International Journal
"Information Models and Analyses"

ISSN 1314-6416 (printed); ISSN 1314-6432 (online)

Edited by the Institute of Information Theories and Applications FOI ITHEA
Published by ITHEA Publishing House, *www.ithea.org*
Sofia, 1000, P.O. Box 775, Bulgaria Phone: (++359 2) 920 19 69 e-mail: *office@ithea.org*

Sofia, 28-10-2019

To KOVAVHEVA, Z.

Dear Kovavheva, Z.

I am glad to inform you that your paper

SOME ASPECTS OF BIG DATA ANALYTICS FOR CYBER-PHYSICAL SYSTEMS (STATISTICAL, MATHEMATICAL, COMPUTATIONAL AND LEGAL VIEW). KOVAVHEVA, Z.

is accepted to be published in the International Journal on Information Models and Analyses (IJ IMA), Vol.8, 2019.

Best regards

Prof. Dr. Krassimir Markov
IJ IMA Editor in chief

Some Aspects of Big Data Analytics for Cyber-Physical Systems (Statistical, Mathematical, Computational and Legal View)

Zlatinka Kovacheva

Abstract*: This paper is focusing on some important aspects of big data analytics for cyber-physical systems from statistical, mathematical, computational and legal point of view. The purpose of the paper is to review the problems generated by the very fast growth of the structured and unstructured data in the modern cyber-physical systems. The main problems from statistical, mathematical, computational and legal point of view have been outlined and some solutions have been discussed.*

Keywords*: Big data, cyber physical systems*

ITHEA Keywords*: D.4.3 File Systems Management*

Introduction

The world is witnessing an unprecedented growth of cyber-physical systems (CPS), which are foreseen to revolutionize our world via creating new services and applications in a variety of sectors such as environmental monitoring, mobile-health systems, intelligent transportation systems and so on. The information and communication technology (ICT) sector is experiencing a significant growth in data traffic, driven by the widespread usage of smartphones, tablets and video streaming, along with the significant growth of sensors deployments that are anticipated in the near future. It is expected to outstandingly increase the growth rate of raw sensed data [Atat at al., 2018].

The aim of this paper is to focus on some important aspects of Big Data Analytics from different points of view:

- Statistical point of view: How to get usable information out of datasets that are too huge and complex for many of the traditional methods to handle?
- Mathematical point of view: How to formalize the nature of Big Data and apply fuzzy logic, neural networks and other mathematical tools?
- Computational point of view: How to solve the problems of data storage and management, communication and computation?
- Legal and ethical point of view: How to ensure the privacy and confidentiality of the data?

1. Some unique characteristics of big data which have to be considered from different points of view

META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Gartner, and now much of the industry, continue to use this "3Vs" model for describing big data [Beyer 2011].

In 2012, Gartner updated its definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." Gartner's definition of the 3Vs is still widely used, and in agreement with a consensual definition that states that "Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" [De Mauro 2016].

According to ORACLE [ORACLE 2016], big data describes a holistic information management strategy that includes and integrates many new types of data and data management alongside traditional data.

Big data has also been defined by the following four Vs [ORACLE 2016] :

- **Volume -** the amount of data. Big data requires processing high volumes of low-density, unstructured data—that is, data of unknown value, such

as Twitter data feeds, click streams on a web page and a mobile app, network traffic, sensor-enabled equipment capturing data at the speed of light, and many more. It is the task of big data to convert such Hadoop data into valuable information. For some organizations, this might be tens of terabytes, for others it may be hundreds of petabytes;

- **Velocity -** The fast rate at which data is received and perhaps acted upon. The highest velocity data normally streams directly into memory versus being written to disk. Some Internet of Things (IoT) applications have health and safety ramifications that require real-time evaluation and action. Other internet-enabled smart products operate in real time or near real time. For example, consumer e-Commerce applications seek to combine mobile device location and personal preferences to make time-sensitive marketing offers. Operationally, mobile application experiences have large user populations, increased network traffic, and the expectation for immediate response;
- **Variety.** New unstructured data types appear nowadays. Unstructured and semi-structured data types, such as text, audio, and video require additional processing to both derive meaning and the supporting metadata. Once understood, unstructured data has many of the same requirements as structured data, such as summarization, lineage, auditability, and privacy. Further complexity arises when data from a known source changes without notice. Frequent or real-time schema changes are an enormous burden for both transaction and analytical environments;
- **Value.** Data has intrinsic value—but it must be discovered. There are a range of quantitative and investigative techniques to derive value from data—from discovering a consumer preference or sentiment, to making a relevant offer by location, or for identifying a piece of equipment that is about to fail. The technological breakthrough is that the cost of data storage and compute has exponentially decreased, thus providing an abundance of data from which statistical analysis on the entire data set versus previously only sample. The technological breakthrough makes much more accurate and precise decisions possible. However, finding

value also requires new discovery processes involving clever and insightful analysts, business users, and executives. The real big data challenge is a human one, which is learning to ask the right questions, recognizing patterns, making informed assumptions, and predicting behavior.

Nowadays some new characteristics such as **variability** and **vitality** can be added.

In [Pyne 2016] we can outline the following specific characteristics of big data which have to be considered:

- Big Data is often nourished by dynamic sources (intense networks of customers, clients and companies) and there is no automatic flow of data that is always available for analysis. There is almost voluntary generation of data;
- The data incorporation in tasks like fitting a suitable statistical model or making a prediction with a required level of confidence is challenging;
- The spontaneous nature of such real- time pro-active data generation can help us to capture complex, dynamic phenomena and enable data-driven decision making process;
- A big data generating mechanism may provide the desired statistical power, but the same may also be the source of limitations;
- Big data have a big potential of being used in unintended manner – e.g. phone records, social networking patterns, etc. The unintended usage of data leads to genuine concerns about individual privacy, data confidentiality and ethics;
- If the data generation costs are low, the people might generate data on as many samples and as many variables as possible;
- The high variety and high dimensionality of big data increase the number of sources of unstructured data such as text, maps, images, audio, video, news, signals and so on.

2. Statistical challenges

From a statistical point of view, the large data could arise in the following cases - either huge numbers of predictors, huge numbers of sample size, or both. New statistical methods like Symbolic Data Analysis and Approximate Stream Regression have been developed. Adopting practices such as dynamic pricing overbooking needs statistical demand forecasting and optimization techniques. The use of statistical techniques on large data bases for business decision making is now a common phenomenon in the developed world and is fast catching up in emerging economies [Pyne 2016].

The methods available now can broadly be categorized into three groups:

- **Divide and conquer method**: First, the original big dataset is divided into K small blocks that are manageable to the current computing facility unit. Then, the intended statistical analysis is performed on each small block. Finally, an appropriate strategy will be used to combine the results from these K blocks. As a result, the computation for the divide and conquer method can easily be done in parallel.
- **Fine to coarse method**: In order to make intended algorithms for the big dataset scalable, statisticians introduced a simple solution: rounding parameters. Hence the continuous real numbers of data are simply rounded from higher decimal places to lower decimal places. A substantial number of observations are degenerated to be identical. This idea was successfully applied to the functional data analysis using smoothing spline ANOVA models.
- **Sampling method**: Another more effective and more general solution for the big data problem is the sampling method. This means that we take a subsample from the original dataset with respect to a carefully designed probability distribution, and use this sample as a surrogate for the original dataset to do model estimation, prediction as well as statistical inference. The most important component for this method is the design of probability distribution for taking the sample.

Overall, the main advantage of the sampling method is its general application to various model settings. Moreover, it will automatically give rise to a random

sketch of the full data as a byproduct, which is useful for the purpose of data visualization. However, the nontrivial part of using sampling method is the construction of sampling probability distribution, which plays a crucial role in sampling methods.

3. Mathematical aspects of Big Data

Mathematicians have also paid increasing attention to the dramatic development of big data and its impact on mathematics through offering courses like mathematics of big data and holding workshops in mathematics of big data.

Zh. Sun and P. Wang [Sun 2017] started to formulate some mathematical foundations of Big Data and the fuzzy logic approach to Big Data. For this purpose, the concept of the discrete mathematics was used to create a theory for big data. The concept of linguistic variables in fuzzy set, and logic theory allow to deal with a higher level of platform for the benefit of a big data theory. In terms of discrete mathematics, they recognize two basic characteristics of big data: the first one is the mathematical operator of being BIG, and the second is to examine the cardinality of big data. Then they provide a mathematical model for searching big data.

Introducing the mathematical operator “BIG”, they discuss its properties. Volume, velocity, variety and veracity are four very basic attributes of data in the world of big data. This can be represented as: data = (volume, velocity, variety, veracity) using the entity-attribute method.

Performing BIG operation “o” on both sides of this equation, they have

o(data) = big data = (o(volume), o(velocity), o(variety), o(veracity))

=(big volume, big velocity, big variety, big veracity)

So-called BIG operation can be considered as an abstraction of technologies, systems and tools of data management and processing that transforms data into big data.

The characteristics of an important concept of ”infinity" was closely associated with the big data, based on the theory of calculus and the set theory.

The relativity of big data is based on the operations of fuzzy subsets theory, recognizing that the linguistic variable “big" is appropriately being modeled as a mathematical linguistic variable suitable for mathematical evaluation.

Fuzzy logic and fuzzy sets have developed a significant methods and techniques to address ambiguity and incompleteness of data, and therefore they will play an important role in overcoming ambiguity and incompleteness of big data.

The other widely used tool for Big Data Analytics is Neural Networks. The layers size (neurons) of the artificial neural network needs to be increased to accommodate the increased dimensions of the input dataset. After certain point, the network size becomes so huge that it becomes almost infeasible to be implemented efficiently because of the increased complexity induced due to the exponential growth of the interconnections among the nodes (neurons) in the network. This phenomenon is generally phrased as the "the curse of dimensionality" in the field of machine learning. Therefore, there is a need to come out with an algorithm to process large dataset efficiently keeping the neural network size considerably small by optimizing the numbers of neurons and the interconnection between them. The future work will be on optimizing the neural networks.

4. Computational solutions

The high amount of traffic driven by the popular use of mobile video and online social media applications along with the scarcity of backhaul resources have pushed researchers and mobile operators to find solutions. Content caching in CPS is of high interest, especially that a big proportion of the traffic load originates from fetching data from different sources such as databases, cache servers and network gateways [Atat at al., 2018].

For computer engineers, a straightforward way to reduce computing time is to resort to more powerful computing facilities. Great efforts have been made to solve the problem of big data by designing supercomputers. Many supercomputers have been built rapidly in the past decade, such as Tianhe-2, Bluewater and Blue Gene. The speed and storage of supercomputers can be hundreds or even thousands of times faster and larger compared to that of a

general-purpose PC. However, the main problem with supercomputers is that they consume enormous energy and are not accessible to ordinary users. Thus, although supercomputers can easily deal with large amounts of data very efficiently, they are still not a panacea.

Instead, cloud computing can partially address this problem and make computing facilities accessible to ordinary users. Nonetheless, the major bottleneck encountered by cloud computing is the inefficiency of transferring data due to the precious low-bandwidth internet uplinks, not to mention the problems of privacy and security concerns during the transfer process. Cloud computing facilitates big data storage, processing and management in CPS, by breaking them down into workflows, which are then distributed over multiple dedicated servers. This allows CPS to provide pervasive sensing services beyond the capacities of individual things, in addition to lower latency and power consumption and larger scalability [Atat at al., 2018].

Cloud computing techniques along with machine learning tools, data mining, artificial intelligence, and fog computing can help the sensed data to be easily stored, processed, and analyzed to uncover hidden patterns, unknown correlations and other useful information [Tsai C. at al. 2014]. That is why big data are referred to as “the 21st century new oil”.

Cloud computing along with data clustering facilitates the parallel processing and execution of tasks and queries. Mapping and scheduling workflows in a multi-cloud environment speeds up the processing and allows for a better big data management. With the large volume of data in the order of exabyte, it becomes almost impractical to process the data on individual machines, no matter how powerful they are. Parallel processing of the data chunks on dedicated servers, such as MapReduce tool proposed by Google, offers advantages over conventional processing methods; however it is still not very effective to handle a large amount of data, mainly due to scalability, latency, availability, and inefficient programming techniques, including but not limited to database management systems [Kraska 2013], [DeWitt 2017].

One attractive solution to dedicated servers is the processing on cloud centers, which offers users the ability to rent computing and storage resources in a

pay-as-you-go manner [Wang 2015]. In addition, even though users will be sharing a common hardware, the shared resources appear exclusive to them through machine virtualization via hiding the platform details [Armbrust at al, 2010]. However, this approach can create problems in the pay-as-you-go environment due to untruthfulness, unfairness and inefficiency of resources and workload transactions [Tang 2014].

Even though cloud computing is an attractive analytics tool for big data applications, it comes with several challenges, mainly concerning security, privacy and data ownership [Atat at al., 2018].

Data clustering refers to partitioning a set of objects comprising of attributes into different groups of similar objects and features. Data clustering becomes very useful in big data applications, where there is a high need to process and analyze large volume of data. Estimating the number of clusters becomes important as clustering facilitates the distribution of the data storage, tasks execution, parallel computing, and queries requests

[Atat at al, 2018].

Another relatively new computational facility proposed is the graphic processing unit (GPU), which is powerful on parallel computing. However, a recently conducted comparison found that even high-end GPUs are sometimes outperformed by general-purpose multi-core processors, mainly due to the huge data transferring time.

In brief, none of the supercomputer, the cloud computing, GPUs solves the big data problem efficiently at this point. Efficient statistical solutions are required, which makes big data problem manageable on general-purpose PCs.

5. Big Data Analytical Tools

After the data are transformed into manageable sizes, data mining tools (e.g. HDFS, MapReduce, R, S), real-time big data analytic tools (e.g. Storm, S-plunk), and cloud-based big data analytic tools (e.g. GFS, BigTable, MapReduce) can be used to extract useful information and make sense of data, which would revolutionize the field of smart cities, environmental monitoring and others [Atat at al, 2018].

We can focus on some typical tools used for the three methods of big data analytics: data mining, real-time big data analytics and cloud-based big data analytics.

5.1. Tools for Data Mining

Apache Hadoop is a popular solution to big data problems. It is of an open-source software framework for distributed storage and distributed processing of very large data sets on computer clusters. The core of Apache Hadoop consists of a storage part (Hadoop Distributed File System - HDFS) and a processing part (MapReduce). Hadoop splits files into large blocks and distributes them amongst the nodes in the cluster [Kovacheva, 2017].

HDFS is developed from an inspiration of GFS, and it is a scalable and distributed storage system, which is an appropriate solution for data-intensive applications, such as Gigabyte and Terabyte scale.

To process the data, Hadoop MapReduce transfers packaged code for nodes to process in parallel, based on the data each node needs to process. The term MapReduce actually refers to two separate and distinct tasks that Hadoop programs perform. The first is the map job, which takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). The reduce job takes the output from a map as input and combines those data tuples into a smaller set of tuples [Kovacheva, 2017.

R is also an open-source software environment for data mining developed by AT&T Bell Labs [Fan, 2013]. Actually, R is a realization of the S language used to explore data, implement statistical analysis and draw plots. Compared with S, R is more popular and supported by a large number of database manufacturers, such as Teradata and Oracle.

5.2. Tools for Real-Time Big Data Analytics

Storm [Fan, 2013] is a distributed real-time computing system for big data analysis. Compared with Hadoop, Storm is easier to operate and more scalable to provide competitive and efficient services. Storm makes use of distinct topologies for different storm tasks in terms of storm clusters, which are composed of master nodes and worker nodes. The master

nodes and worker nodes play two kinds of roles in the fields of big data analysis, namely nimbus and supervisor, respectively. The functions of these two roles are in agreement with jobtracker and tasktracker of the MapReduce framework [Atat at al, 2018].

Nimbus takes charge of code distribution across the storm cluster, the schedule and assignment of worker nodes tasks, and the whole system surveillance. The upervisor compiles tasks given by nimbus. Splunk [Zadrozny, 2013] is also a real-time platform designed for big data analytics. Based on the web interface, Splunk is available to search, monitor and analyze machine-generated big data, and the results are exhibited in different varieties including graphs, reports, alerts and so on. Compared with other real-time analytical tools, Splunk provides various smart services for commercial operations, system problem diagnosis, and so on [Atat at al, 2018].

5.3. Tools for Cloud-Based Big Data Analytics

The most popular tool for cloud-based big data analytics is Google's cloud computing platform. GFS is a distributed file system and it is enchanced to meet the requirements of big data storage and usage demands of Google Inc. In order to deal with the commodity component failure problem, GFS facilitates continuous surveillance, errors detection and component faults tolerance. GFS adopts clustered approach that divides data chunks into 64-KB blocks and stores a 32-bit hecksum for each block. BigTable supplies highly adaptable, reliable, applicable and dynamic control and management in the field of big data placement, representation, indexing and clustering for enormous and distributed commodity servers, and it constitutes of a row, column, record tablet and time stamp [Atat at al, 2018].

6. Legal and ethics aspects of Big Data

In the realm of cyber physical systems, the tight interaction among physical objects which collect and transmit a large volume of data place security threats under the spotlight of attention. With the enormous amount of data that are constantly flowing through the network, it becomes essential to protect the system from cyber attacks.

While data storage in the cloud offers several advantages in terms of data storage, availability, scalability and processing, it increases the chance of malicious attacks, that in addition to potential privacy invasion by cloud operators who can have accesses to sensitive data. All this puts a question mark on whether cloud data storage is feasible, especially for governmental agencies and financial industries [Atat at al, 2018].

Several works have attempted to solve the security challenges of cloud storage. For instance, Gai et al. proposed a method that splits files into encrypted parts and store them in distributed cloud servers without users' data being directly reached by cloud service operators [Gai 2016]. In [Kang 2016], the authors optimized the data placement on cloud servers that minimizes the retrieval time of data files while guaranteeing their security based on the distance between nodes that store the data chunks, such that the malicious attacker cannot guess the locations of all the data chunks. In [Sekar 2016], the authors suggested that data should be encrypted and decrypted before being sent to clouds [Atat at al, 2018].

We can outline the following critical factors for the success of the Big Data projects:

- Understanding the legal framework for Big Data and how it applies to the organization concerned;
- Effectively bringing together the organization's IT and legal functions in the Big Data project;
- A clear understanding of the organization's objectives for its Big Data operations;
- A structured approach to the strategy, policy and process aspects of Big Data governance.

Big data and intellectual property rights and data protection are the key issues which have to be considered.

General Data Protection Regulation ((EU 2016/679) (GDPR) (NIS Regulations) (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679)

came into effect on 25 May 2018 pursuant to the Cyber security Directive (also known as the Network and Information Security or NIS Directive).

Any organizational conversation about big data ethics should relate to four basic principles that can lead to the establishment of big data norms:

- Privacy;
- Confidentiality;
- Transparency;
- Identity.

Enabling security and privacy aspects of big data analytics has attracted a great attention from the scientific community mainly due to different reasons. First, the data are more likely stored, processed and analyzed in several cloud centers leading to security issues due to the the random locations of data. Second, big data analytics treats sensitive data in similar way to other data without taking security measures such as encryption or blind processing into consideration [Gahi 2016]. Third, big data computations need to be protected from malicious attacks in order to preserve the integrity of the extracted results. In the realm of CPS, an enormous amount of data make the surveillance of security-related information for anomaly detection a challenging task for analysts. In healthcare, for instance, the security issues of information extraction from massive amount of data and accurate analytics are of high importance. Sensitive data recorded in databases need to be protected via monitoring which applications and users get accesses to the data [Rao 2015]. In order to guarantee a strong secure big data analytics, the following tasks can be performed [Mahmood 2013]:

- Surveillance and monitoring of real-time data streams;
- Implementation of advanced security controls such additional authentication and blocking suspicious transactions;
- Anomaly detection in behavior, usage, access and network traffic;
- Defending the system against malicious attacks in real time;

– Adoption of visualization techniques that give a full overview of network problems and progress in real time.

Conclusion

As a result of unpredictable growth of data, generated by many sources, nowadays, the attention of the managers, scientists, analysts, etc. is more and more directed to the methods for retrieval and analyzing the relevant and useful information instead of data gathering and storage. This paper considers a very large subject area related to the problems concerning the Big Data Analytics for Cyber Physical Systems. The aim of the presented review is to focus on the most important aspects of this subject from different points of view. Some solutions of the problems have been pointed out and discussed. It will be useful for future research and more detailed analyses of the problems.

Acknowledgements

This paper is partially supported by the Task 1.2.5 of the Bulgarian National Scientific Program "ICT in Science, Education and Security", funded by the Ministry of Education and Science (MES) (Contract MES DOI-205/23.11.2018).

Bibliography

[Atat at al., 2018] Atat R. Liu L. Wu J. at al., Big Data Meet Cyber-Physical Systems: A Panoramic Survey. DOI: 10.1109/ACCESS.2018.2878681

https://www.researchgate.net/publication/328924647_Big_Data_Meet_Cyber-Physical_Systems_A_Panoramic_Survey [accessed Sep 30 2019]

[Beyer 2011] Beyer, Mark, Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data, *Gartner.* Archived *from the original on 10 July 2011*. Retrieved 13 July 2011.

[De Mauro 2016] De Mauro Andrea, Greco Marco, Grimaldi Michele, A Formal definition of Big Data based on its essential Features, Library Review *65: 122–135.* doi:10.1108/LR-06-2015-0061.

[ORACLE 2016], An Enterprise Architect's Guide to Big Data, Reference Architecture Overview, ORACLE Enterprise Architecture white paper, March

2016: https://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf

[Pyne 2016] Pyne S., Rao B.L.S. Rao, S.B. (Eds), Big Data Analytics, Methods and Applications, Springer 2016:

https://www.springer.com/us/book/9788132236269

[Sun 2017], Sun Zh., Wang P., A Mathematical Foundation of Big Data, New Mathematics and Natural Computation 13(02)p July 2017, 83-99, DOI: 10. 1142/S1793005717400014:

https://www.researchgate.net/publication/309035353_A_Mathematical_Foundation_of_Big_Data

[Tsai C. at al. 2014] C. W. Tsai, C. F. Lai, M. C. Chiang, and L. T. Yang, "Data mining for internet of things: A survey," IEEE Communications Surveys Tutorials, vol. 16, no. 1, pp. 77–97, First 2014

[Kraska 2013] T. Kraska, "Finding the needle in the big data systems haystack," IEEE Internet Computing, vol. 17, no. 1, pp. 84–86, Jan 2013.

[DeWitt 2017] D. J. DeWitt and M. Stonebraker. Mapreduce: A major step backwards. Accessed on July 10, 2017. [Online]: http://databasecolumn.vertica.com/database-innovation/mapreduce-a-major-step-backwards

[Wang 2015] D. Wang and J. Liu, "Optimizing big data processing performance in the public cloud: opportunities and approaches," IEEE Network, vol. 29, no. 5, pp. 31–35, September 2015.

[Armbrust at al, 2010] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, Apr. 2010. (Online): http://doi.acm.org/10.1145/1721654.1721672

[Tang 2014] S. Tang, B. S. Lee, and B. He, "Towards economic fairness for big data processing in pay-as-you-go cloud computing," in Proc. 2014 IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom), Dec 2014, pp. 638–643

[Kovacheva, 2017] Kovacheva Zl., Naydenova, I., Kaloyanova, K., Markov, Kr.. Big Data Mining: In-Database Oracle Data Mining over Hadoop. AIP Conference Proceedings of ICNAAM 2016 (Rhodes, Greece), 1863, 1, American Institute of Physics, 2017, ISBN:978-0-7354-1538-6, ISSN:0094-243X, DOI:10.1063/1.4992195

[Fan, 2013] W. Fan and A. Bifet, “Mining big data: Current status, and forecast to the future,” SIGKDD Explor. Newsl., vol. 14, no. 2, pp. 1–5, Apr. 2013. (Online): http://doi.acm.org/10.1145/2481244.2481246

[Zadrozny, 2013] P. Zadrozny and R. Kodali, Big data analytics using Splunk: Deriving operational intelligence from social media, machine data, existing data warehouses, and other real-time streaming sources, Apress, 2013

[Gai 2016] K. Gai, M. Qiu, and H. Zhao, “Security-aware efficient mass distributed storage approach for cloud systems in big data,” in Proc. 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), April 2016, pp. 140–145

[Kang 2016] S. Kang, B. Veeravalli, and K. M. M. Aung, “A security-aware data placement mechanism for big data cloud storage systems,” in Proc. 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), April 2016, pp. 327–332

[Sekar 2016] K. Sekar and M. Padmavathamma, “Comparative study of encryption algorithm over big data in cloud systems,” in Proc. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIA-Com), March 2016, pp. 1571–1574

[Gahi 2016] Y. Gahi, M. Guennoun, and H. T. Mouftah, “Big data analytics: Security and privacy challenges,” in 2016 IEEE Symposium on Computers and Communication (ISCC), June 2016, pp. 952–957

[Rao 2015] S. Rao, S. N. Suma, and M. Sunitha, “Security solutions for big data analytics in healthcare,” in Advances in Computing and communication Engineering (ICACCE), 2015 Second International Conference on, May 2015, pp. 510–514

[Mahmood 2013] T. Mahmood and U. Afzal, “Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools,” in Information Assurance (NCIA), 2013 2nd National Conference on, Dec 2013, pp.129–13

Authors' Information

Zlatinka Kovacheva – *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences; University of Mining and Geology, Sofia, Bulgaria; e-mail: zkovacheva@math.bas.bg*

Major Fields of Scientific Research: Neural networks, Big data