ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ
СПбГУТ))) Декабрь 2019 Том 7 № 2 ISSN 2307-1303 http://www.itt.sut.ru
TELECOM IT
December 2019 Vol. 7 Iss. 2 ISSN 2307-1303 http://www.itt.sut.ru

ИДЕНТИФИКАЦИЯ МУЛЬТИМЕДИЙНОГО КОНТЕНТА НА ОСНОВЕ АРХИТЕКТУРЫ ЦИФРОВЫХ ОБЪЕКТОВ

В. В. Зеленов[1], Н. И. Шустов[1], Р. В. Киричек*[1], А. С. Бородин[2]

[1]Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М. А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация
[2]ПАО «Ростелеком», Москва, 115172, Российская Федерация

*Адрес для переписки: kirichek.sut@mail.ru

Аннотация
Предмет исследования. Статья посвящена разработанному методу идентификации мультимедийного контента на основе архитектуры цифровых объектов и стенду, который демонстрирует работу разработанного метода. **Метод.** Рассмотрен принцип разделения идентификационных данных, которые необходимо скрыть, на составляющие части и внедрение этих частей в случайном порядке в разные области файла. Проведен анализ существующих решений. **Основные результаты.** Разработан стенд по идентификации мультимедийного контента на основе архитектуры цифровых объектов, в котором используется реализованный алгоритм записи идентификатора в файл, а также его последующего извлечения из файла. **Практическая значимость.** Разработанный стенд по идентификации мультимедийного контента на основе архитектуры цифровых объектов может быть использован для защиты от пиратского контента.

Ключевые слова
Стеганография, стеганографическая инъекция, DOA, REST API.

Информация о статье
УДК 004.7
Язык статьи – русский.
Поступила в редакцию 15.11.19, принята к печати 30.12.19.

Ссылка для цитирования: Зеленов В. В., Шустов Н. И., Киричек Р. В., Бородин А. С. Идентификация мультимедийного контента на основе архитектуры цифровых объектов // Информационные технологии и телекоммуникации. 2019. Том 7. № 2. С. 85–95. DOI 10.31854/2307-1303-2019-7-2-85-95.

IDENTIFICATION OF MULTIMEDIA CONTENT BASED ON DIGITAL OBJECT ARCHITECTURE

V. Zelenov[1], I. Shustov[1], R. Kirichek[*1], A. Borodin[2]

The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation
[2]PJSC "Rostelecom", Moscow, 115172, Russian Federation

*Corresponding author: kirichek.sut@mail.ru

Abstract—**Research subject.** The article is devoted to the method of identification of multimedia content based on digital object architecture and a stand which demonstrates the work of the method. **Method.** The principle of identification data division and injecting data parts in random order to a different area of the file is considered. The analysis of existing solutions is performed. **Core results.** Stand for identification of multimedia content based on digital object architecture have been developed. This stand can write identifier to the multimedia file and read from it. **Practical relevance.** The developed stand for identification of multimedia content based on digital object architecture can be used to protect digital content from piracy.

Keywords—Steganography, steganography injection, DOA, REST API.

Article info
Article in Russian.
Received 15.11.19, accepted 30.12.19.

For citation: Zelenov V., Shustov I., Kirichek R., Borodin A.: Identification of Multimedia Content Based on Digital Object Architecture // Telecom IT. 2019. Vol. 7. Iss. 2. pp. 85-95 (in Russian). DOI 10.31854/2307-1303-2019-7-2-85-95.

Введение

С каждым днем в Сети Связи Общего Пользования количество цифрового контента: видео, аудио, изображения, и т.д. непрерывно растет. В связи с этим увеличивается количество случаев мошенничества, связанных с незаконным копированием цифрового контента и размещением его в публичный доступ от своего лица. В качестве одного из решений применяются технологии идентификации цифрового контента.

Методы идентификации позволяют однозначно идентифицировать цифровой контент, что защищает от незаконного копирования [1, 2].

По статистике экспертов в 2017 году рост пиратства составил 21 % по сравнению с 2016 годом и достиг 85 млн долларов[1]. Даже учитывая доступность стриминговых сервисов по подписке, таких как Netflix и Spotify, пользователи предпочитают пиратский контент.

[1] Аналитика экспертов Group-IB [Электронный ресурс] // Group-IB, URL: https://www.group-ib.ru/media/antipiracy-statistics-record-2018/

В связи с этой проблемой возникает актуальная задача однозначно идентифицировать мультимедийный контент.

Данная статья посвящена решению этой проблемы, для чего необходимо реализовать процедуру идентификации цифрового контента.

Решения по защите от нелегального копирования цифрового контента

Далее рассмотрим решения возможных аналогов разрабатываемой системы.

Одним из таких решений является система компании Group-IB. Group-IB – одна из значимых компаний на международном рынке, которая занимается предотвращением и расследованием киберпреступлений и мошенничества, используя собственные технологии.

Anti-Piracy Group-IB (рис. 1)[2] – это система защиты цифрового контента и противодействия пиратству. В данной системе используется передовая технология «цифрового отпечатка», при помощи которой возможно сравнение цифровых копий по различным критериям и выявление по ним пиратского контента, даже если этот контент был значительно изменен.

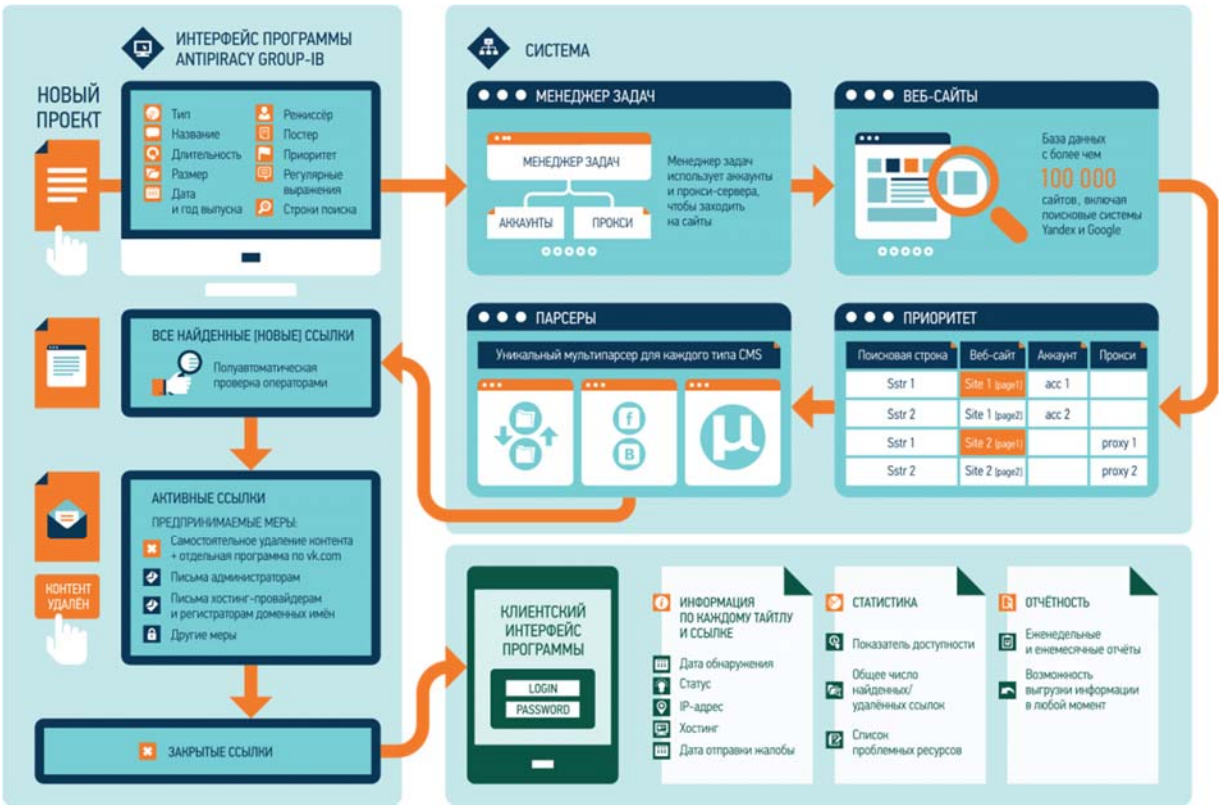

Рис. 1. Система Anti-Piracy Group-IB

[2] Anti-Piracy Group-IB // Group-IB, URL: https://www.group-ib.ru/antipiracy.html

Данная система проверяет около 110 тысяч сайтов на русском и английском языках, в том числе торрент-трекеры, стриминговые сервисы, группы социальных сетей и пиратские площадки на просторах даркнета.

Также данная система определяет владельца ресурса и обращается к нему напрямую, благодаря чему уведомления о нарушении авторского права приходят модераторам гарантированно. Модераторский аккаунт в каждом из крупных сервисов позволяет этой системе блокировать неправомерный мультимедиа-контент оперативно.

Недостатком данной системы является то, что она допускает существование пиратского контента, который в дальнейшем она пытается обнаружить.

Другой системой защиты от пиратства является система StarForce (рис. 2)[3]. Их система защиты видео и аудио от копирования позволяет предотвратить неправомерное распространение аудио и видео при помощи привязки к ПК, оптическим дискам или USB накопителям. Преимуществами этой системы являются: защита от несанкционированного использования, распространения, захвата экрана и его трансляции, утечки конфиденциальной информации.

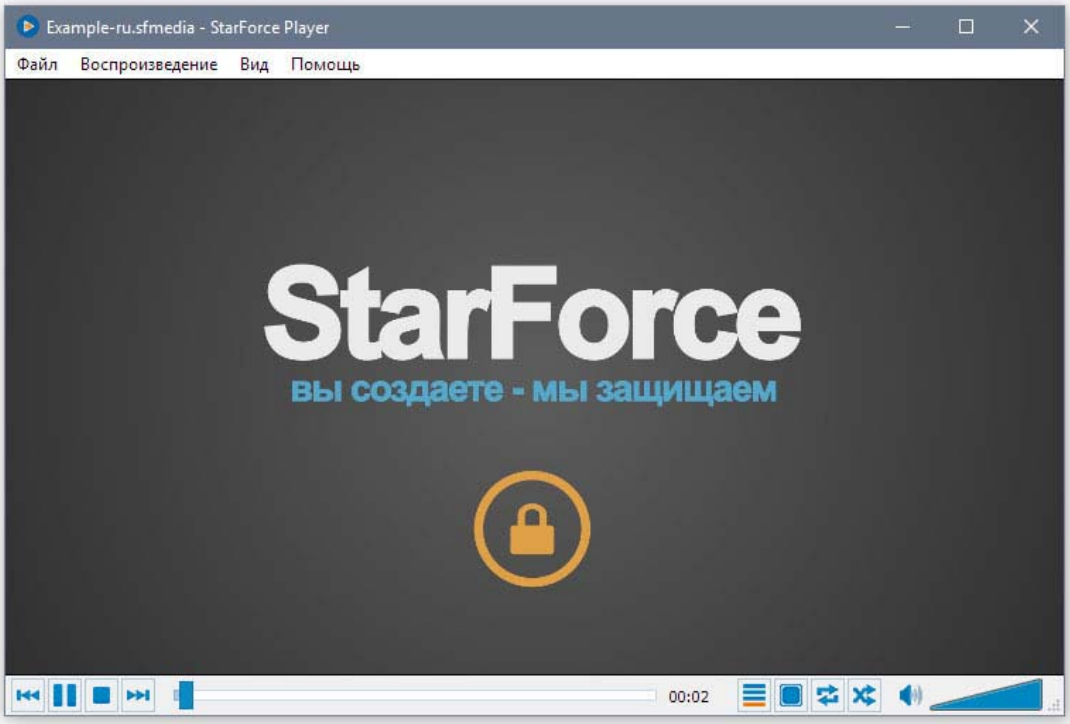

Рис. 2. Интерфейс ПО StarForce Player

В то же время недостатком этой системы является тот факт, что аудио и видеофайлы, которые были защищены системой StarForce, возможно воспроизводить только при помощи специального программного обеспечения StarForce Player.

[3] StarForce Player. URL: http://www.star-force.ru/

Еще одним решением является система фирмы BrandSecurity[4], которая также позволяет защитить интеллектуальную собственность от неправомерного использования. Система BrandSecurity анализирует все возможные источники нелицензионного контента. Далее, каждый обнаруженный случай анализируется сотрудниками фирмы BrandSecurity. После чего происходит рассылка уведомлений о блокировке контента напрямую контент-провайдерам. По итогам работы ведется отчетность, а клиентам предоставляется статистика.

Минус данной системы в том, что приходится вручную искать источники распространения нелицензионного контента, чтобы впоследствии разработать алгоритм борьбы с данным источником.

Исходя из всех рассмотренных на рынке систем, было принято решение разработать систему идентификации контента, которая должна идентифицировать подлинный мультимедийный контент однозначно.

Архитектура цифровых объектов

Для хранения информации, связанной с идентификатором контента, была выбрана архитектура цифровых объектов.

Архитектура DOA [3, 4, 5] (англ. Digital Object Architecture – архитектура цифровых объектов) – это система, используемая для идентификации, аутентификации и авторизации цифровых объектов в Сети Связи Общего Пользования. При помощи данной системы можно реализовать процедуры AAA (от англ. *Authentication* – аутентификация, *Authorization* – авторизация, *Accounting* – хранение данных о пользователе) для физических или виртуальных устройств Интернета Вещей посредством получения информации по уникальному ключу объекта – идентификатору.

LHR [6, 7, 8, 9, 10, 11] – локальный реестр. LHR содержит в себе хранилище данных о цифровых объектах. Цифровой объект – это набор проиндексированных структур, каждая из которых может указывать на другой цифровой объект, например, URL, адрес электронной почты и т. д.

GHR [6, 7, 8, 9, 10,11] – глобальный реестр. GHR содержит в себе данные публичных ключей сервисов. Эти ключи могут быть получены с правами администратора.

Сервера DOA — GHR. Данный сервер представляет собой базу данных, содержащих записи о местоположении серверов LHR по идентификаторам DOI.

Сервера DOA — LHR. Данный сервер представляет собой базу данных, содержащую подробную информацию о идентифицируемом с помощью DOI объекте.

Далее для аутентификации и получения дополнительной информации о мультимедийном контенте REST API [13] сервер отправляет запрос на удаленный сервер DOA GHR, содержащий идентификатор DOI, полученный из видеоконтента при помощи алгоритма изъятия. В свою очередь GHR перенаправляет запрос на удаленный сервер DOA LHR, который и возвращает REST API серверу информацию видеофайле.

Далее рассмотрим метод идентификации контента.

[4] BrandSecurity. URL: https://brandsecurity.ru/

Метод идентификации мультимедийного контента на основе архитектуры цифровых объектов

Для того, чтобы записать идентификатор в файл видеоконтента, был разработан алгоритм стеганографической инъекции [12].

Суть метода заключается в разделении идентификационных данных (стеганограммы), которые необходимо скрыть, на составляющие части и внедрение этих частей в случайном порядке в разные области файла.

Стеганографическая инъекция – это способ сокрытия данных в файле при помощи разделения их на составные части и запись этих частей по определенному алгоритму в разные части файла. Таким образом, информация, записанная в файл, не будет доступна без знания алгоритма извлечения идентификатора.

Алгоритм стеганографической инъекции:

1. Извлечение WAV-аудиодорожки из оригинального видео;
2. Представление аудиофайла в последовательность байт;
3. Запись идентификатора в последовательность байт;
4. Создание нового WAV-файла с записанным идентификатором;
5. Добавление модифицированной аудиодорожки к видео.

Метод, используемый в этой статье, предполагает извлечение аудиодорожки из видеофайла, что позволяет нам произвести инъекцию не в сам видеофайл, а в аудиодорожку формата WAV. Аудиодорожка раскладывается на последовательность байт, после чего при помощи алгоритма части идентификатора записываются в определенной последовательности в разные части видеофайла, после чего аудиодорожка с идентификатором добавляется в видеофайл.

Алгоритм изъятия идентификатора:

1. Извлечение WAV-аудиодорожки из оригинального видео;
2. Представление аудиофайла в последовательность байт;
3. Изъятие идентификатора из массива байт при помощи оригинального видео.

Для извлечения идентификатора необходимо иметь оригинальный видеофайл без идентификатора и сам алгоритм извлечения, который связан с алгоритмом записи, так как только при наличии оригинала можно вычислить, в каких местах находятся части идентификатора. Преимущества этого метода в том, что для извлечения идентификатора необходим оригинальный файл, который хранится только на недоступном извне сервере и нигде не публикуется и алгоритм поиска идентификатора, который без оригинального файла не сможет найти идентификатор, что обеспечивает высокий уровень безопасности, что может являться недостатком, так как требуется хранить оригинальный файл.

Стенд по идентификации мультимедийного контента на основе архитектуры цифровых объектов

Далее, рассмотрим работу собранного стенда по идентификации мультимедийного контента на основе архитектуры цифровых объектов.

Стенд (рис. 5) состоит из компьютера с операционной системой Windows 10, на который предварительно загружено программное обеспечение Web API [13] сервера, на котором реализован алгоритм по записи и извлечению идентификатора, веб-сайта, где представлен видеоконтент, а также установлено разработанное расширение для браузера Google Chrome. Цифровой контент на данном стенде представлен видеофайлом, который маркируется и идентифицируется с помощью разработанного алгоритма записи идентификатора.

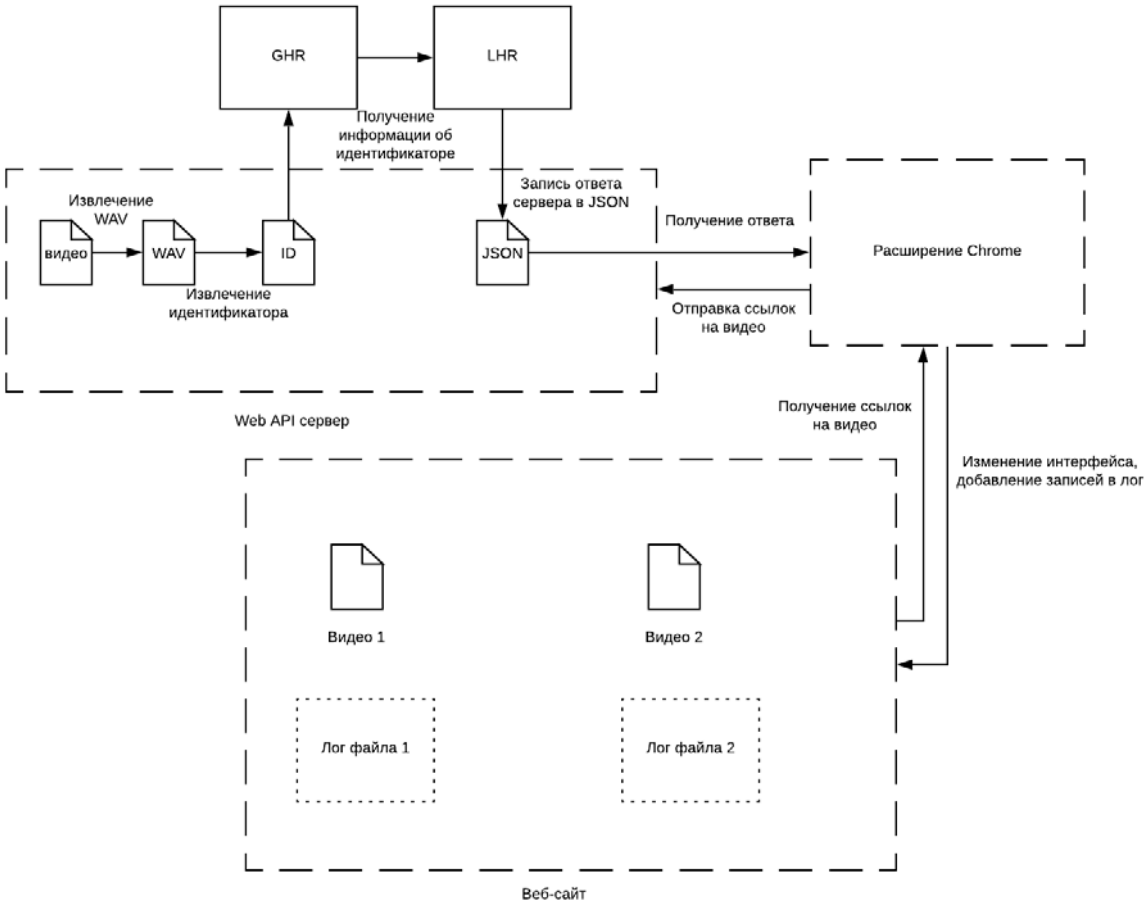

Рис. 5. Схема работы стенда по идентификации мультимедийного контента

Веб-сайт (рис. 6) отображает цифровой видеоконтент, который впоследствии пройдет проверку идентификации, а также отображает информацию о запросах на Web API сервер в виде двух лог-списков, которые отображают последовательность запросов и ответов REST API [13] сервера, в наглядной для восприятия форме.

Расширение для браузера Chrome выполняет поиск видеоконтента на веб-странице. Если видеоконтент был найден, то расширение отправляет в Web API сервер ссылку для загрузки видео. Далее, расширение получает ответ от сервера в виде JSON-данных, в которых содержится информация о подлинности видеоконтента. Если видеоконтент прошел идентификацию, то он помечается галочкой, что информирует пользователя о подлинности данного видеоконтента. В противном случае, если видеоконтент не прошел идентификацию, то оно блокируется путем вывода поверх него надписи о нарушении авторского права.

Web API сервер реализован на основе архитектуры REST API, что позволяет обращаться к его методам через URL-запросы. Сервер выполняет две функции: запись идентификатора в исходный файл при помощи вышеописанного алгоритма и извлечение идентификатора из файла при его наличии.

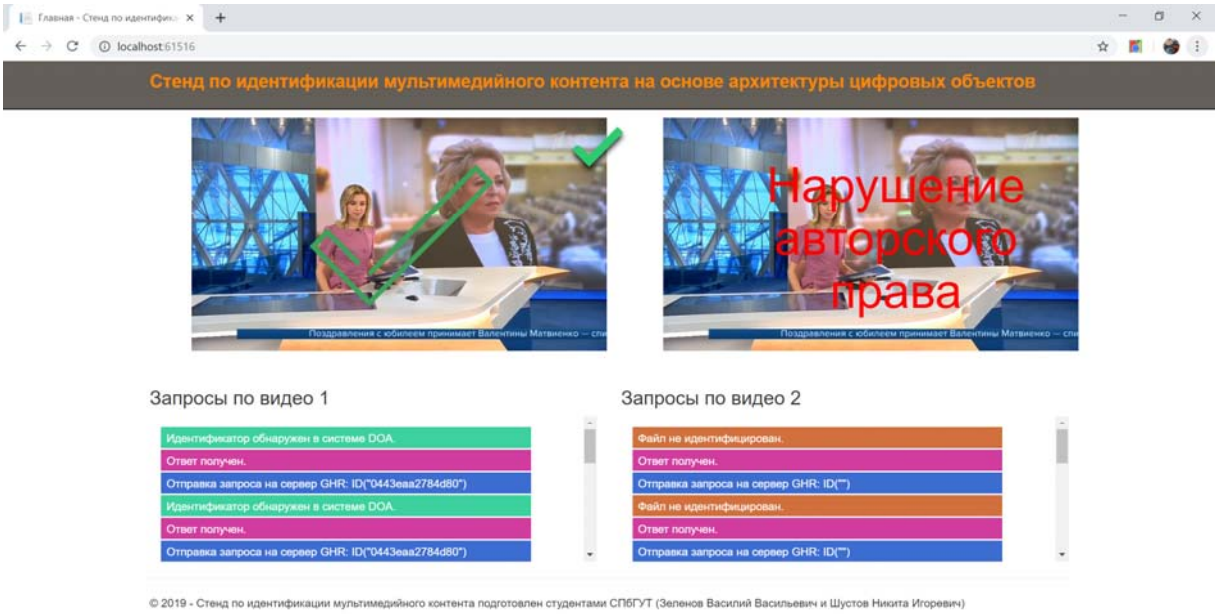

Рис. 6. Стенд по идентификации мультимедийного контента на основе архитектуры цифровых объектов

Процесс маркировки происходит следующим образом:

• Web API сервер получает оригинальный файл;

• происходит вызов функции отправки запроса на сервер GHR (*Global Handle Register*), после чего GHR запрашивает у LHR (*Local Handle Register*) локальный идентификатор DOI (*Digital Object Identifier*); функция возвращает полученный идентификатор DOI;

• открывается новый маркированный файл в режиме записи, производится операция записи маркера (DOI) в этот файл по разработанному алгоритму, который описан выше.

Процесс идентификации файла происходит следующим образом:

• открывается веб-сайт, на котором воспроизводится видеоконтент;

• далее, установленное расширение выполняет поиск видео и по результату поиска отправляет ссылки на найденные видео на REST API сервер;

• REST API сервер получает n-количество ссылок на видео, скачивает их и выполняет процедуру извлечения идентификатора из видео, если он там имеется;

• если идентификатор найден, то выполняется запрос на GHR, который делает запрос на LHR, на котором хранятся данные, связанные с идентификатором. Если найденный идентификатор существует, то происходит возврат JSON с информацией о видеоконтенте;

• далее, расширение получает ответ от сервера, и если видео идентифицировано, то видео маркируется галочкой как подлинное, иначе, пользователю сообщается, что видео не является подлинным.

Заключение

В данной статье были рассмотрены различные решения, у каждого из которых есть свои преимущества, но ни у одного из них нет возможности заранее предотвратить незаконное распространение мультимедиа контента.

Исходя из этого, был разработан метод идентификации мультимедийного контента, на основе которого был собран стенд, демонстрирующий работу метода. Разработанный метод позволяет незаметно для пользователя хранить информацию об идентификаторе в видео, при этом обеспечивает высокий уровень безопасности, так как для извлечения идентификатора из видеофайла необходим файл оригинала видео без идентификатора, который недоступен извне.

В дальнейшем, предполагается совершенствование системы, а именно:

1) привязка видео к аккаунту видеохостинга;

2) улучшение алгоритма инъекции стеганограммы, а именно – поиск метода, обеспечивающего высокий уровень безопасности, не требующего наличия оригинального файла.

Литература

1. Коржик В. И., Небаева К. А., Герлинг Е. Ю., Догиль П. С., Федянин И. А. Цифровая стеганография и цифровые водяные знаки : монография в 2 ч. / под общ. ред. проф. В. И. Коржика; СПбГУТ. – СПб., 2016. – С. 226.

2. Бескид П. П., Татарникова Т. М. О некоторых подходах к решению проблемы авторского права в сети интернет // Ученые записки Российского государственного гидрометеорологического университета. 2010. № 15. С. 199–210.

3. Karim Farhat. Digital Object Architecture and the Internet of Things: Getting a ‘handle’ on techno-political competition, 2017. 22 p.

4. S. Sun, L. Lannom, B. Boesch, RFC3650 "Handle System Overview", CNRI, 2003. 94 p.

5. S. Sun, L. Lannom, S. Reilly RFC3651 “Handle System Namespace and Service Definition”, CNRI, 2003. 3 p.

6. Аль-Бахри М. С., Киричек Р. В., Бородин А. С. Архитектура цифровых объектов как основа идентификации в эпоху цифровой экономики // Электросвязь. 2019. № 1. С. 12–22.

7. Аль-Бахри М. С. Метод идентификации устройств и приложений интернета вещей в гетерогенных сетях связи на базе архитектуры цифровых объектов // Электросвязь. 2019. № 4. С. 41–47.

8. Аль Бахри М. С., Киричек Р. В., Сазонов Д. Д. Моделирование системы идентификации устройств интернета вещей на базе архитектуры цифровых объектов // Труды учебных заведений связи. 2019. Т. 5, № 1. С. 42–47.

9. Al-Bahri M., Ateya A. A., Muthanna A., et al. Combating Counterfeit for IoT System based on DOA // Proceedings of the 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT) 2018, St. Petersburg, Russia, November 5-9, 2018. IEEE, 2018. pp. 338–342.

10. Al-Bahri M., Yankovsky A., Borodin A., Kirichek R. Smart System Based on DOA and IoT for Products Monitoring and Anti-counterfeiting // 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC). IEEE, 2019. pp. 25–31.

11. Al-Bahri M., Yankovsky A., Borodin A., Kirichek R. Testbed for Identify IoT Devices Based on Digital Object Architecture // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Proceedings of 18th International Conference, NEW2AN 2018, and 11th Conference, ruSMART 2018, St. Petersburg, Russia, August 27–29, 2018. – Cham: Springer, 2018. pp. 129–137.

12. Пономарев К. И. Некоторые математические модели стеганографии и их статистический анализ. Москва: Московский государственный институт электроники и математики (технический университет), 2010. 59 с.

13. Eve Andersson, Philip Greenspun, Andrew Grumet. Software Engineering for Internet Applications. MIT Press, 2006. 31 p. ISBN 0262511916.

References

1. Korzhik, V., Nebaeva, K., Gerling, E., Dogil', P., Fedyanin, I. Digital Steganography and Digital Watermarking. SPb.: SPbGUT, 2016. 226 p.

2. Beskid, P., Tatarnikova T. About Some Approaches to the Copyright Solution of a Problem in the Internet // RSHU Proceedings Journal. 2010. No. 15. Pp. 199–210.

3. Farhat, K. Digital Object Architecture and the Internet of Things: Getting a 'handle' on techno-political competition. Georgia Institute of Technology 2017. 22 p.

4. Sun, S., Lannom, L., Boesch, B. RFC3650 "Handle System Overview". CNRI, 2003. 94 p.

5. Sun, S., Lannom, L., Reilly, S. RFC3651 "Handle System Namespace and Service Definition". CNRI, 2003. 3 p.

6. Al-Bahri, M., Kirichek, R., Borodin, A. The Digital Object Architecture as a Basis for Identification in the Era of the Digital Economy // Elektrosvyaz'. 2019. No. 1. Pp. 12–22.

7. Al-Bahri, M. Method of Identification of Devices and Applications of the Internet of Things in Heterogeneous Communication Networks based on Digital Object Architecture // Elektrosvyaz'. 2019. No. 4. Pp. 41–47.

8. Al-Bahri, M., Kirichek, R., Sazonov, D. A Digital Object Architecture Based Internet of Things Devices Identification System Modeling // Proceedings of Telecommunication Universities. 2019. Vol. 5. Iss. 1. Pp. 42–47.

9. Al-Bahri, M., Ateya, A. A., Muthanna, A., et al. Combating Counterfeit for IoT System based on DOA // 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). 2018. Pp. 338–342.

10. Al-Bahri, M., Yankovsky, A., Borodin, A., Kirichek, R. Smart System Based on DOA and IoT for Products Monitoring and Anti-counterfeiting // 4th MEC International Conference on Big Data and Smart City (ICBDSC). 2019. Pp. 25–31.

11. Al-Bahri, M., Yankovsky, A., Borodin, A., Kirichek, R. Testbed for Identify IoT Devices Based on Digital Object Architecture // Lecture Notes in Computer Science. 2018. Vol. 11118. Pp. 129–137.

12. Ponomaryov, K. Some Mathematical Models of Steganography and Their Statistical Analysis. M.: MIEM, 2010. 59 P.

13. Andersson, E., Greenspun, P., Grumet, A. Software Engineering for Internet Applications. MIT Press, 2006. 31 p.

Зеленов Василий Васильевич – магистрант, СПбГУТ, Санкт-Петербург, 193232, Российская Федерация, ze_vs@outlook.com

Шустов Никита Игоревич – студент, СПбГУТ, Санкт-Петербург, 193232, Российская Федерация, niiiksh@outlook.com

Киричек Руслан Валентинович – доктор технических наук, профессор, СПбГУТ, Санкт-Петербург, 193232, Российская Федерация, kirichek.sut@mail.ru

Бородин Алексей Сергеевич – кандидат политических наук, представитель в Женеве, Ростелеком, Москва, Российская Федерация, borodin.msk@mail.ru

Zelenov Vasily – Undergraduate student, SUT, St. Petersburg, 193232, Russian Federation, ze_vs@outlook.com

Shustov Nikita – Student, SUT, St. Petersburg, 193232, Russian Federation, niiiksh@outlook.com

Kirichek Ruslan – Doctor of Engineering Sciences, Professor, SUT, St. Petersburg, 193232, Russian Federation, kirichek.sut@mail.ru

Borodin Alexey – Candidate of Political Sciences, Representative in Geneva, Rostelecom, Moscow, Russian Federation, borodin.msk@mail.ru