ISSN 2226-3780

9. Eeckhoudt L., Gollier C., Schlesinger H. Economic and financial decisions under uncertainty. Princeton: Princeton University Press, 2004. 248 p.
10. Penman S. H. Financial Forecasting, Risk and Valuation: Accounting for the Future. Abacus. 2010. Vol. 46, No. 2. P. 211–228. doi: http://doi.org/10.1111/j.1467-6281.2010.00316.x
11. Clements M. P., Franses P. H., Swanson N. R. Forecasting economic and financial time-series with non-linear models // International Journal of Forecasting. 2004. Vol. 20, No. 2. P. 169–183. doi: http://doi.org/10.1016/j.ijforecast.2003.10.004
12. Sekerke M. Bayesian Risk Management: A Guide To Model Risk And Sequential Learning In Financial Markets. Hoboken: John Wiley & Sons, Inc., 2015. 240 p. doi: http://doi.org/10.1002/9781118864784
13. Bidyuk P., Gozhyj O., Korshevnyuk L. Development of Decision Support Systems. Mykolaiv: Petro Mogyla Black Sea National University, 2012. 379 p.
14. Petersen I. R., Savkin A. V. Robust Kalman filtering for signals and systems with large uncertainties. Boston: Birkhouser, 1999. 202 p. doi: http://doi.org/10.1007/978-1-4612-1594-3
15. Statistical publication // The official website of the State Statistics Service of Ukraine. 2016. URL: http://www.ukrstat.gov.ua/

Bidyuk Petro, *Doctor of Technical Sciences, Professor, Department of Mathematical Methods of Systems Analysis, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, ORCID: https://orcid.org/0000-0002-7421-3565, e-mail: pbidyuke_00@ukr.net*

Prosyankina-Zharova Tatyana, *PhD, Department of Physical and Mathematical Modelling, Institute of Telecommunications and Global Information Space, Kyiv, Ukraine, ORCID: https://orcid.org/0000-0002-9623-8771, e-mail: t.pruman@gmail.com*

Terentiev Oleksandr, *PhD, Junior Researcher, Department of Mathematical Methods of Systems Analysis, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, ORCID: https://orcid.org/0000-0002-4288-1753, e-mail: o.terentiev@gmail.com*

Medvedieva Mariia, *PhD, Associate Professor, Head of the Department, Department of Informatics and Information and Communication Technologies, Pavlo Tychyna Uman State Pedagogical Universit, Uman, Cherkasy region, Ukraine, ORCID: https://orcid.org/0000-0001-9330-5185, e-mail: medvedeva-masha25@ukr.net*

UDC 004.915
DOI: 10.15587/2312-8372.2018.141299

Petrenko A.,
Kyslyi R.,
Pysmennyi I.

DESIGNING SECURITY OF PERSONAL DATA IN DISTRIBUTED HEALTH CARE PLATFORM

Об'єктом дослідження є розробка системи електронної медкарти (EHR), призначеної одночасно як для взаємодії пацієнт-лікуючий лікар, так і для обміну анонімізованими даними між різними медичними організаціями для їх подальшої обробки та побудови аналітичних моделей. Постійний моніторинг стану пацієнта, а також кількість і якість оброблених даних є ключовими факторами, що впливають на точність постановки діагнозу і подальші лікарські рекомендації. Слід зауважити, що більшість сучасних підходів до проектування EHR-систем є вразливими до атак цілісності даних і не дозволяють обмінюватися інформацією з іншими організаціями, зберігаючи при цьому лікарську таємницю, що призводить до наявності у окремих акторів лише невеликих фрагментованих датасетів. Важливим напрямком для поліпшення існуючих рішень є безпека обміну інформацією між натільними смарт-сенсорами.

У даній роботі пропонується розбити архітектуру на шари з виділеними зонами безпеки. Ця фрагментація дозволяє ефективно сегментувати інфраструктуру, дозволяючи кожному елементу застосовувати свої власні вимоги до аутентифікації і авторизації, використовуючи різні підходи до захисту інформації. Додатковим ефектом цього підходу є зниження навантаження на мережу і уникнення проблем безпеки шляхом мінімізації передачі конфіденційних даних (наприклад, проводити базовий збір та обробку даних на смарт-сенсорах). Пропонується використання блокчейн-технологій для забезпечення цілісності даних з використанням офф-чейн бази даних для оптимізації зберігання та швидкості транзакцій. Застосування MPC-протоколу дозволяє обмінюватися даними між партнерськими організаціями для спільних розрахунків і навчання ml-моделей, не показуючи фактичні дані.

Пропоновані підходи дозволяють створювати надійну, гнучку і в той же час безпечну платформу для збору конфіденційних даних, їх аналізу і обробки розподіленою багатоакторною системою, використовуючи переваги туманних обчислень, блокчейна та MPC.

Ключові слова: *безпека електронної карти пацієнта, блокчейн в медицині, безпека особистих даних, безпека мереж натільних датчиків.*

1. Introduction

Fast development of different wearable health monitoring and tracking capable devices allows collection and processing of patient's data during his everyday activity. By integrating these data sources with Personal Health Record (PHR) systems and partly automating inferences and decision-making process for physician assistance that can be applied to the new connected e-Health paradigm. It allows simultaneous integration of PHRs information, body sensor networks' (BSN) streams, activity data and context for better outcomes in preventive health as well as

chronic or follow-up care disease management and monitoring. This comes with a cost of exposing big amounts of sensitive personal health information to external processing facilities and therefore, a lot of data transfer and storing. Each of these operations is vulnerable to different set of threats which will be discussed in detail in following section with possible solutions such as blockchain and secure MPC examined in section 6. Worth mentioning, blockchain implementation in different areas, including IoT, is widely supported by big industry players like IBM or Amazon. These platforms are built for different purposes but also share much common features. For example, data from different devices can be included into private blockchain ledgers that at the same time can be in one shared transaction. To add new data, smart contracts are used, so data from devices comes to API and then, only data that is related to the transaction and specific contract requirements is included.

In the given scope challenges can be divided into security (protection from malicious leaks and modifications of data during transfer and storage) and privacy (access to the information is given only to the authorized parties) domains [1].

Considering all above, research of security options for healthcare systems in general and usage of blockchain for data storage in particular is highly actual at the current stage of automation and integration of patient-doctor interaction.

2. The object of research and its technological audit

The object of research is the design of EHR system with preservation of sensitive user's data privacy in distributed big data processing environment with maintaining system flexibility and scalability.

When designing a system, especially with sensitive data, it is important to understand the potential threats to that system and use appropriate defence techniques. Security model of the product need to be designed at the very beginning of the development, because finding, eliminating and preventing appearance of possible ways by which malicious adversary might be able to break into the system helps to secure from data leaks. For proper utilization of common security architecture best practices, applying layered approach to designing system architecture as well as grouping into the security zones is suggested [2]. These zones are listed below and effectively apply to the layered infrastructure discussed in research results section:

– device;
– field gateway;
– cloud gateways;
– protocol of data transfer.

This fragmentation allows to effectively segment infrastructure, allowing each element to apply to its own authentication and authorization requirements as well as secure data individually. Positive side effect of this approach is failure and breach isolation and restricting it's consequences to particular trust level. Inside each of these zones next types of attacks can be highlighted [2]:

– dpoofing;
– denial of Service;
– tampering;
– information Disclosure;
– elevation of Privilege.

Breaches listed above can cause leaks of sensitive data bound with user identity with tampering might lead to mistreatment resulting in user's severe injuries in addition to legislative responsibility as health care data is protected by governmental regulation acts.

Problems listed above rise only for classical layered architecture design model that requires cloud-based back-end. Usage of blockchain and server-less back-end can solve several issues [3]:

– Reduce costs and capacity requirements relinquishing from the centralized entity as devices can securely exchange information and value with each other, and execute actions via smart contracts with confidence of computed result's and other input's validity.
– Security of communication between nodes as all transactions are cryptographically encoded and signed ensuring sender device's identity and protecting against man-in-the-middle attacks. This can also helps to protect against distributed DOS attacks as messages from the unauthorized and unsigned origins can be rejected by protocol (proof-of-work concept is used).
– Replication of records between different nodes in the network eliminates the single failure point threat and, as a result, system is safe from downtime or data losses.

3. The aim and objectives of research

The aim of research is development of EHR platform architecture with respect to users' data privacy as well as implementing modern data analyses capabilities.

Achieving given aim can be split into following tasks:

1. Designing application layers and isolating processes inside them.
2. Preserving data integrity and security during multiparty computations.
3. Implementation of hyperledger.

4. Research of existing solutions of the problem

Due to its practical importance subject is an active research area for scientific and commercial teams. Ability to automate and analyse PHR data flow brings questions of patient records' privacy and security [4]. Article questions the whole way of treating health records («medical vs patient record») also arguing that badly anonymized data used in researches can lead to sensitive information leakage. Paper [5] is rising security and privacy issues about BSN usage in e-health. To overcome specified threats authors propose usage of authorization and encryption which is not enough in the case that data is shared between different parties for analysis. Further research in wireless sensor security is performed in [1]. Authors focus on different security weaknesses of body networks dividing them into threats from device compromise and threats from network dynamics and argue requirements for distributed data storage security.

In [6] authors describe recent security breaches which result in sensitive data loss with generic advises acquired from them but without proposing any technical platform requirements to prevent future hacks.

In paper [7] authors suggest usage of blockchain as a standard decentralized security framework. [8] is further

ISSN 2226-3780

extending this thesis and applying it to m-health by providing a proof-of-concept system, demonstrating how principles of decentralization and blockchain architectures could contribute to secure, interoperable EHR systems using Etherum smart contracts.

Usage of trusted external verification utility allows to significantly reduce development time and provides good security level but is not intended for a use on the network edge in BSN.

[9] addresses different challenges of Hyperledger usage in medicine such as lack of anonymity, speed and scalability, vulnerability to «51 % attack». Authors suggest different approaches for solving raised issues including sharing only already trained predictive models between parties (which vastly reduces potential outcomes' accuracy) and moving data off-chain without implementation details. Storage capacity advances of moving actual data off-chain are described in [10]. Implementation of off-chain database to optimize blockchain transactions was suggested by [11]. Authors propose technology called Enigma which allows to eliminate the need of trusted third party. Enigma has a decentralized off-chain distributed hash-table that makes use of blockchains to store data references (not to the actual data) [12].

Another set of challenges is distributing data for computations and exchanging it with different parties. In [13] authors advocate Arithmetic Cryptographic Protocol that allows joint computation without compromising with the privacy and confidentiality of individual and provides data security and privacy along with the confirmation for the correctness of results, but doesn't specify low-level implementation details and run-time measurements.

Authors of paper [14] suggest method for MPC of linear regression on high-dimensional data. Proposed protocol implementation allows operation even in semi-honest environments and can be used as a component in larger system.

Thus, results of the analysis let get to the conclusion that potential leakage of sensitive data can be a big threat to the automation of patient-doctor interaction. However, proper usage of blockchain technology can be used to minimize these threats and protect patients' data.

5. Methods of research

Different scientific methods were used during the research process:

– analysis, while analysing data sources and previous solutions. Analysis of the previous solution was used to identify potential problems of already created systems and solutions;

– empirical, while trying to solve founded security gaps and testing solutions. An at this point hypothesis testing was done to prove concepts and proposed solutions;

– classification, while classifying data on different layers. In proposed solutions, data is classified in sensitive and not sensitive, so it can be divided and put into different layers of the storage;

– proof of work was used while testing system, based on the blockchain database to put data into Hyperledger layer;

– data encryption, while putting data into blockchain database and for securing MPC.

6. Research results

6.1. System architecture design. The most vulnerable part of every modern system is data transfer as it is performed through public networks which might be compromised and are vulnerable to man-in-the-middle security attacks.

To isolate sensitive information transfer and prevent cloud storage of de-anonymized data splitting architecture into loosely coupled layers with strong cohesion inside each level is suggested. This approach forces user to partly process private data on the edge of the infrastructure and expose only generalized and anonymised data to the cloud storage system.

System consists of following layers as shown on Fig. 1:

– Expert Layer is the physician assigned to patient, capable for assigning treatment approving analysis layer's recommendations.

– Cloud layer – capable for data gathering, storage, analysis and integration with external data. While being scalable, cost and productivity-efficient, enhancing ease-of-access to the information it is very vulnerable to most kinds of attacks from denial-of-service (DDOS) to person-directed hacking [15].

– Private client's layer consisting of mist-computing BSN layer, cloud gateway and user interaction (UI) layer. These sub-layers are capable, namely, for processing sensor output and retrieving context, integrating with cloud and providing feedback to and from the patient.

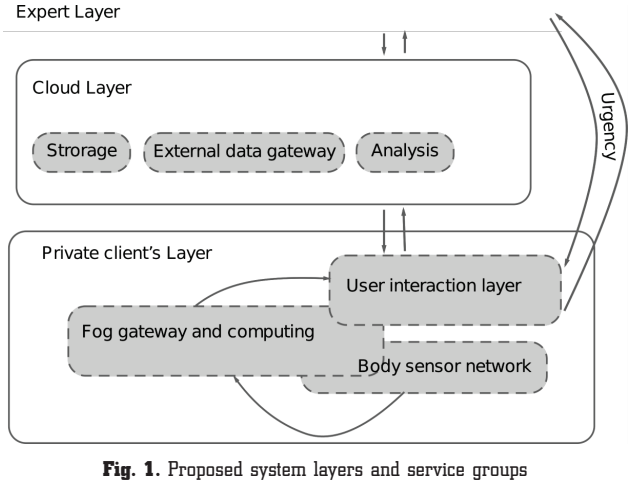

Fig. 1. Proposed system layers and service groups

User's smartphone is usually used as cloud gateway as it provides availability of internet connection with cloud layer and Bluetooth for communication with sensor networks coupled with enough computing power for data processing and is always aware of user's location and other context data.

Basic data gathering and processing is being performed on the smart sensors reducing network load and eliminating security and privacy issues. Moreover, modern smart phones can even take part in analysis process [16]. For example the pre trained deep learning model allows to acquire human breath frequency only from one accelerometer sensor output with 90 percent probability which can assist in asthmatic attack detection and prevention. With most edge devices being context-aware, this opens wide range of possibilities for further integration with smart-city health care modules (e. g. automatically call

closest ambulance on heart attack event, find the drug store with the needed treatment available).

Cross-layer communication is performed in hierarchical fashion with an exception for urgent situations (exacerbation of the disease, attacks and other defined events) where user can communicate directly to the assigned expert. Protocol peculiarities for data privacy and security are further described in following chapter.

6.2. Trusted cross-organization multi-party computations. Sometimes joint calculations by akin organizations must be performed with the user's privacy being preserved [13]. Several most common examples are listed below:

– Trust funds or insurance companies may want to ensure patient's honesty.

– Need to check if the treatment is eligible for pavement from the insurance company without revealing the whole health record (the problem known as «Private data joins»).

– Need to transfer patient records to external facilities for treatment rectification (for example, getting predictions and recommendations based on DNA test data) with preserving data privacy.

– Gathering data for statistical analysis and training of deep learning and machine learning models.

In these situations, each party is interested in computational result without revealing its data to the others.

In Fig. 2 the common approach to this problem is shown. Each of the parties share encrypted data with Trusted Third Party (TTP) which sends computed outputs back.

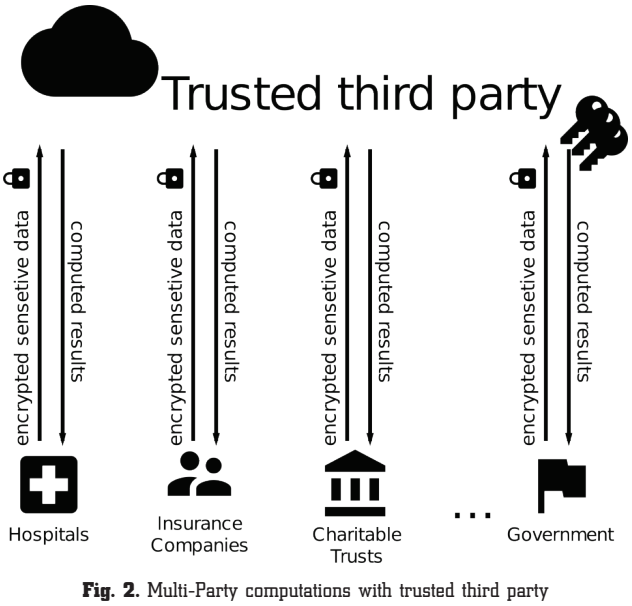

Fig. 2. Multi-Party computations with trusted third party

Usage of Private Set Intersection cryptographic protocols, for example, using garbled circuits, can help to eliminate the need of TTP relying verification and anonymization of transferred data on the protocol itself [17]. This approach is described in Fig. 3.

In [14] expands the concept even further arguing that MPC approach can be applied to data mining process by suggesting and implementing protocol for «Privacy-Preserving Distributed Linear Regression on High-Dimensional Data» can be applied.

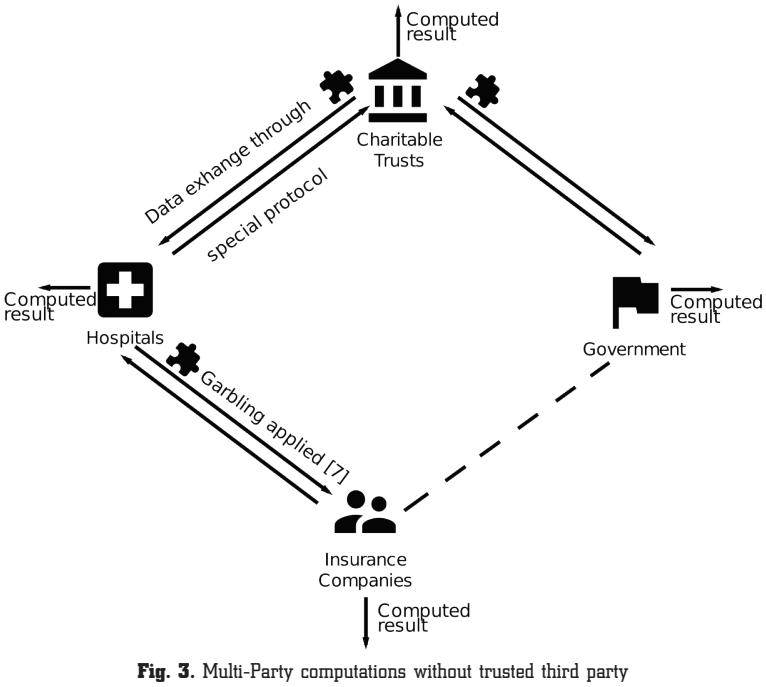

Fig. 3. Multi-Party computations without trusted third party

MPC approach can also be used on system edge level resulting in building effective distributed mist-computing network and utilizing maximum of limited smart sensor's computing power [12].

6.3. Blockchain and off-chain database. Hyperledger is a blockchain framework that is often used for production systems. However even it has its own limitations – as much more data is added to the chain, the more time transaction takes. It was counted, that using public blockchain from IBM, it takes around 40 seconds to finish 100 transactions in parallel with 96 % utilization of nodes CPU [18], even considering that PBFT consensus protocol is used.

As the system for healthcare records storage needs to be available for most people (in ideal case for everyone), such resource consumption may be a huge bottleneck.

To overcome this, it is proposed to add some kind of load-balancer that will create a queue of incoming transactions and maintain suitable level of resource consumption. This will create following advantages:

– No transaction will be lost (all transactions moving through the queue in FIFO mode).

– Highly predictable usage of infrastructure.

– Avoiding the bottleneck of running out of memory on the nodes.

At the same time, to maintain data consistent and to trust source of incoming transaction, let's propose to verify source of transaction and k-nearest nodes with double handshake, also comparing timestamps of the transactions.

As medical data is extremely sensitive, it is not the thing that has to be available for everyone on distributed

ISSN 2226-3780

system, so it is making sense to store only metadata in the blockchain (like digest of the data for verification and reference to the actual storage), while keeping data records itself in the trusted place. Also, worth mentioning that storing large volumes of data in blockchain will enormously increase it and make very expensive to store the whole chain, even with the fact that storage space becomes cheaper through the time. Even it is feasible to store that huge amounts of data, another problem will arise – the bigger chain will be, the slower system will work (as block size and it's computational time is defined, there will be less new transactions in every new block) [10].

But pushing some data into off-chain DB will also cause some troubles, main of which is losing ACID (Atomicity, Consistency, Isolation, Durability) guarantee. As it is known, main idea of blockchain is that everyone keeps the same data, so no one can change it. When data is pushed out of chain, persistence-from-the-box benefits are lost. Of course, saved hash still can be used to verify that data wasn't changed. In case of data alterations proof tail will be available, but in the end it is not so secure as to rely on replicas demanded by the protocol itself, because there is no possibility of reverting data to original state if no backups were made on client level [10].

As it was already discussed, blockchain cannot include all patient's information, especially very sensitive, therefore it has to be stored off-chain. As a result all patient's data is being stored in traditional DB system and traditional methods of data security (like encrypting the whole DB, storage of users personification separately from the data, network access limitations and other methods) must be applied.

7. SWOT analysis of research results

Strengths. Proposed approach of maximizing computations with sensitive data on edge layer helps to reduce private information transfer.

Utilization of blockchain protocol stack bri ngs additional benefits:

– Proof-of-work concept protects system from DDOS attacks.

– System becomes more node-failure tolerant as there is no central node.

– Increasing transparency and audibility of the architecture.

Weaknesses. Proposed approach requires significant computing resources on the edge of the network and therefore is not compatible with some of the current smart sensor solutions. Also, Hyperledger increases required computing and storage requirements for the designed system.

To overcome performance and storage size limitations of blockchain usage of off-chain database and writing only transactions information and data signatures to the ledger is proposed.

Opportunities. Further research in current domain aimed at further enhancement of data transfer and management techniques and consists of following topics:

– Designing MPC protocol for distributed deep learning with privacy preservation (both for cross-organizational and fog levels).

– Keeping system tamper-proof in environment with prevailing malicious adversaries.

– Optimizing deep learning models for running on low-power edge devices in distributed fashion.

– Applying multi-agent paradigm on processing health care data.

Threats. One of the main threats for proposed platform is a legislation modification which would ban processing of anonymised medical data.

8. Conclusions

1. Layered approach for system security design is discussed and implementation for EHR systems with isolated layers and security groups is proposed. As a result of the research, a secure system with layered architecture is proposed. Due to the usage of the blockchain technology, sensitive information stores in the encrypted layer, while its modification needs much efforts and computation power.

2. Described MPC mechanisms for calculations shared by akin organizations with the user's privacy being preserved. While saving data in the Hyperledger, proof of work concept is used. This guarantee that data that came into system during one transaction will be saved in one block of the block-chain. At the same time, this means that data integrity will be kept, because all sensitive patient data that came into the system is put in one transaction.

3. Usage of Hyperledger for data integrity preservation is investigated. To overcome performance and storage size limitations of blockchain usage of off-chain database and writing only transactions information and data signatures to the ledger is proposed. Proposed novel infrastructure usage approaches allow more efficient utilization of computation capabilities.

Suggested solutions considerably improve sensitive health data management and computation security as well as preservation of user's privacy.

References

1. Li M., Lou W., Ren K. Data security and privacy in wireless body area networks // IEEE Wireless Communications. 2010. Vol. 17, Issue 1. P. 51–58. doi: https://doi.org/10.1109/mwc.2010.5416350
2. Internet of Things Security Architecture. URL: https://docs.microsoft.com/en-us/azure/iot-suite/iot-security-architecture
3. Kshetri N. Can Blockchain Strengthen the Internet of Things? // IT Professional. 2017. Vol. 19, Issue 4. P. 68–72. doi: https://doi.org/10.1109/mitp.2017.3051335
4. Anderson R. J. A security policy model for clinical information systems // Proceedings 1996 IEEE Symposium on Security and Privacy. 1996. doi: https://doi.org/10.1109/secpri.1996.502667
5. Al Ameen M., Liu J., Kwak K. Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications // Journal of Medical Systems. 2010. Vol. 36, Issue 1. P. 93–101. doi: https://doi.org/10.1007/s10916-010-9449-4
6. He Y., Johnson C. W. Generic security cases for information system security in healthcare systems // 7th IET International Conference on System Safety, incorporating the Cyber Security Conference 2012. 2012. doi: https://doi.org/10.1049/cp.2012.1507
7. The Blockchain as a Decentralized Security Framework [Future Directions] / Puthal D. et. al. // IEEE Consumer Electronics Magazine. 2018. Vol. 7, Issue 2. P. 18–21. doi: https://doi.org/10.1109/mce.2017.2776459
8. A Case Study for Blockchain in Healthcare: 'MedRec' prototype for electronic health records and medical research data / Ekblaw A. et. al. // White Paper. 2016. 13 p.
9. Kuo T.-T., Kim H.-E., Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications // Journal of the American Medical Informatics Association. 2017. Vol. 24, Issue 6. P. 1211–1220. doi: https://doi.org/10.1093/jamia/ocx068

10. Going off chain for storage. 2017. URL: http://goo.gl/xwauRC
11. Zyskind G., Nathan O., Pentland A. Enigma: Decentralized Computation Platform with Guaranteed Privacy. URL: https://enigma.co/enigma_full.pdf
12. Fog Computing in the Internet of Things / Rahmani A. M., Liljeberg P., Preden J.-S., Jantsch A. (Eds.). Springer, 2018. doi: https://doi.org/10.1007/978-3-319-57639-8
13. Jangde P., Mishra D. K. A Secure Multiparty Computation Solution to Healthcare Frauds and Abuses // 2011 Second International Conference on Intelligent Systems, Modelling and Simulation. 2011. doi: https://doi.org/10.1109/isms.2011.75
14. Privacy-Preserving Distributed Linear Regression on High-Dimensional Data / Gascón A. et. al. // Proceedings on Privacy Enhancing Technologies. 2017. Vol. 20117, Issue 4. P. 345–364. doi: https://doi.org/10.1515/popets-2017-0053
15. Gupta A. K., Mann K. S. Sharing of Medical Information on Cloud Platform-A Review // IOSR Journal of Computer Engineering. 2014. Vol. 16, Issue 2. P. 08–11. doi: https://doi.org/10.9790/0661-16270811
16. Baidu Mobile Deep Learning. URL: https://github.com/baidu/mobile-deep-learning
17. Huang Y., Evans D., Katz J. Private set intersection: Are garbled circuits better than custom protocols // Network and Distributed System Security Symposium (NDSS). 2012. P. 5–8.
18. Towards Blockchain-enabled Wireless Mesh Networks / Selimi M. et. al. // Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems – CryBlock'18. 2018. doi: https://doi.org/10.1145/3211933.3211936

Petrenko Anatolii, *Doctor of Technical Sciences, Professor, Head of the Department, Department of System Design, Institute of Applied Systems Analysis, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, e-mail: tolja.petrenko@gmail.com, ORCID: http://orcid.org/0000-0001-6712-7792*

Kyslyi Roman, *Postgraduate Student, Department of System Design, Institute of Applied Systems Analysis, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, e-mail: kvrware@gmail.com, ORCID: http://orcid.org/0000-0002-8290-9917*

Pysmennyi Ihor, *Postgraduate Student, Department of System Design, Institute of Applied Systems Analysis, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, e-mail: ihor.pismennyy@gmail.com, ORCID: http://orcid.org/0000-0001-7648-2593*