comprehensive search was done to collect data from the Web of Science(WoS) directory. Several keywords were identified based on the title pattern of the published articles. The study was kept strictly under cybersecurity domain. A total of 2000 records were retrieved using keywords, “cybersecurity” or “cyber-security” or “cyber security”. The collected data points were further analyzed using certain tools, methods, and metrics to reveal the insights prevailing in cybersecurity field.

3.2. Data analysis methodology

The collected data wereanalyzed using Excel, VOS viewer, and Tableau for cleaning, sorting and visualization. The data was screened for duplicity using Excel. Few papershad same titles but

different authors and published in different journals. So such data points have been considered for analysis. Rest all duplicate data was removed.
Excel was used for all calculations related analysis using pivots and mathematical functions.
For different types of visualizations using graphs,VOS viewer and Tableau was used. VOS viewer helped in text analysis. Various clusters and network visualization maps were generated to analyze titles and abstracts. To find the most contributing countries, the same software was used. Tableau was used to generate other types of visualizations.

4. Results and findings

4.1. Publications per year

From the graph for 1998-2019, itcan be noted that there has been a positive progress in the number of articles published and citation with years advancing. A maximum of 381 articles waspublished in 2018. For citations, the highest numberof citations was in the year 2012 which is 2168.
Average citation per paper(ACPP) is calculated as:

$$\text{ACPP} = \frac{\text{Total number of citations}}{\text{Total number of papers published in a particular year}}$$

ACPP is a helpful metric to understand the influence of a journal or an author. It can be seen that the year 2008-2012 was a very influential time frame in the timeline of cyber security.

Year	Total citations	Total articles published	ACPP
1998-2002	79	24	3.29
2003-2007	399	93	4.29
2008-2012	3809	225	16.93
2013-2017	6283	947	6.63
2018-2022	1837	895	2.05
Grand Total	**12407**	**2184**	**5.68**

Table 1: Year-wise publication output

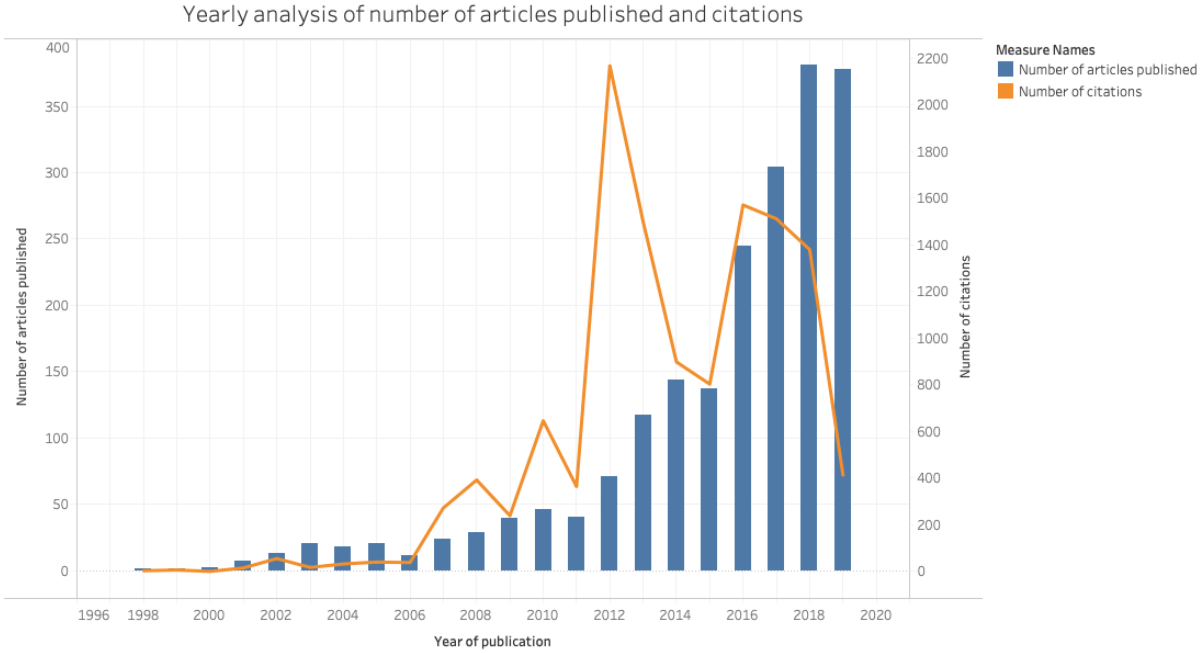

Figure 1: Graph showing total articles published and citations from 1998-2019

4.2. Country-wise output

The analysis of top 20 countries' contribution in terms of publications shows that the USA leads in country wise output contributing to maximum number of papers and citations followed by UK and Netherlands.

PEI i.e. Publication Efficiency Index is one metrics that measures the quality of research (Guan and Ma, 2007). PEI is calculated as:

$$PEI = \frac{TNC_i / TNC_t}{TNP_i / TNP_t}$$

Where,

TNCi = Total citations by a country
TNCt = Total citations by all countries
TNPi = Total papers of a country
TNPt = Total papers of all countries

The value of PEI more than 1shows a greater impact of publication than the research efforts devoted to it.Two countries US and Netherlands have PEI more than 1 which means their publications have more impact.

S.No.	Countries	Articles	Citations	PEI
1	USA	974	7175	1.30
2	UK	486	2621	0.95
3	Netherlands	167	1730	1.82
4	Switzerland	101	261	0.45
5	Germany	97	118	0.21
6	Canada	37	111	0.53
7	South Korea	31	99	0.56
8	Romania	31	17	0.10
9	Spain	26	19	0.13
10	France	19	7	0.06
11	Austria	18	26	0.25
12	Singapore	16	15	0.17
13	India	15	9	0.11
14	Japan	13	28	0.38
15	Poland	11	13	0.21
16	Australia	11	11	0.18
17	Ukraine	9	7	0.14
18	Portugal	8	23	0.51
19	China	6	31	0.91
20	Lithuania	6	21	0.62

Table 2: Country-wise number of articles published and citations

4.3. Authorship pattern

The study of the collaboration of authors led to the conclusion that the maximum contribution to the literature have been done by single authors followed bypapers having two authors.Degree of collaboration has been calculated using(Subramanyam, 1983)formula which is:

$$DC = 1 - \left[\frac{f_1}{N}\right]$$

Where,
f_1 =Papers having one author
N = Total papers published in a year

The collaboration index(CI) is calculatedas (Lawani, 1980):

$$CI = \frac{\sum_{j=1}^{A} jfj}{N}$$

Where,
j = Number of authors in a paper
fj = Number of j authors in a paper
N = Total papers published in a year
A = Total number of authors per paper
The application of the theory can be applied (Yadav et al., 2019) like collaboration index for 1998-2002 is :

$$CI = \frac{(15*1) + (2*2) + (3*0) + (2*4) + (0*5)}{19} = 1.42$$

The collaboration coefficient (CC) is calculated using(Ajiferuke, 1988) formula:

$$CC = 1 - \frac{\sum_{j=1}^{A} \left(\frac{1}{j}\right) fj}{N}$$

Where,
j = Number of authors in a paper
fj = Number of j authors in a paper
N = Total papers published in a year
A = Total number of authors per paper
For instance, the collaboration coefficient for 1998-2002 is :

$$CC = 1 - \left\{ \frac{\left[\frac{15*1}{1}\right] + \left[\frac{2*1}{2}\right] + \left[\frac{3*1}{0}\right] + \left[\frac{2*1}{4}\right] + \left[\frac{0*1}{5}\right]}{19} \right\} = 0.13$$

Higher the value of collaboration coefficient,better the collaboration rate. If the value is nearer to0, it means authors have a weak collaboration (Elango and Rajendran, 2012).

Year	Single author paper	Two author people	Three author people	Four author paper	Five author paper	More than five author paper	Total Papers	Degree of collaboration	Collaboration Index	Collaboration Coefficient
2018-2022	218	189	186	139	77	75	884	0.75	2.99	0.51
2013-2017	342	219	171	89	55	46	922	0.63	2.47	0.41
2008-2012	105	45	32	17	5	5	209	0.50	1.99	0.31
2003-2007	58	8	4	4	3	0	77	0.25	1.52	0.16
1998-2002	15	2	0	2	0	0	19	0.21	1.42	0.13

Table3: Authorship pattern of paper published

4.4. Contribution of authors country wise

In this diagram, each bubble represent a country or a territory. The size of each bubble depicts the number of papers contributed by a country and the colour tells the number of citations referring to RGB colour scheme. The line connecting any two bubbles shows the cooperative relationship between them.

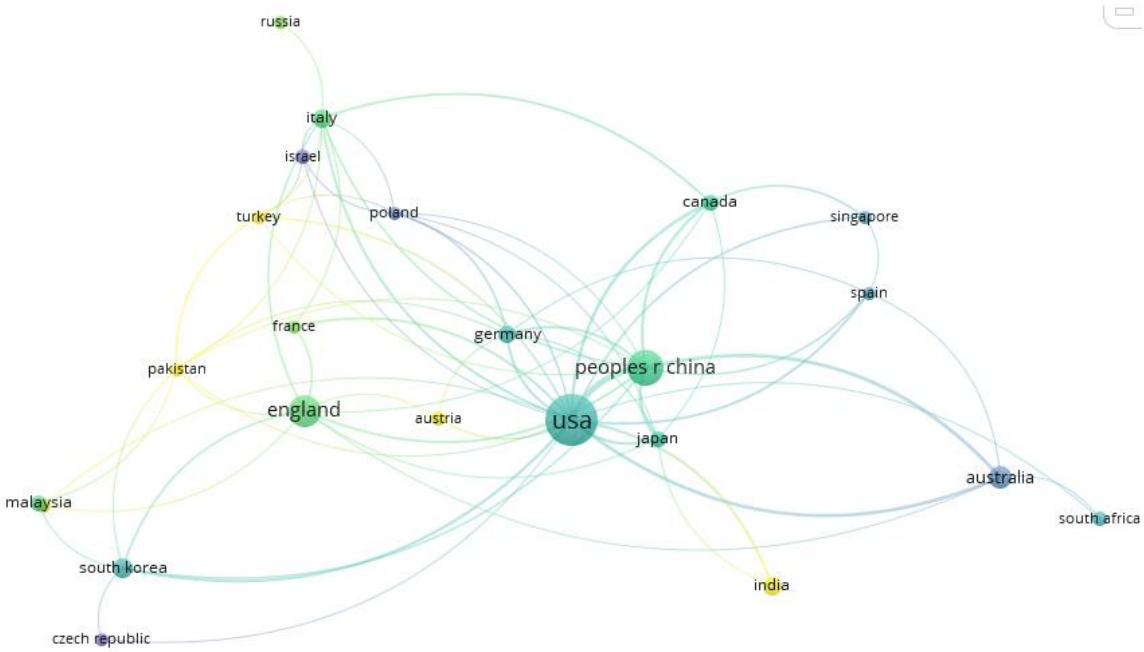

Figure2: Network visualization map of co-authorship and collaborations between countries

4.5. Keyword analysis

The network map of author keyword occurrencesis shown below. In this graph each bubble represents a keyword. The size of each bubble depicts the frequency of occurrence of that keyword. The colour of the bubble shows the cluster to which the keyword belongs. Keywords of same color were commonly listed together. So, for instance, from the size of the bubble, the most frequently used keywords arecyber security, security, information security, internet, and attacks. From the color, it can be inferred that cyber security, cyber-crime, networks, machine-learning, and challenges are usually quoted together due to similar nature of expression (Gochhait, and Rimal, 2019).

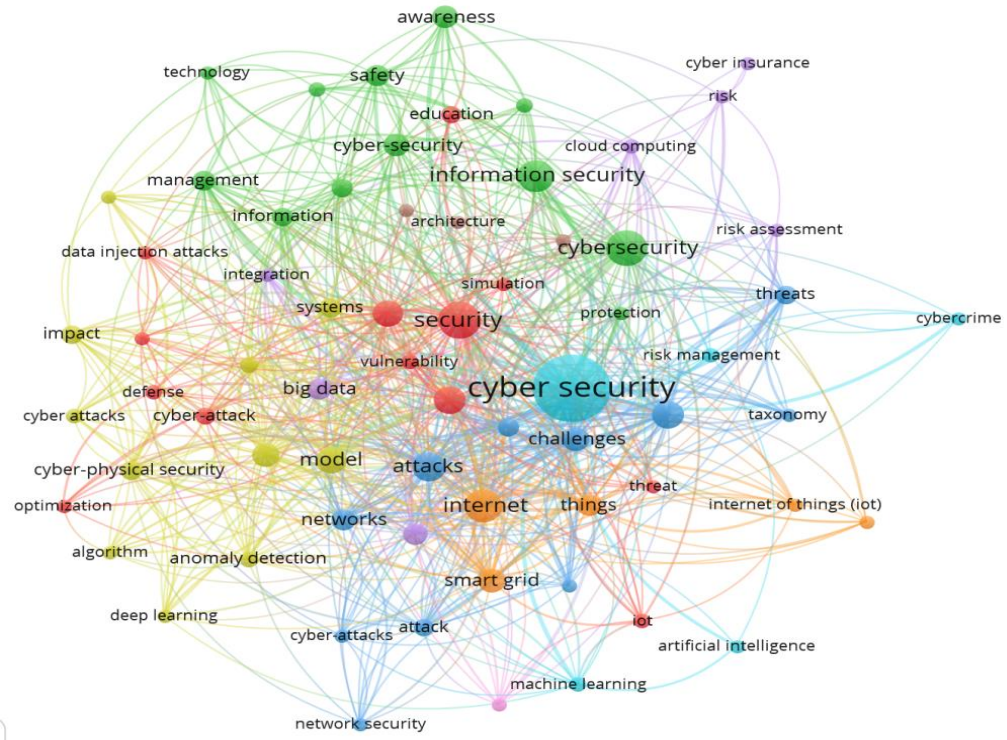

Figure3: Network visualization map on high-frequency keywords

4.6. Percentage of Papers Publishers

It was found that IEEE is the most well-recognized publisher with a total contribution of 40%, publishing 313 articles with 5198 citations in total. If the contribution of IEEE is considered domain wise then IEEE – INST ELECTRICAL ELECTRONICS ENGINEERS INC has contributed the most in terms of maximum papers published and total citations in the field of cybersecurity. Therefore it can be inferred that IEEE contributes to the maximum research followed by Elsevier, Routledge Journal, Springer, and Willey respectively.

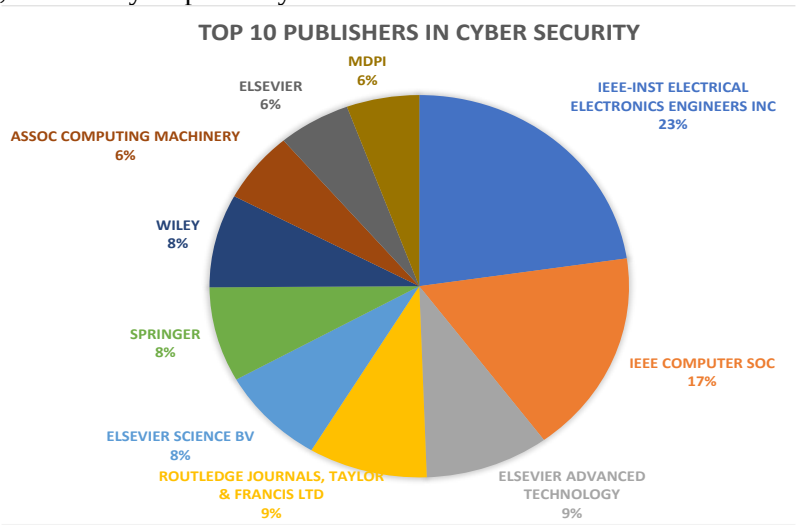

Figure 4: Top 10 publishers in cyber security

4.7. Most cited articles

S.No	Authors	Document Title	Publication Name	Total times cited	Year published
1	Mo, Y; Kim, Tiffany; Brancik, K; Dickinson, D; Lee, H; Perrig, A; Sinopoli, B	Cyber-Physical Security of a Smart Grid Infrastructure	PROCEEDINGS OF THE IEEE	466	2012
2	Sridhar, S; Hahn, A; Govindarasu, M	Cyber-Physical System Security for the Electric Power Grid	PROCEEDINGS OF THE IEEE	436	2012
3	Wang, W; Lu, Z	Cyber security in the Smart Grid: Survey and challenges	COMPUTER NETWORKS	402	2013
4	Buczak, A. L.; Guven, E	A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection	IEEE COMMUNICATIONS SURVEYS AND TUTORIALS	334	2016
5	Ten, C.W; Liu, C.C; Manimaran, G	Vulnerability Assessment of Cybersecurity for SCADA Systems	IEEE TRANSACTIONS ON POWER SYSTEMS	268	2008
6	Yan, Y; Qian, Y; Sharif, H Tipper, D	A Survey on Cyber Security for Smart Grid Communications	IEEE COMMUNICATIONS SURVEYS AND TUTORIALS	260	2012
7	Liu, J; Xiao, Y; Li, S; Liang, W; Chen, C. L. P	Cyber Security and Privacy Issues in Smart Grids	IEEE COMMUNICATIONS SURVEYS AND TUTORIALS	233	2012
8	Ericsson, GN.	Cyber Security and Power System Communication-Essential Parts of a Smart Grid Infrastructure	IEEE TRANSACTIONS ON POWER DELIVERY	200	2010
9	Ten, CW; Manimaran, G; Liu, CC	Cybersecurity for Critical Infrastructures: Attack and Defense Modeling	IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS PART A-SYSTEMS AND HUMANS	161	2010
10	Hahn, A; Ashok, A; Sridhar, S; Govindarasu, M	Cyber-Physical Security Testbeds: Architecture, Application, and Evaluation for Smart Grid	IEEE TRANSACTIONS ON SMART GRID	138	2013

Table 4: Top 10 most cited papers incyber security till date

4.9. Objective-wise paper published in 2019

When investigatingthe papers published in 2019, it was seen security, machine-learning, cyber-attack, Internet of Things (IoT), and privacy were among the few common objectives of the articles. The data depicts the trend from Jan-Dec 2019.

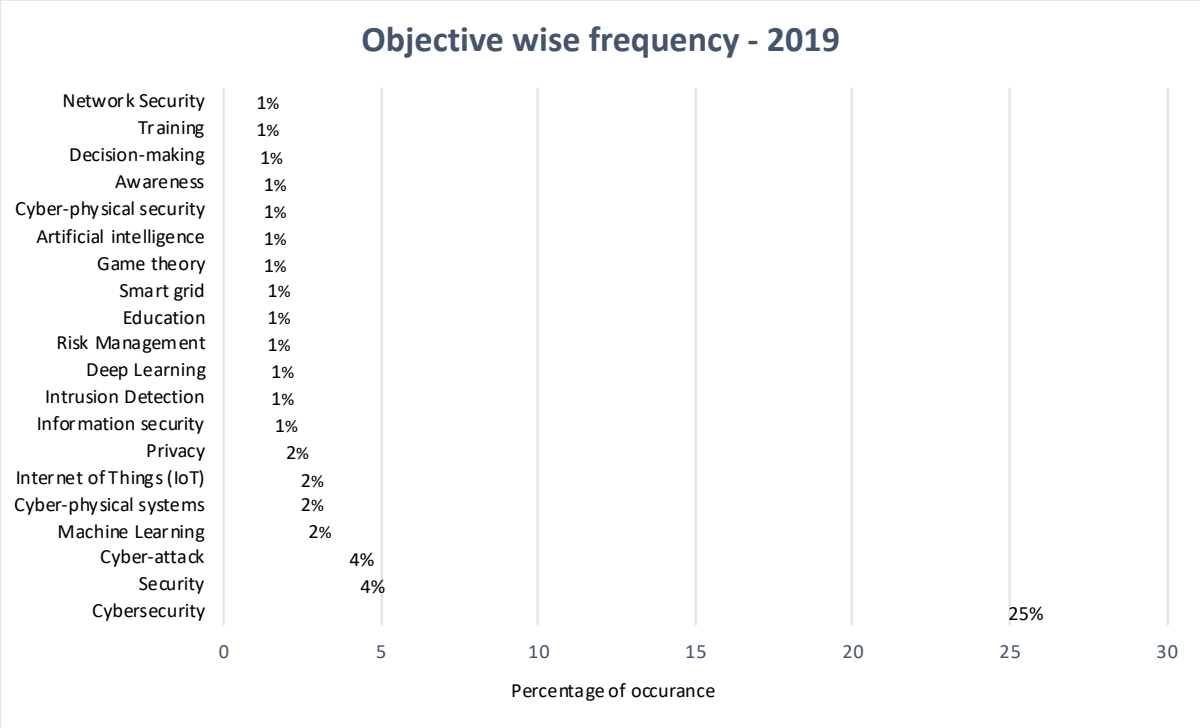

Figure 5: Percentage of objective wise papers published in 2019

Conclusion

The paper tries to analyzeliterature published under the Cyber Security domain. The dataextracted from Web of Science databasewas analyzed on multiple parameters.From the researchit is evident that USA and Netherlands aretwo countries that have significantly contributedto cyber security field with impact score of 1.30 and 1.82 respectively.Maximum research papers have been published by single authors. IEEE and Elsevier are therichest sourcesof cyber security literature. Analysis of articles from January 2019 - December 2019, revealed that the current trend in cyber security is theautomation of cyber behavior through machine learning algorithms. The implementation of security in various facets of the digital ecosystem is another. USA and China are two countries ahead in co-authorship and collaboration for research works.

This research has been done with cyber security as a separate field considering other factors like government support, law, and comparison of one nation's progress when compared with others as constant. Considering such factors in mind, the bibliometric analysis has good potential for future research in those areas.

References

1. AjiferukeI., BurellQ., and Tague, J. (1988). Collaborative Coefficient: A SingleMeasure of the Degree of Collaboration in Research. *Scientometrics, 14.* 421-433. doi: 10.1007/BF02017100
2. Bhuyan, S.S., Kabir, U., Escareno, J.M. et al. (2020). Transforming Healthcare Cybersecurity from Reactive to Proactive: Current Status and Future Recommendations. *Journal of Medical System,44(98*). doi: https://doi.org/10.1007/s10916-019-1507-y
3. Chang, ES; Jain, AK; Slade, DM; Tsao, SL. (1999). Managing cyber security vulnerabilities in large networks. *Bell Labs Technical Journal, 4(4),*252-272. doi: 10.1002/bltj.2202
4. Cobo JM.,Jūrgens B. ,Solana HV.,Martínez AM.,Viedma HE. (2018). Industry 4.0: a perspective based on bibliometric analysis. *Procedia Computer Science, 139.* 364-371. doi: https://doi.org/10.1016/j.procs.2018.10.278.
5. Daniel M. Gallagher.(1998). Move over Tickertape, Her tape, Here Comes the CyberExchange: The Rise change: The Rise of Internet-Based Securities Trading Systems.

Catholic University law review,*47*(3), 1009-1056. Available at: https://scholarship.law.edu/lawreview/vol47/iss3/10.

6. Elango B.& Rajendran, Dr. (2012). Authorship Trends and Collaboration Pattern in the Marine Sciences Literature : A Scientometric Study. *International Journal of Information Dissemination and Technology, 2(3)*. 166-169.Available at: http://www.ijidt.com/index.php/ijidt/article/view/91
7. Gochhait,S., Rimal, Y. (2019). "Machine Learning Neural Analysis Noisy Data", International Journal of Engineering and Advanced Technology , ISSN: 2249-8958, 8(6),08/2019.
8. Gochhait, S., Shou, D. T., & Fazalbhoy, S. (2020). Cloud Computing Applications and Techniques for E-Commerce. IGI Global. http://doi:10.4018/978-1-7998-1294-4.
9. Gill J., Okere I., HaddadPajouh H., Dehghantanha A. (2018) Mobile Forensics: A Bibliometric Analysis. In book: Cyber Threat Intelligence. *Advances in Information Security,70*.doi: https://doi.org/ 10.1007/978-3-319-73951-9_15.
10. Guan, J. and Ma, N. (2007). A bibliometric study of China's semiconductor literature compared with other major asian countries.*Scientometrics 70(1)*,107–124.doi: https://doi.org/10.1007/s11192-007-0107-7
11. Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S. (2019). Health Care and Cybersecurity: Bibliometric Analysis of the Literature**.***Journal of Medical Internet Research.21(2)*. doi: https://doi.org/10.2196/12644.
12. Makawana, P &Jhaveri, R. (2018). A Bibliometric Analysis of Recent Research on Machine Learning for Cyber Security.*Intelligent Communication and Computational Technologies.* 213-226. doi: 10.1007/978-981-10-5523-2_20.
13. Rousseau, R. (2014). Forgotten founder of bibliometrics. *Nature, 510(218)*.doi: https://doi.org/10.1038/510218e
14. Sakhnini J.,Karimipour H.,Dehghantanha A.,Parizi MR.,Srivastava G. (2019). Security aspects of Internet of Things aided smart grids: A bibliometric survey. *Science Direct*.doi: https://doi.org/10.1016/j.iot.2019.100111
15. Sofaer, AD; Goodman, SE. 2001. *Cybercrime and security - The transnational dimension.* 1-32. Available at: https://media.hoover.org/sites/default/files/documents/0817999825_1.pdf
16. Ten, Chee-Wooi& Liu, Chen-Ching &Govindarasu, Manimaran. (2008).Vulnerability Assessment of Cybersecurity for SCADA Systems. *IEEE Transactions on Power Systems, 23(4)*, 1836 - 1846.doi: 10.1109/TPWRS.2008.2002298
17. Van EckN.J. and Waltman L. (2007) VOS: A New Method for Visualizing Similarities Between Objects. In: Decker R., Lenz H.J. (eds) Advances in Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization. 299-306. doi: https://doi.org/10.1007/978-3-540-70981-7_34
18. Van EckN.J.& Waltman L. (2010). Software survey: VOSviewer, a computer programfor bibliometric mapping. *Scientometrics, 84(2)*, 523–538. doi: https://doi.org/10.1007/s11192-009-0146-3
19. Waltman L and Van Eck N.J.(2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics, 9(4)*, 872–894.doi: https://doi.org/10.1016/j.joi.2015.08.001
20. Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review.*Journal of InformationScience, 6(1)*,33–38.doi: 10.1177/016555158300600105
21. YadavS., Singh S.&Verma M. (2019). Authorship and Collaboration Pattern in SRELS Journal of Information Management during 2008-2017: An Evaluation.*Library of Philosophy and Practice (e-journal)*, 1-15. Available at: http://digitalcommons.unl.edu/libphilprac/2119.