Архитектура информационной системы для проверки подлинности медицинских данных в архиве DICOM

Ш.Г. Магомедов

***Аннотация*— В настоящее время персональные медицинские устройства играют все более важную роль в экосистемах здравоохранения как оборудование для жизнеобеспечения пациентов. И все больший интерес вызывает у злоумышленников программное обеспечение и протоколы, взаимодействующие с данным оборудованием. Наиболее распространённым медицинским протоколом является протокол DICOM, передаваемые посредством которого данные будут рассматриваться в контексте корректности с медицинской точки зрения, чтобы минимизировать возможный вред от поддельных файлов DICOM. Для достижения приемлемой точности на практике учитываются два аспекта: корректность периодичности и корректность самих данных изображения (временного ряда) для рассматриваемой модальности. В данной работе предлагается архитектура информационной системы и интегрированный в нее сетевой фильтр для проверки подлинности данных в архиве DICOM, обеспечивающий возможности анализа и управления предупреждениями. Архитектура предлагаемой системы предназначена как для работы с наборами архивных данных, так и для выполнения анализа входящих потоков больших данных. Компоненты программного обеспечения работают в кластере, что обеспечивает горизонтальную масштабируемость и отказоустойчивость.**

***Ключевые слова*— большие данные, архив DICOM, проверка данных, распределенные системы, фильтрация сетевого трафика.**

I. Введение

Сегодня медицинские учреждения представляют собой обширные экосистемы, состоящие из большого количества сетевых устройств, оборудования и систем, которые часто требуют подключения к внешним системам. Медицинские данные очень чувствительны к изменениям, которые представляют реальную угрозу для здоровья и жизни пациентов. Необязательно обладать специальными навыками, чтобы ознакомиться с потенциальными уязвимостями, с которыми может столкнуться медицинское учреждение. Таким образом, безопасность медицинских данных должна быть обеспечена на каждом этапе получения, передачи, обработки, хранения информации, чтобы обеспечить конфиденциальность данных пациента, а также доступность и устойчивость медицинских услуг одновременно [1]. Исходя из этого, производителям медицинских систем, а также организациям, которые их организуют, необходимо реализовать меры по обеспечению необходимого уровня защиты от киберугроз, что повысит уровень безопасности пациентов и инфраструктуры медицинского учреждения в целом. Рассмотрим три направления защиты медицинских данных.

1. Обеспечение соответствия входящих данных стандарту DICOM. Это должно быть реализовано на стороне сервера - сервер DICOM или компонент фильтрации, работающий в качестве внешнего интерфейса (frontend) для сервера DICOM. Представляют интерес данные IOD (Information Objects Definition). Идея предложить формальный язык для выражения IOD не нова [2]. Современное программное обеспечение, такое как dcm4che [3,17], поддерживает эту проверку.
2. Нанесение водяных знаков на медицинские изображения. Метод, предложенный в [4], основанный на методике обратимого водяного знака, обеспечивает аутентификацию и самокоррекцию путем разделения изображения на две области: интересующая область (ROI) и область не интересующая (RONI). Затем ROI встраивается в RONI, поэтому любое изменение изображения может быть обнаружено и может быть самовосстановлено обратно к исходному изображению путем извлечения ROI из RONI. В работе [5] предлагается метод обеспечения безопасности с аутентификацией пациента, а также обеспечения конфиденциальности и проверки целостности информации на основе обратимого водяного знака. Для проверки целостности вычисляется MD5-хэш изображения. Обратимость достигается с помощью сжатого R–S-вектора, определяемого по изображению. Водяной знак, обеспечивающий конфиденциальность и сервисы аутентификации, создается путем агрегирования сжатого R–S-вектора, хэш-значения и идентификатора пациента. Он зашифрован с помощью AES и встроен в медицинские изображения.
3. Шифрование файлов DICOM. В работе [6] предлагается следующий алгоритм обеспечения конфиденциальности, проверки целостности и аутентичности данных

Магомедов Шамиль Гасангусейнович, кандидат технических наук, доцент, заведующий кафедрой КБ-4 «ИСИБ», МИРЭА – Российский технологический университет. (email: msgg@list.ru)

заголовка и пикселей изображений DICOM: процедура шифрования и создания подписи; расшифровка и процедура проверки подписи. Сингла и Сингх [7] разработали структуру, предлагающую два разных подхода к обеспечению безопасности облачных данных: расширяемый протокол аутентификации и алгоритм шифрования Rijndael, используемый для шифрования конфиденциальных данных. Усама и др. [8] предлагает структуру для безопасной передачи и хранения медицинских изображений в облаке с использованием гибридных (комбинация симметричного и асимметричного) алгоритмов шифрования. Их схема состоит из отдельных этапов хеширования заголовка DICOM с помощью SHA-3 и шифрования данных пикселей с результатом предыдущего этапа с использованием алгоритма XTEA.

Все вышеперечисленные приемы необходимы для обеспечения безопасности данных, передаваемых в архив и хранимых в нем. Поскольку любое программное обеспечение может иметь уязвимости, возникает проблема проверки подлинности медицинских данных: легально ли добавлен файл в архив DICOM, не сфабрикован ли он. Работа организована следующим образом: сначала представлена формализация задачи, в которой обсуждаются два аспекта: правильность периодичности и правильность самих данных изображения (временных рядов) для рассматриваемой модальности; в следующем разделе описывается архитектура информационной системы и предлагаемый сетевой фильтр, интегрированный в нее, чтобы обеспечить средства для анализа и управления предупреждениями; в заключении обсуждаются плюсы и минусы предложенного подхода.

II. Формализация задачи

Рассмотрим ряды данных с периодически отправляемыми в архив отсчетами (измерениями). Они могут генерироваться не только стационарным оборудованием, но и носимыми устройствами, такими как биомедицинские датчики ЭКГ. Аппаратные средства предоставляют данные по запросу программного обеспечения, которое обычно выполняет периодические запросы и отправляет данные в облако [9-11]. Если периодичность входящего потока данных нарушена, высока вероятность атаки, если простая потеря сетевого соединения не имеет места. В следующем подразделе предлагается модель корректности периодичности для входящих потоков данных. Она позволяет иметь некоторое отклонение от фиксированной величины периодичности для моделирования редких проблем с сетевым подключением.

Второй принимаемый во внимание аспект - это подлинность самих медицинских данных [12]. Для любой рассматриваемой модальности должны быть разработаны специальные методы и алгоритмы, чтобы выполнить анализ файлов и оценить возможность фальсификации медицинских данных с правильной меткой времени. Для любой рассматриваемой модальности необходимо разработать метод и алгоритм, предназначенные для обнаружения резкого изменения состояния здоровья пациента [13-16].

III. Модель корректности периодичности

Обозначим $X = \{X_1, X_2, ..., X_1^r, X_2^r, ..., X_n^r, ..., X_1^d, X_2^d, ..., X_m^d, ...\}$ - временной ряд рассматриваемой модальности. Он разделен на две части: контрольную часть (отсчеты обозначены как $X_\bullet^r$) и часть для анализа (отсчеты обозначены как $X_\bullet^d$).

Пусть $time(X_\bullet^r)$ - функция, возвращающая отметку времени отсчета $X_\bullet^r$, переданного в качестве аргумента.

Пусть $t_i^r = time(X_{i+1}^r) - time(X_i^r), i = 1, ..., n-1$ - временной интервал для двух отсчетов в контрольной части.

Тогда средний временной интервал для контрольной части $\frac{1}{n-1}\sum_{i=1}^{n-1} t_i^r$ должен принадлежать $\left[T^r - \Delta^r; T^r + \Delta^r\right]$, где T^r - ожидаемый период, Δ^r - допустимое изменение ожидаемого периода. Обе величины задаются определяемыми пользователем внешними параметрами, представленными в виде положительных значений даты и времени.

Рассмотрим $t_j^d = time(X_{j+1}^d) - time(X_j^d), j = 1, ..., m-1$ как временной интервал двух отсчетов в анализируемой части. Среднее значение $\frac{1}{m-1}\sum_{j=1}^{m-1} t_j^d$ также должно принадлежать $\left[T^r - \Delta^r; T^r + \Delta^r\right]$ и должно выполняться $\forall j \in \{1, 2, ..., m\}\left(k_l \min_{i=1,2,...,n}\left(t_i^r\right) \le t_j^d \le k_u \max_{i=1,2,...,n}\left(t_i^r\right)\right)$, чтобы период анализируемой части считался корректным, где k_l и k_u - определяемые пользователем внешние параметры, представленные положительными действительными значениями.

IV. Архитектура сетевого фильтра

Сетевой фильтр предназначен для перехвата DICOM-трафика и автономного анализа поступающих серий данных наблюдаемых пациентов. Программные компоненты низкого уровня полагаются на функциональность libpcap для захвата трафика с сетевых интерфейсов на машинах Linux. Все функции Linux и возможности libpcap для простоты обозначены как компонент LinuxKernel на представленных UML-диаграммах.

Диаграмма классов на рис. 1 иллюстрирует анатомию компонентов фильтрации и анализа.

Рис. 1 - Диаграмма классов анализатора медицинских данных

1. Класс DICOMInterceptor предназначен для анализа пакетов DICOM и запуска проверки содержимого. Это основной класс, взаимодействующий с LinuxKernel.

inspectPacket	1. Извлекает метку времени из пакета данных. 2. Если данные являются действительным файлом DICOM с меткой времени, запускает задачу анализа асинхронно.
passThroughAndLog	Пропускает пакет дальше и регистрирует это событие.
extractTimestamp	Извлекает метку времени из файла DICOM.
postAlert	Публикует оповещение для указанного файла DICOM.
analysisTaskProceduce	1. Ожидает появления файла с указанной меткой времени в архиве DICOM. 2. Запускает анализ этого файла. 3. Если есть нарушения, публикует предупреждения.

2. Класс LegalityCheckingProvider предназначен для управления конфигурацией и обеспечения функции проверки для компонента фильтрации.

retrainPatientSeriesModel	Переобучить модель машинного обучения для указанной серии с указанным интервалом данных для обучения.
testPatientSeries	Возвращает ассоциативный массив с идентификаторами отсчетов в качестве ключей и их корректностью в виде значений <boolean, boolean>. Первое логическое значение в кортеже - это корректность периодичности, второе логическое значение в кортеже - корректность содержимого.
testPatientFile	Возвращает корректность периодичности и содержимого для отдельного указанного файла пациента.
getSeriesOfPatient	Возвращает идентификаторы наблюдаемых серий для пациента.
releasePatientSeries	Удалить серию пациента из списка наблюдений.
lookOnPatientSeries	Добавить серию пациента в список наблюдений.
observations	Ассоциативный массив с идентификаторами пациентов в качестве ключей и списками идентификаторов серий в качестве значений.

3. Класс SeriesTester загружает и использует различные плагины для проверки корректности данных для различных модальностей (типов серий данных). Он также использует экземпляр PeriodAnalyzer для проверки периодичности входящих данных.

testSeries	Возвращает ассоциативный массив с идентификаторами отсчетов в качестве ключей и их корректностью в виде значений <boolean, boolean>. Первое логическое значение в кортеже - это корректность периодичности, второе логическое значение в кортеже - корректность содержимого.
getCheckerForSeries	Получить точку входа в соответствующем плагине для указанной серии данных.
periodAnalyzer	Экземпляр PeriodAnalyzer.

4. Класс PeriodAnalyzer предназначен для анализа периодичности поступления данных.

setEthalonPeriod	Устанавливает значение периода, с которым проверяется соответствие для временного ряда.
getSuspiciousFiles	Возвращает ассоциативный массив с идентификаторами отсчетов в качестве ключей, и корректностью отсчетов в виде логических значений.

5. ILegalityChecker - это общий интерфейс, который должен реализовывать каждый плагин проверки для любой модальности.

trainModel	Обучить модель машинного обучения для указанной серии с указанным интервалом данных для обучения.

testSeries	Возвращает ассоциативный массив с идентификаторами отсчетов в качестве ключей, и корректностью отсчетов в виде логических значений.

6. MRChecker - это плагин проверки модальности MR, реализующий интерфейс ILegalityChecker.
7. RGChecker - это плагин проверки модальности RG, реализующий интерфейс ILegalityChecker.
8. ECGChecker - это плагин для проверки модальности ЭКГ, реализующий интерфейс ILegalityChecker.

Все классы, кроме классов проверки корректности для разных модальностей (MRChecker, RGChecker, ECGChecker), реализованы в компоненте DICOMAnalyzer. Классы проверки корректности для разных модальностей реализованы в отдельных компонентах. Диаграмма компонентов на рис. 2 иллюстрирует взаимосвязь компонентов в системе и другого программного обеспечения, с которым она работает, например dcm4che, предоставляющий функциональность архива DICOM.

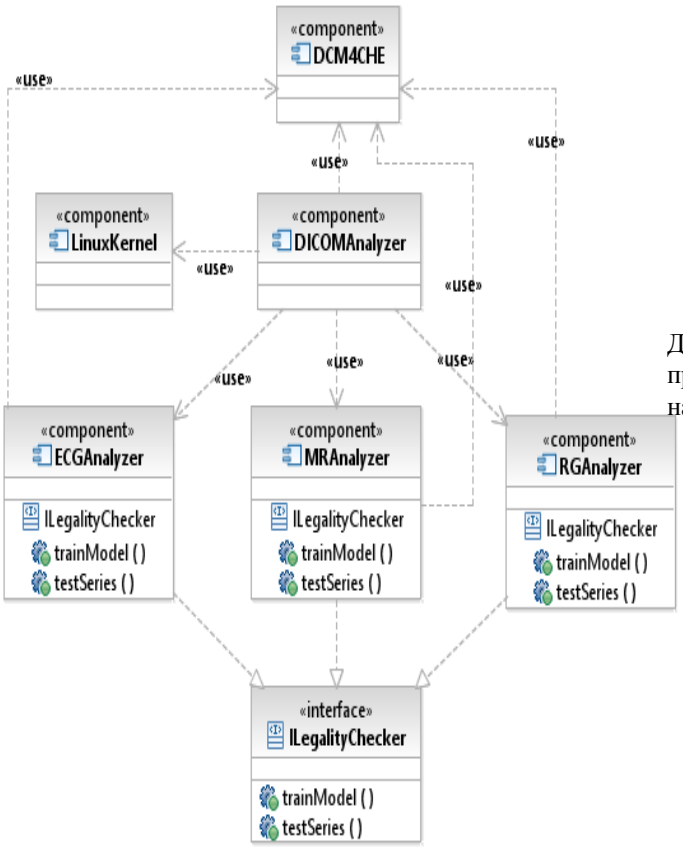

Рис. 2 - Диаграмма компонентов анализатора медицинских данных

DICOMAnalyzer, MRAnalyzer, ECGAnalyzer, RGAnalyzer (и другие компоненты проверки корректности для разных модальностей) должны быть упакованы как отдельные артефакты (с такими же именами, рис. 3). Все они должны быть развернуты на узле FilteringFrontend. Эти узлы должны быть кластеризованы для обеспечения высокой доступности. Сервер Nginx - это надежное решение с открытым исходным кодом для балансировки нагрузки HTTP, которое будет развернуто на узле HTTPLoadBalancer. dcm4che, предоставляющий функциональность архива DICOM, и PostgreSQL, предоставляющий для него базу данных, также должны быть развернуты на отдельных узлах.

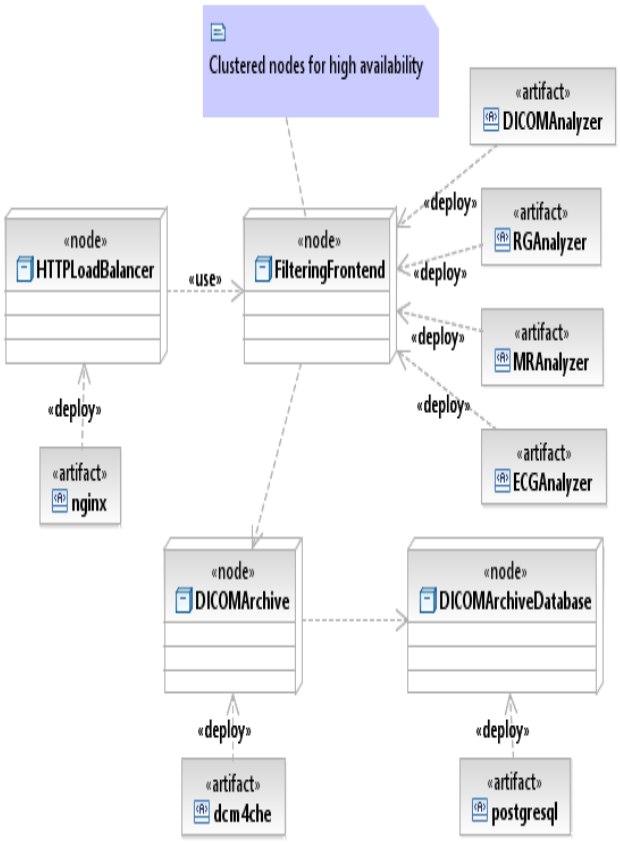

Рис. 3 - Диаграмма развертывания анализатора медицинских данных

Диаграмма последовательности на рис. 4 иллюстрирует процесс проверки и анализа пакетов DICOM в системе на фильтрующем внешнем узле.

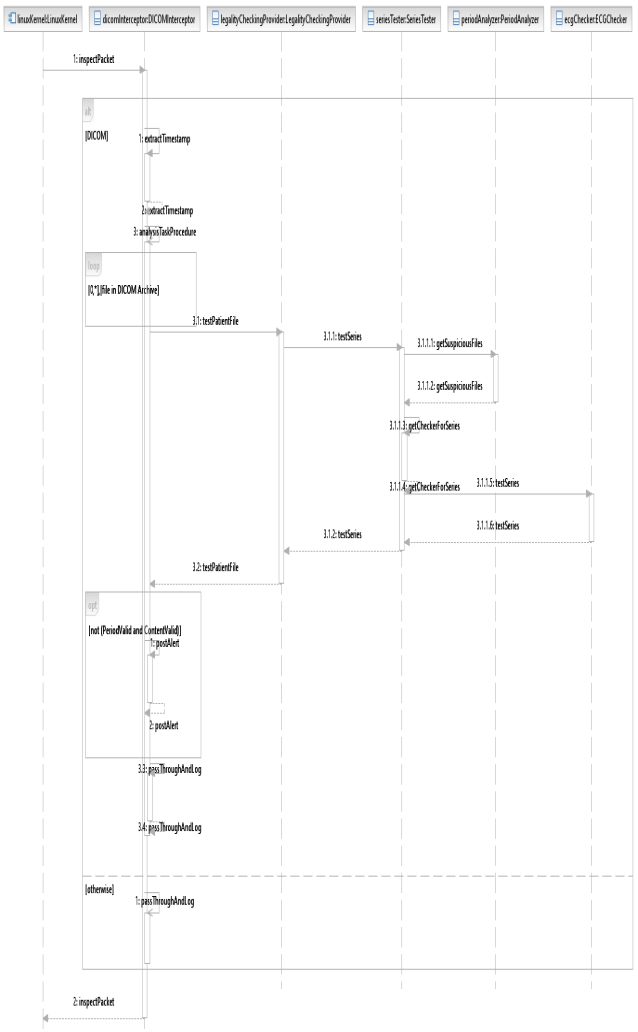

Рис. 4 - Анализ входящих файлов DICOM

Процесс, показанный на диаграмме, предполагает, что модель для наблюдения за рядами уже обучена (обучение было инициировано вручную или автоматически с помощью средств планирования). Могут быть проанализированы только действительные пакеты DICOM. Когда пакет прибывает, он проверяется и извлекается значение временной метки. Если проверяемый пакет является допустимым пакетом DICOM, процедура анализа выполняется асинхронно. Происходит ожидание, пока файл DICOM не появится в архиве DICOM, и тогда начинается анализ на предмет корректности периодичности и содержимого с помощью экземпляров PeriodAnalyzer и [Modality]Checker. Какую программу проверки выбрать для анализа содержимого, решает getCheckerForSeries класса SeriesTester. Во всех случаях сетевая активность журналируется.

V. Заключение

Предложена архитектура системы, готовой к обработке больших данных при автономной проверке корректности медицинских записей в архивах DICOM. Условие отказоустойчивости - это наличие кластеризованных экземпляров компонентов DICOMAnalyzer, развернутых на разных физических машинах с наличием или отсутствием уровня виртуализации или контейнеризации. Анализаторы для различных модальностей должны развертываться с такой же избыточностью. Их реализация относится к масштабным исследованиям, которые являются предметом будущей работы..

Библиография

[1] Магомедов Ш.Г. Анализ защиты компьютерных сетей и приложений информационных процессов учреждений здравоохранения. Cloud of Science. 2020. Т. 7. № 3. С. 685-704.

[2] Hewett, Andrew J., et al. "Conformance testing of DICOM image objects." Medical Imaging 1997: PACS Design and Evaluation: Engineering and Clinical Issues. Vol. 3035. International Society for Optics and Photonics, 1997.

[3] https://www.dcm4che.org/ (Дата обращения 20.08.2020)

[4] Coatrieux, Gouenou, et al. "Mixed reversible and RONI watermarking for medical image reliability protection." 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2007.

[5] Abd-Eldayem, Mohamed M. "A proposed security technique based on watermarking and encryption for digital imaging and communications in medicine." Egyptian Informatics Journal 14.1 (2013): 1-13.

[6] Al-Haj, Ali. "Providing integrity, authenticity, and confidentiality for header and pixel data of DICOM images." Journal of digital imaging 28.2 (2015): 179-187.

[7] Singla, Sanjoli, and Jasmeet Singh. "Cloud data security using authentication and encryption technique." Global Journal of Computer Science and Technology (2013).

[8] Dorgham, Osama, et al. "Enhancing the security of exchanging and storing DICOM medical images on the cloud." International Journal of Cloud Applications and Computing (IJCAC) 8.1 (2018): 154-172.

[9] Kassinen O. et al. Guidelines for the implementation of cross-platform mobile middleware. International Journal of Software Engineering and Its Applications. – 2010. – Т. 4. – №. 3. – С. 43-58.

[10] Петров, Александр Викторович, et al. "Метод шаблонов приложений для повышения мобильности распределенных систем сбора и ретрансляции информации с биомедицинских датчиков." Журнал радиоэлектроники 5 (2013): 7-7.

[11] Лебедев, А. С., О. С. Большаков, and А. В. Петров. "Проектирование распределенной системы ретрансляции данных с мобильными клиентами на основе кроссплатформенных методов разработки программного обеспечения." Современные проблемы науки и образования 1 (2013): 133-133.

[12] Карпов О. Е. и др. Цифровое здравоохранение в цифровом обществе. Экосистема и кластер. – 2017.

[13] Комисарук О. В., Никульчев Е. В., Малых С. Б. Разработка нейросетевой модели выявления артефактов в электроэнцефалограмме мозга. Cloud of Science. 2020. Т. 7. № 3. С. 631-654.

[14] Benssalah M., Rhaskali Y. A Secure DICOM Image Encryption Scheme Based on ECC, Linear Cryptography and Chaos //020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP). – IEEE, 2020. – С. 131-136.

[15] Mortajez S. et al. A novel chaotic encryption scheme based on efficient secret keys and confusion technique for confidential of DICOM images //Informatics in Medicine Unlocked. – 2020. – С. 100396.

[16] Shini S. G., Thomas T., Chithraranjan K. Cloud based medical image exchange-security challenges //Procedia Engineering. – 2012. – Т. 38. – С. 3454-3461.

[17] ISO 12052:2017. Health informatics — Digital imaging and communication in medicine (DICOM) including workflow and data management. [Электронный ресурс] URL: https://www.iso.org/standard/72941.html

Шамиль Гасангусейнович МАГОМЕДОВ,
доцент кафедры КБ-4 «ИСИБ» ФГБОУ ВО «МИРЭА – Российский технологический университет» (https://www.mirea.ru/),
email: msgg@list.ru
ORCID: 0000-0001-8560-1937

Architecture of information system for authenticating medical data in DICOM archive

S.G. Magomedov

Abstract - Nowadays personal medical devices play an increasingly important role in health care ecosystems as equipment for life support of patients. And the malicious software and protocols interacting with this equipment cause more and more interest. Any data transmitted via DICOM (the most common medical protocol) will be considered in the context of medical correctness to minimize the possible harm from fake DICOM files. To achieve acceptable accuracy in practice, two aspects are taken into account: periodicity correctness and image data (time series) correctness for the modality under consideration. This paper proposes the architecture of the information system and an integrated network filter for data validation in the DICOM archive that provides opportunities for analysis and warning management. The architecture of the proposed system is designed to work with archive data sets as well as to perform analysis of incoming big data streams. The software components operate in a cluster, which provides horizontal scalability and fault tolerance.

Keywords - big data, DICOM archive, data verification, distributed systems, network traffic filtering.

REFERENCES

[1] Magomedov S.G. Security analysis of computer networks and applications of the healthcare organizations information processes. Cloud of Science. 2020. T. 7. № 3. C. 685-704.

[2] Hewett, Andrew J., et al. "Hewett, Andrew J., et al. "Conformance testing of DICOM image objects. Medical Imaging 1997: PACS Design and Evaluation: Engineering and Clinical Issues. Vol. 3035. International Society for Optics and Photonics, 1997.

[3] https://www.dcm4che.org/ (Date of address 20.08.2020).

[4] Coatrieux, Gouenou, et al. "Mixed reversible and RONI watermarking for medical image reliability protection". 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2007.

[5] Abd-Eldayem, Mohamed M. "A proposed security technique based on watermarking and encryption for digital imaging and communications in medicine. Egyptian Informatics Journal 14.1 (2013): 1-13.

[6] Al-Haj, Ali. "Providing integrity, authenticity, and confidentiality for header and pixel data of DICOM images. Journal of digital imaging 28.2 (2015): 179-187.

[7] Singla, Sanjoli, and Jasmeet Singh. "Cloud data security using authentication and encryption technique. Global Journal of Computer Science and Technology (2013).

[8] Dorgham, Osama, et al. "Enhancing the security of exchanging and storing DICOM medical images on the cloud. International Journal of Cloud Applications and Computing (IJCAC) 8.1 (2018): 154-172.

[9] Kassinen O. et al. Guidelines for the implementation of cross-platform mobile middleware. International Journal of Software Engineering and Its Applications. – 2010. – T. 4. – №. 3. – C. 43-58.

[10] Petrov, Alexander Viktorovich, et al. "Method of application templates for increasing mobility of distributed systems for collection and retransmission of information from biomedical sensors". Journal of Radioelectronics 5 (2013): 7-7.

[11] Lebedev, A. S., O. S. Bolshakov, and A. V. Petrov. "Design of a distributed system of data retransmission with mobile clients based on cross-platform methods of software development". Modern Problems of Science and Education 1 (2013): 133-133.

[12] Karpov O. E. et al. Digital Healthcare in the Digital Society. Ecosystem and Cluster. – 2017.

[13] Komisaruk O. V., Nikulchev E. V., Malykh S. B. Development of a neural network model for detecting artefacts in the electroencephalogram of the brain. Cloud of Science. 2020. T. 7. № 3. C. 631-654.

[14] Benssalah M., Rhaskali Y. A Secure DICOM Image Encryption Scheme Based on ECC, Linear Cryptography and Chaos //020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP). - INTERNATIONAL CONFERENCE ON COMMUNICATIONS, CONTROL SYSTEMS AND SIGNAL PROCESSING (CCSSP), IEEE, 2020. - – C. 131-136.

[15] Mortajez S. et al. A novel chaotic encryption scheme based on efficient secret keys and confusion technique for confidential of DICOM images //Informatics in Medicine Unlocked. – 2020. – C. 100396.

[16] Shini S. G., Thomas T., Chithraranjan K. Cloud based medical image exchange-security challenges //Procedia Engineering. - – 2012. - – T. 38. - – C. 3454-3461.

[17] ISO 12052:2017. Health informatics - Digital imaging and communication in medicine (DICOM) including workflow and data management. [Electronic resource] URL: https://www.iso.org/standard/72941.html.

Shamil Gasanguseynovich MAGOMEDOV,
Associate Professor of the Department of CS-4 "ISIB" "MIREA - Russian Technological University". (https://www.mirea.ru/),
email: msgg@list.ru
ORCID: 0000-0001-8560-1937