DATA SECURITY AS A TOP PRIORITY IN THE DIGITAL WORLD: PRESERVE DATA VALUE BY BEING PROACTIVE AND THINKING SECURITY FIRST

Anastasija Nikiforova

University of Tartu, Institute of Computer Science

European Open Science Cloud Task Force «FAIR metrics and data quality»

Email: Nikiforova.Anastasija@gmail.com,

Website: https://anastasijanikiforova.com/

Research and Innovation Forum (Rii Forum), April 20-22, 2022

BACKGROUND AND MOTIVATION

Today, in the age of information and Industry 4.0, large amounts of data are being continuously produced, collected, processed, and exchanged between different systems.

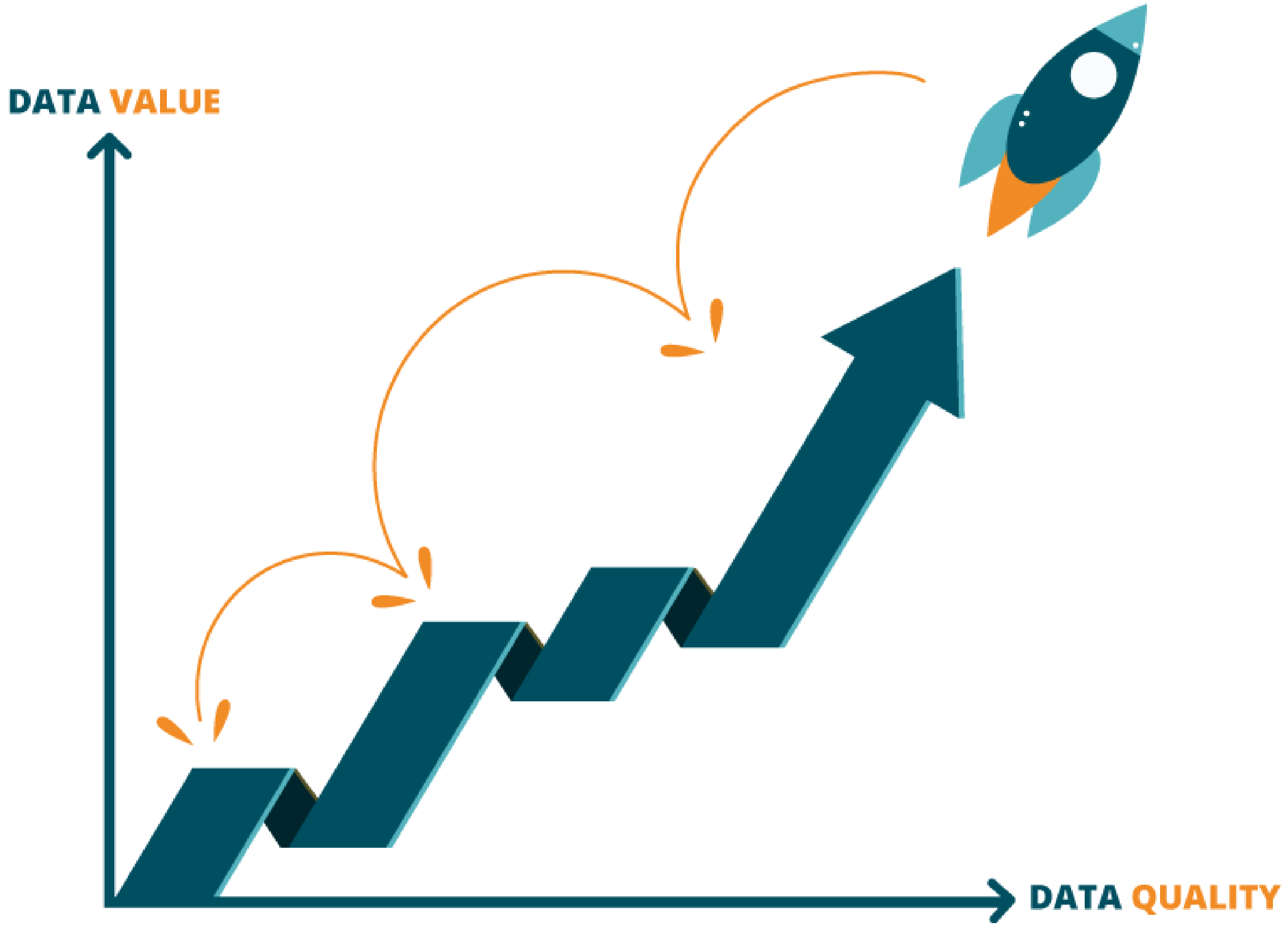
Due to the digitization and variety of data being continuously produced and processed with a reference to Big Data, their value, is also growing.

This is all the more relevant in times of COVID-19 pandemic, which has affected not only the health and lives of human beings' but also the lifestyle of society, i.e. the digital environment has replaced the physical.

an increase in cyber security threats of various nature (including but not limited to security breaches and data leaks)

BACKGROUND AND MOTIVATION

The value of data depends on several factors, where data quality and data security that can affect the data quality, are the most vital.



Data serve as the basis for decision-making, input for models, forecasts, simulations etc., which can be of high strategical and commercial / business value.

This has become even more relevant in terms of COVID-19 pandemic, when in addition to affecting the health, lives, and lifestyle of billions of citizens globally, making it even more digitized, it has had a significant impact on business.

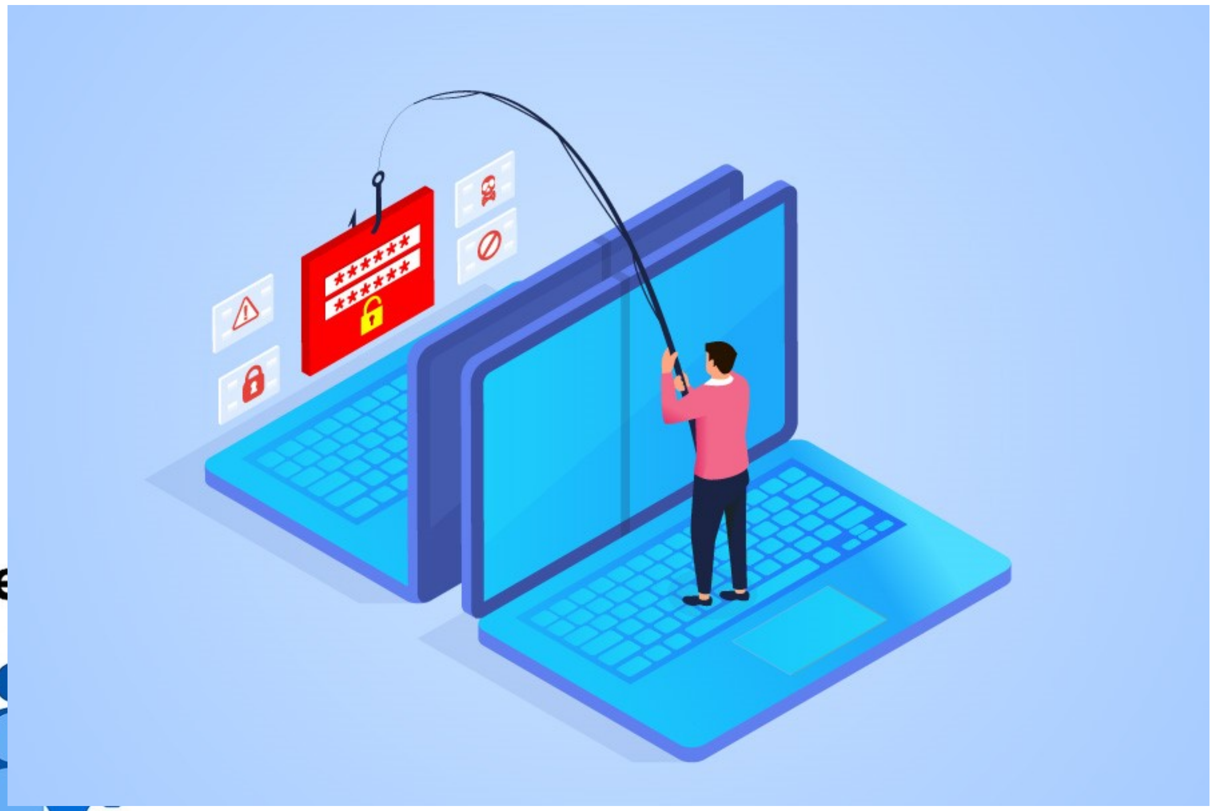
↓

In addition to those cybersecurity threats that are caused by changes directly related to the pandemic and its consequences, many previously known threats have become even more desirable targets for intruders, hackers.

Source: https://blog.devolutions.net/2021/03/how-remote-working-is-permanently-changing-the-business-landscape/

CYBERPANDEMIC

The current state of cyber-security horizon during the pandemic clearly indicate a very significant increase of cybersecurity threats.



- ✓ **600% increase in phishing attacks in 2020, when some countries were not even affected (Shi, 2020)**
- ✓ **a record-breaking number of data compromises, where "*the number of data compromises was up more than 68% when compared to 2020*" (Miles, 2022)**
- ✓ **73 million records were exposed in March 2022, and 358 vulnerabilities were identified as having a public exploit that had not yet been provided with CVE IDs (Risk Based Security Monthly Newsletter).**
- ✓ **LinkedIn was the most exploited brand* in phishing attacks last quarter (TechRepublic , 2022)**

** DHL, Google, Microsoft, FedEx, WhatsApp, Amazon, Maersk, AliExpress and Apple are also in the list of top targets*

While Risk based security & Flashpoint suggests that vulnerability landscape is returning to normal, there is another trigger closely related to cyber-security that is now affecting the world - geopolitical upheaval.

A Data Breach Investigations Report (2021) revealed that one of the most prominent and growing problems is the misconfiguration of DBMS - this is even more the case for NoSQL.

mongoDB

elasticsearch

Source: https://abacode.com/how-to-protect-your-business-from-phishing-attacks/, Shi, F. (2020). Threat spotlight: Coronavirus-related phishing. Barracuda Networks, https://blog. barracuda. com/2020/03/26/threat-spotlight-coronavirus-related-phishing, Miles B. (2022) How to minimize security risks: Follow these best practices for success, https://www.techrepublic.com/article/minimizing-security-risks-best-practices/?utm_source=email&utm_medium=referral&utm_campaign=techrepublic-news-special-offers, Risk based security & Flashpoint (2021) 2021 Year End Report Vulnerability QuickView, https://www.techrepublic.com/article/linkedin-most-exploited-brand-phishing/?utm_source=email&utm_medium=referral&utm_campaign=techrepublic-news-special-offers Verizon. 2021 Data Breach Investigations Report (DBIR). 2021. 119 Pages, https://www.verizon.com/business/resources/reports/2021/2021-data-breach-investigations-report.pdf, last accessed 2022/03/31

BACKGROUND AND MOTIVATION

While security breaches and security protection mechanisms of different nature have been widely covered in the literature, the concept of a “primitive” artifact such as data management system seems to have been more neglected by researchers and practitioners.

But are data management systems always secure and protected by default?

Previous research and regular updates on data leakages suggest that the number and nature of these vulnerabilities are high.
It also refers to little or no DBMS protection, especially in case of NoSQL databases and Big Data respectively, which are thus vulnerable to attacks.

CYBERPANDEMIC AND SEARCH ENGINES FOR INTERNET OF EVERYTHING

A recent research demonstrated that weak data and database protection in particular is one of the key security threats.

!!! Moreover !!! recent advances such as search engines for Internet connected devices*** decreased a level of complexity of searching for connected devices on the internet and easy access even for novices due to the widespread popularity of step-by-step guides on how to use IoT search engine to find and gain access (if insufficiently protected) to webcams, VoIP phones, routers, databases and in particular non-relational (NoSQL) databases, and other more «exotic» artifacts such as power plants, wind turbines or refrigerators.

*** also known as Internet of Things Search Engines, Internet of Everything (IoE) or Open Source Intelligence (OSINT) Search Engines

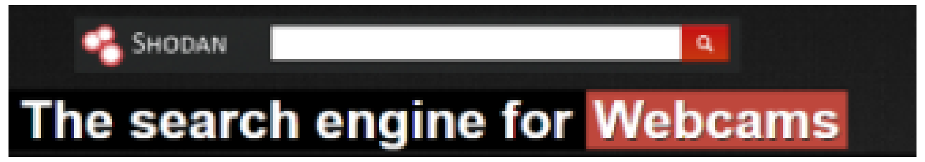
SHODAN SEARCHES

censys

ZoomEye

IOTCRAWLER

GREYNOISE INTELLIGENCE


SHODAN

The search engine for Webcams

The search engine for Refrigerators

Shodan is the world's first search engine for Internet-connected devices.

Create a Free Account   Getting Started

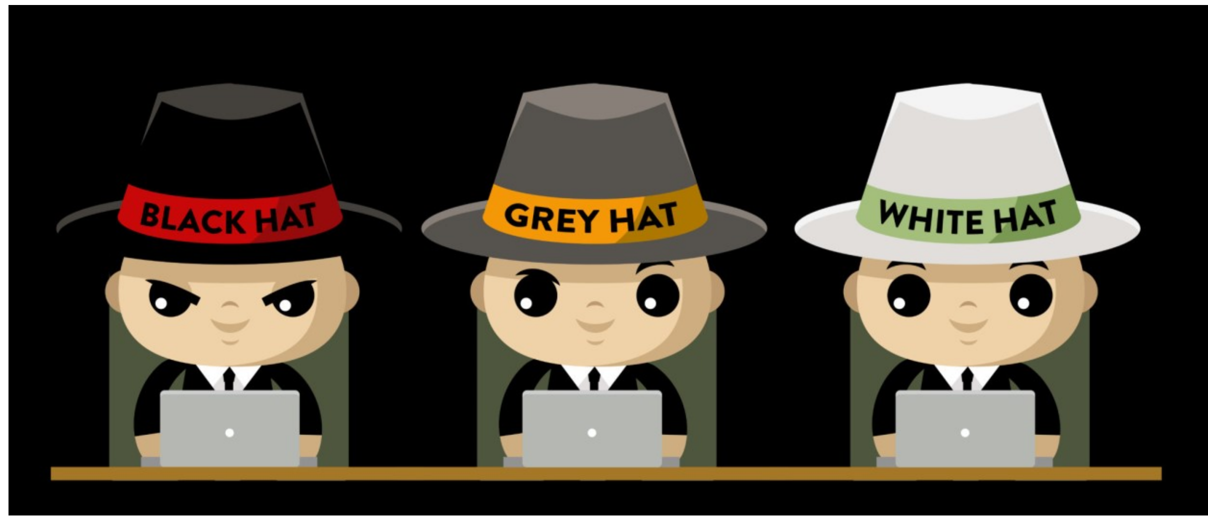
The search engine for Power Plants

IOTCRAWLER   HOME   ABOUT   FOLLOWER GROUP   NEWS   PARTNERS   SCENARIOS

IoTCrawler – a search engine for Internet of Things devices

SHODAN

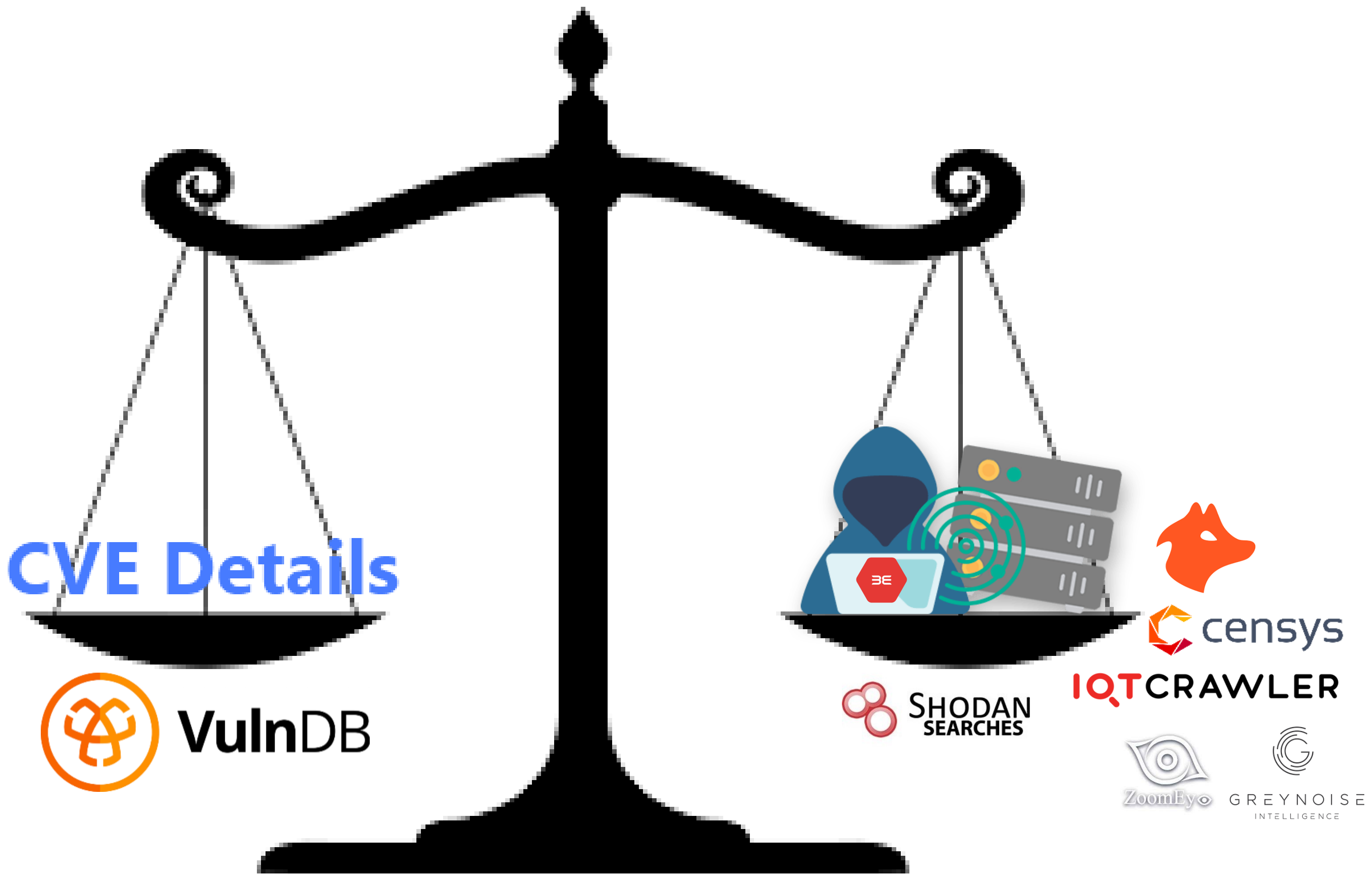
The Search Engine for Everything



DATA AND DATABASE PROTECTION AS A TOP PRIORITY

- **In the past, vulnerability databases such as CVE Details were considered useful resources for monitoring the security level of a product being used.**
- **BUT! they are static and refer to very common vulnerabilities in the product and are registered when a vulnerability is detected.**
- **Moreover, there is an opinion that they *tend to be inaccurate and incomplete*.**

Advances in ICT, including the power of the IoTSE, require the use of more advanced techniques for this purpose.

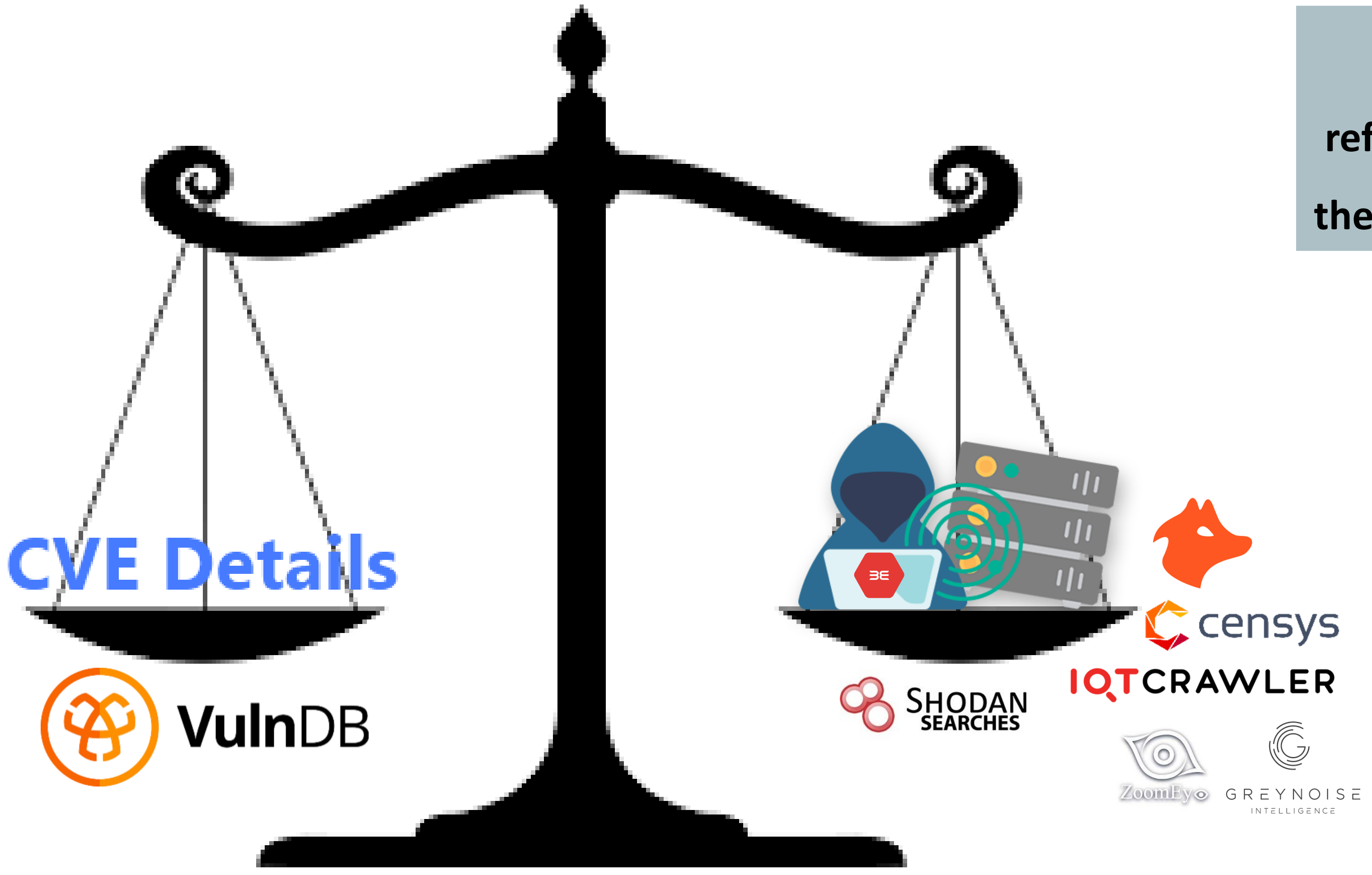


OBJECTIVE

…to examine current data security research and to analyze *whether "traditional" vulnerability registries provide a sufficiently comprehensive view of DBMS security, or they should be rather inspected by using IoTSE-based and respective passive testing* or dynamically inspected by DBMS holders conducting an active testing.***

*****this study refers to Shodan- and Binary Edge- based vulnerable open data sources detection tool – ShoBeVODSDT (Daskevics & Nikiforova, 2021)**

ShoBEVODSDT uses mainly the passive assessment (non-intrusive testing), thus refering to the most likely and potentially existing bottlenecks or weaknesses which, if the 4th stage of the penetration testing - the attack would take place, could be exposed



Daskevics, A., & Nikiforova, A. (2021, December). IoTSE-based open database vulnerability inspection in three Baltic countries: ShoBEVODSDT sees you. In *2021 8th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)* (pp. 1-8). IEEE.
Daskevics, A., & Nikiforova, A. (2021, November). ShoBeVODSDT: Shodan and Binary Edge based vulnerable open data sources detection tool or what Internet of Things Search Engines know about you. In *2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)* (pp. 38-45). IEEE.

GENERAL DATABASE-WISED STATISTICS OF THEIR VULNERABILITY (based on CVE Details)

Database	Type of database	1st vulnerability registered	last vulnerability registered	Total # of vulnerabilities	Most popular vulnerability	TOP-3 vulnerabilities in 2018-2022
Oracle	Relational, multi-model	2008	2021	44	DoS	DoS, Code Execution, Gain Information
MySQL	Relational, multi-model	2001	**2015**	**152**	DoS	--
Microsoft SQL Server	Relational, multi-model	1999	2021	**87**	Code Execution	Code Execution
PostgreSQL	Relational, multi-model	1999	2022	**134**	DoS	Code Execution, Overflow, Sql Injection
MongoDB	Document, multi-model	2013	2022	38	DoS	DoS, Code Execution, Overflow, Bypass Something
Redis	Key-value, multi-model	2015	2021	23	Overflow	Overflow, Code Execution, Memory corruption, Bypass something
IBM Db2	Relational, multi-model	2004	2021	**106**	DoS	Code Execution, Overflow, Gain Information
Elasticsearch	Search engine, multi-model	2018	2022	22	Gain Information	Gain Information, DoS, Gain privilege, Code execution
Microsoft Access	Relational	1999	**2020**	17	Code execution	Code execution, Overflow
SQLite	Relational	2009	2022	**48**	DoS	Code execution, DoS, Overflow
Cassandra	Wide column store	2015	2022	6	Code execution	Code execution, DoS, Bypass Something
Memcached	Key-value store	2013	**2020**	14	DoS	DoS, Overflow
CouchDB	Document, multi-model	2010	2021	15	Code Execution	Code Execution, Bypass Something, Gain Privileges

BRIEF INSIGHT INTO CVE DETAILS

- Despite the undeniable popularity of NoSQL databases, relational databases remain popular, and TOP-5 consists of 4 RDBMS and MongoDB. However, all the most popular relational DBMS, taking the highest places are multi-model.
- The highest number of discovered vulnerabilities are in MySQL, although this is the only DB for which data are no longer provided. It is followed by PostgreSQL and IBM Db2, with Cassandra, Memcached, CouchDB, Microsoft Access, Elasticsearch and Redis reporting the fewest vulnerabilities.

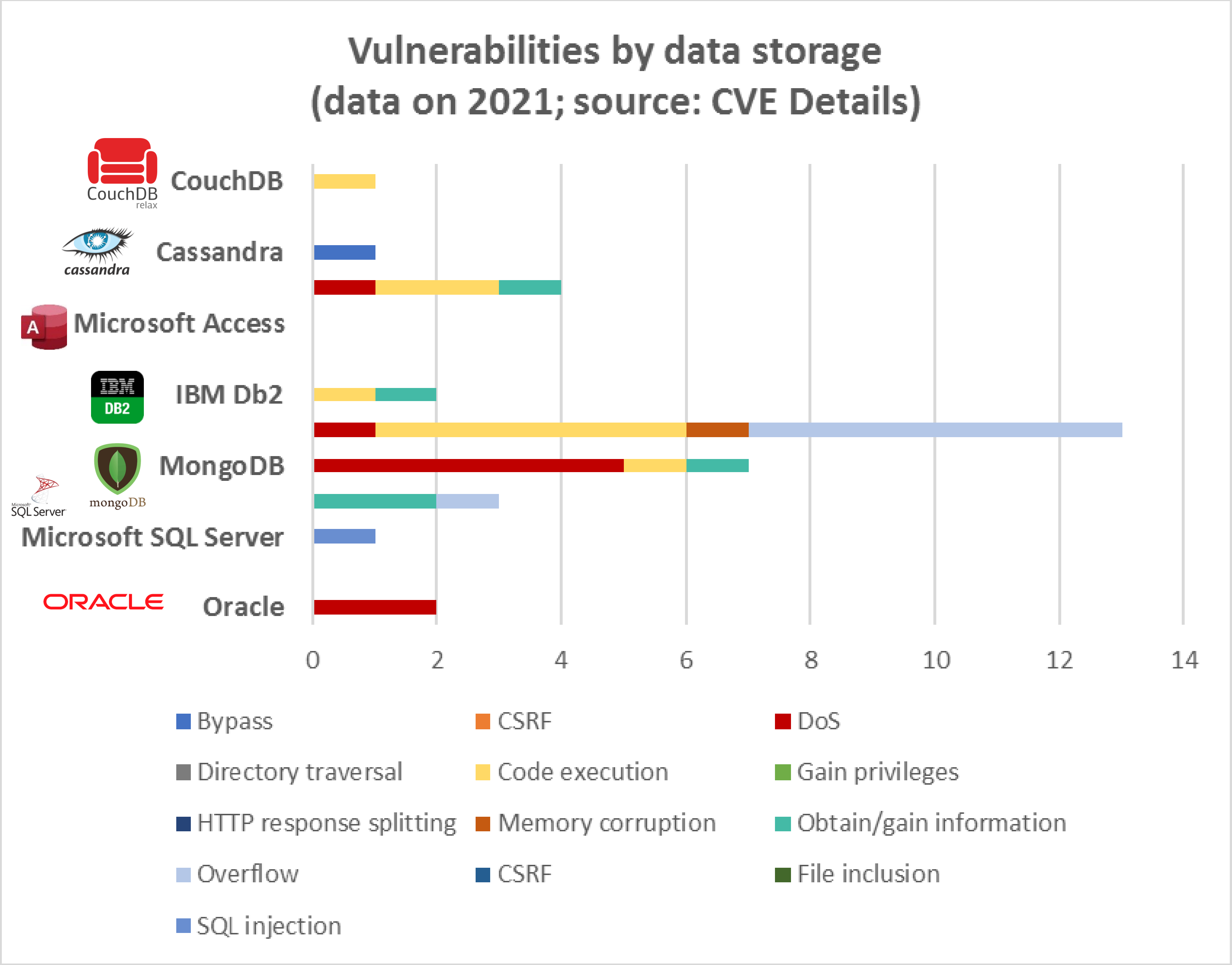
BUT!

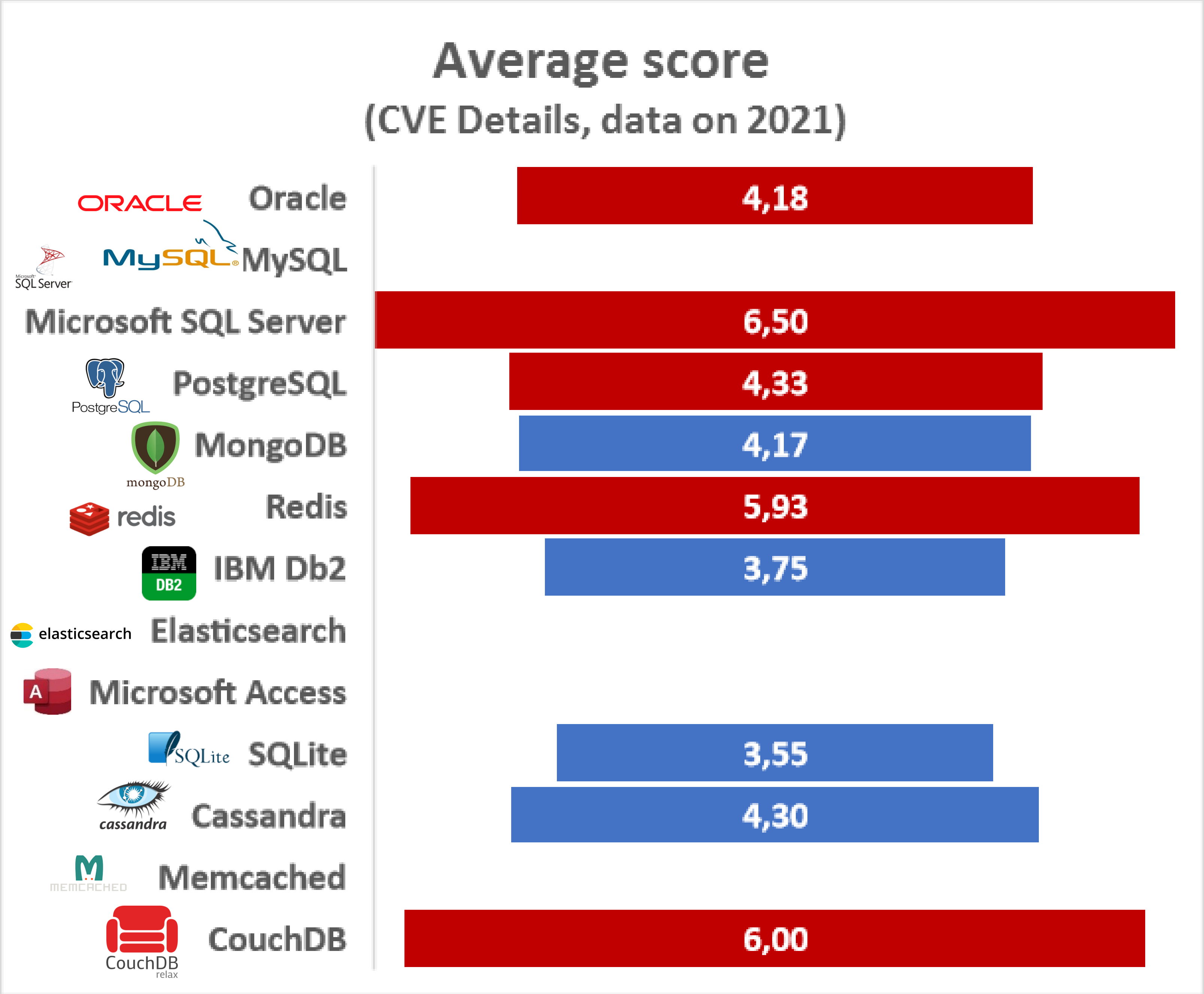
- The number of revealed vulnerabilities does not necessarily mean that the level of the relevant DB is definitely higher or lower, which may depend on their popularity, users and community involvement, this suggests such an assumption.
- E.g. the aforementioned DB with fewer reported vulnerabilities have come under the spotlight in some of recent data leakages, with Elasticsearch and MongoDB dominating.

The most common and major vulnerabilities encountered over time are DoS, although code execution is also a widespread vulnerability.

A database-wised analysis of the most frequently reported vulnerabilities over the past 5 years demonstrate that Code Execution is the most common and is in the TOP-3 for 11 databases, followed by overflow, DoS, bypassing something, gaining information.

VULNERABILITY OF DATA STORAGES IN 2021 AND THEIR SCORE (based on CVE Details)





- The very obvious and strong conclusions cannot be drawn from the data provided **BUT** it can be speculated that MongoDB is weak against DoS, but Redis against code execution and overflow.
- To get more supported results, this paper addresses the call made in (Daskevics & Nikiforova, 2021) and maps the results obtained in their study to the data obtained from CVE Details.

CVE DETAILS- AND IOTSE- STATISTICS ON DB VULNERABILITY

	CVE Details			IoTSE tool			
Database	**Total # of vulnerabilities**	**Total registered**	**Ratio (Info gained/total)**	**Total DBMS found**	**# DBMS connected**	**Gathered data or compromised**	**Ratio (Info gained/ connected)**
Oracle	**11**	**2**	0%	-	-	-	-
MySQL	-	0	0%	**13452**	0,13%	0%	0%
Microsoft SQL Server	1	1	0%	-		-	-
PostgreSQL	**5**	**3**	**67%**	**1187**	0,17%	0%	0%
MongoDB	**13**	**7**	14%	**177**	8%	**79%**	7%
Redis	**8**	**13**	0%	122	10%	**83%**	**83%**
IBM Db2	2	2	**50%**	-		-	-
Elasticsearch	-	0	0%	86	**90%**	**27%**	**9%**
Microsoft Access	-	0	0%	-		-	-
SQLite	2	1	**50%**	-		-	-
Cassandra	1	1	0%	7	**14%**	0%	0%
Memcached	-	0	0%	116	**80%**	26%	**24%**
CouchDB	1	1	0%	14	0	0	0

RESULTS

MySQL *!!! the data on which are not updated by CVE !!!* accounts more than half of all databases found on the Internet. But the number of DB that it was able to connect to is not very high similar to PostgreSQL where the number of found DB is 1187 with only 2 DB could be connected.

The absolute leader in this negative trend is Memcached - possible to connect to 93 of 116 DB with more than 20% of the databases, from which data can be gathered or they were found to be already compromised.

Elastisearch - possible to connect to 90% of all DB found, and 27% already compromised or data could be gathered. *!!! CVE Details does not provide details of its vulnerabilities in 2021 !!!*

MongoDB and Redis showed the worst results for both data sources - MongoDB was inferior to data gatherings and has a large number of compromised DB according to ShoBeVODSDT and is subject to both DoS, code execution and data gatherings according to CVE Details.

Redis with being relatively difficult to connect to (every 10th DB), is characterized by a high ratio of information gatherings. According to CVE Details, both DoS, code execution, overflow, and memory corruption have been detected for it.

5 vulnerabilities in PostgreSQL registered by CVE Details, with 2 of them related to information gaining that was not found by ShoBeVODSDT.

Oracle was one of the most frequently reported databases in CVE Details, with 10 vulnerabilities in total, while only two of them have a comprehensive description - both related to DoS.

RESULTS

- All in all, the results in most cases are rather complimentary, and one source cannot completely replace the second*

  * not only due to scope limitations of both sources - CVE Details cover some databases not covered by ShobeVODSDT, while not providing the most up-to-date information with a very limited insight on MySQL

BUT!

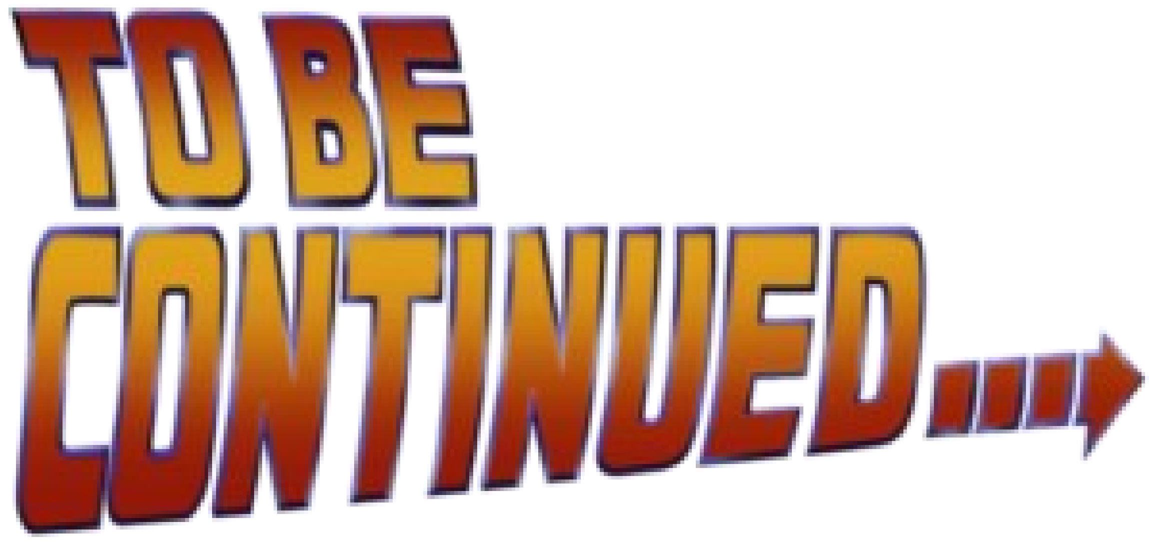
there are cases when both sources refer to a security-related issue and their frequency, which can be seen as a trend and treated by users respectively taking action to secure the database that definitely do not comply with the “secure by design” principle. This refers to MongoDB, PostgreSQL and Redis.

CouchDB, however, can be considered relatively secure by design, as is less affected, as evidenced by both data sources, where only 1 vulnerability was reported in CVE Details in 2021, and it was the only data source, to which ShoBeVODSDT was not able to connect**

**could be because CouchDB proved to be less popular, with only 14 of nearly 15 000 instances found

CONCLUSIONS

- Obviously, data security should be the top priority of any information security strategy. Failure to comply with the requirements for security and protection of data can lead to significant damage and losses of a different nature - commercial, reputation, operational etc.
- However, despite the undeniable importance of data security, the current level of data security is relatively low – data leaks occur regularly, data become corrupted, in many cases remaining unnoticed for IS owners.
- This study provided a brief insight of the current state of data security provided by CVE Details – the most widely known vulnerability registry, considering 13 DB. Although the idea of CVE Details is appealing, i.e., it supports stakeholder engagement, it is obviously not sufficiently comprehensive - can be used to monitor the current state of vulnerabilities, but this static approach, which sometimes provides incomplete or inconsistent information even about revealed vulnerabilities, must be complemented by other more dynamic solutions.
- This includes not only the use of IoTSE-based tools, which, while providing valuable insight into unprotected DB seen or even accessible from outside the organization, are also insufficient.



CONCLUSIONS

- While this may seem ridiculous in light of current advances, the first step that still needs to be taken thinking about date security is to make sure that the DB uses the basic security features: authentication, access control, authorization, auditing, data encryption and network security

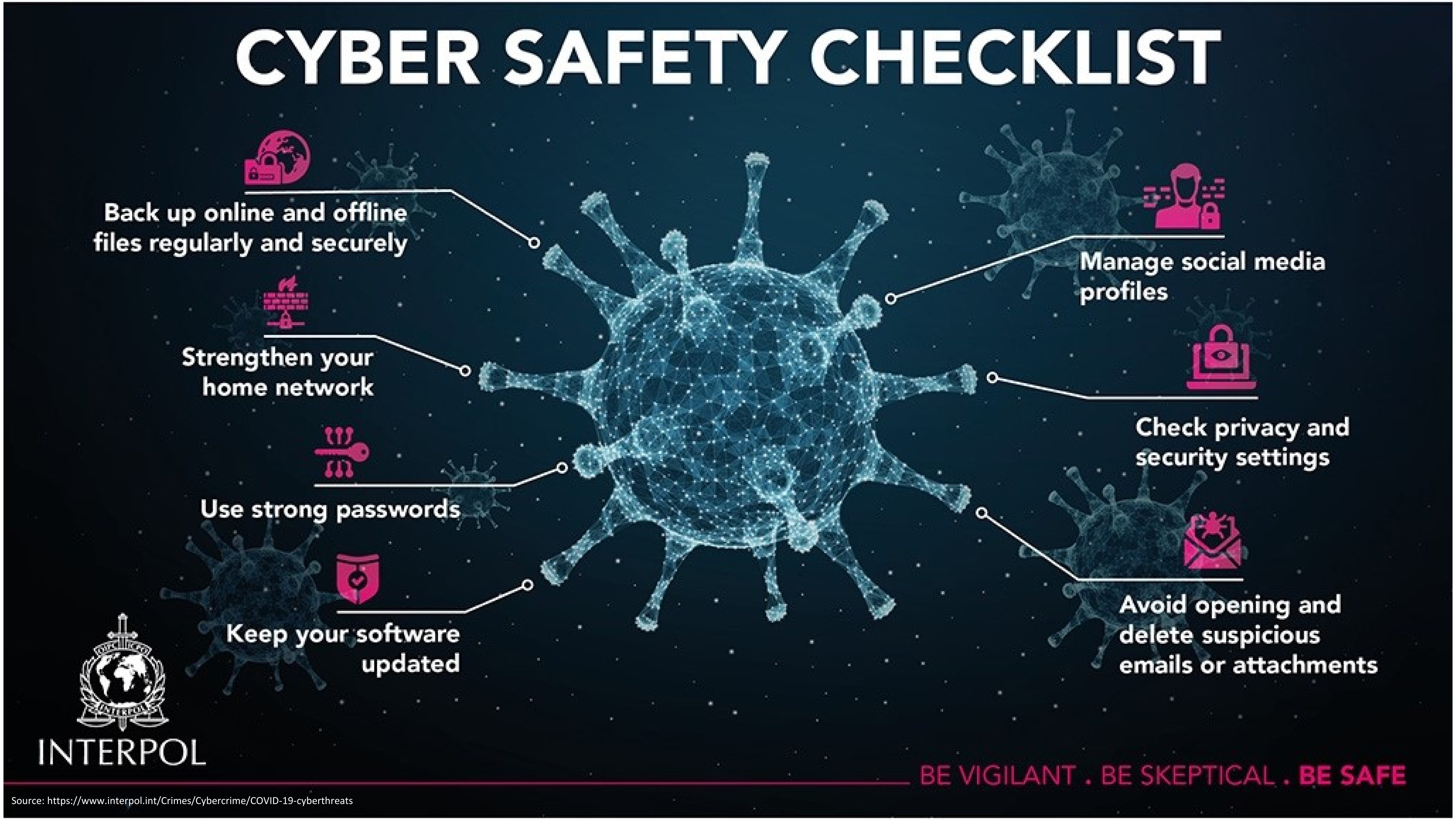
- Data security and appropriate database configuration is not only about NoSQL, which is typically considered to be much less secured, but also about RDBMS. This study has shown that RDBMS are also relatively inferior to various types of vulnerabilities.

- Moreover, there is no “secure by design” database, which is not surprising since it is absolute security is known to be impossible. However, this does not mean that actions should not be taken to improve it - it should be a continuous process consisting of a set of interrelated steps, sometimes characterized as “reveal-prioritize-remediate”.

- 85% of breaches in 2021 were due to a human factor, with social engineering recognized as the most popular pattern (Verizon, 2021)➔ even in the case of highly developed and mature data and system protection mechanism, the human factor remains difficult to control ➔ education and training of system users regarding digital literacy, as well as the definition, implementation and maintaining security policies and risk management strategy, must complement various technical advances.

All in all, cyber hygiene is the answer!!!

Source: Verizon. 2021 Data Breach Investigations Report (DBIR). 2021. 119 Pages, https://www.verizon.com/business/resources/reports/2021/2021-data-breach-investigations-report.pdf, last accessed 2022/03/31



CYBER SAFETY CHECKLIST

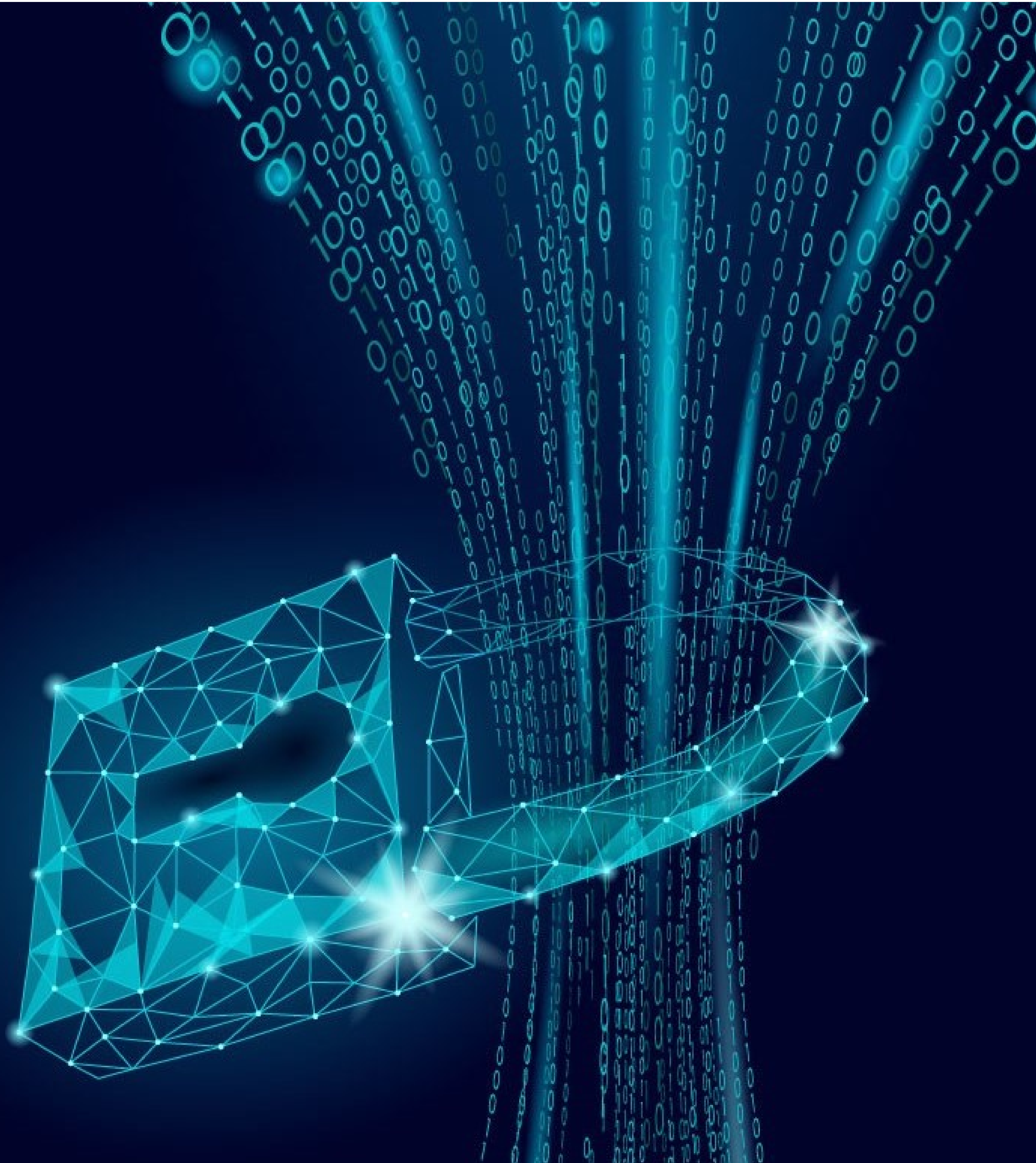
- Back up online and offline files regularly and securely
- Strengthen your home network
- Use strong passwords
- Keep your software updated
- Manage social media profiles
- Check privacy and security settings
- Avoid opening and delete suspicious emails or attachments

BE VIGILANT . BE SKEPTICAL . **BE SAFE**

Source: https://www.interpol.int/Crimes/Cybercrime/COVID-19-cyberthreats

THANK YOU FOR ATTENTION!

QUESTIONS?



For more information, see ***ResearchGate***

See also ***anastasijanikiforova.com***

*For questions or any other queries, contact me via email - **Anastasija.Nikiforova@lu.lv***