Раздел V. Информационные технологии и защита информации

УДК 004.056

Р.М. Алгулиев, Я.Н. Имамвердиев, Б.Р. Набиев

О МЕТОДЕ СОЗДАНИЯ ПРОФИЛЯ ДЛЯ ВЕБ-ПОЛЬЗОВАТЕЛЕЙ

Существует множество средств для обеспечения безопасности компьютерных сетей и оптимизации процессов. Известно, что одной из основных причин возникновения опасности в сетевом трафике является генерация аномального и непрофильного трафика. Все это, создаёт ненужную нагрузку на компьютерную сеть, что в свою очередь, снижает доступность полезной нагрузку на каналах связи. Это событие, является одним из тех событий, с которыми рано или поздно могут столкнуться корпоративные сети, неадаптированные к правилу поведения. Учитывая это, для определения профиля поведения трафика в сети, разработан специальный подход. Для определения профиля поведения применён метод кластеризации K-средних. Причиной выбора алгоритма K-средних является то, что для решения задачи кластеризации этот метод является очень быстрым и простым. Данные для анализа собраны в сетевой среде AzScienceNet состоящей из более чем 5000 IP адресов (персональных компьютеров), и эта сеть также разделяется на несколько маленьких подсетей. С целью обеспечения сохранности конфиденциальности пользователей, учтены политика AzScienceNet об использовании Интернета и дополнительные ограничения, конфиденциальности личных данных пользователей. В результате применения модели кластеризации были сформированы определённые кластеры. Кластеры, в основном, формируют социальные сети, видео-ресурсы и научно-практические ресурсы. Результат получен для 20 кластеров с помощью bigml.com ресурса. Наиболее часто обращаемый кластер состоит из научно-практических ресурсов. 2-ой по порядку обращаемый кластер-это социальные сети. Третий кластер состоит из обращений к видео-рурсам. Обращение к другим кластерам значительно меньше.

Сетевой трафик; кластеризация; профиль поведения; аномальный трафик.

R.M. Alguliev, Y.N. Imamverdiyev, B.R. Nabiyev

ABOUT THE METHOD OF CREATING A PROFILE FOR WEB USERS

There are some tools for securing computer networks and optimizing processes. It is known that one of the main causes of the danger in network traffic is the generation of anomalous and non-core traffic. All this, creates an unnecessary load on the computer network, which in turn, reduces the availability of payload on the communication channels. This event is one of those events, which sooner or later may face corporate networks that are not adapted to the rule of behavior. Considering this, to determine the behavior profile of traffic on the network, a special tool has been developed. To determine the behavior profile, the K-means clustering method was applied. The reason for choosing the K-means algorithm is that this method is very fast and simple for solving the clustering problem. Data for analysis is collected in AzScienceNet network environment consisting of more than 5000 IP addresses (individual computers) and this network is also divided into several small subnets. In order to ensure that users privacy is not violated, AzScienceNet is based on user policy and additionally limited data on the identity of users. As a result of the application of the clustering model, certain clusters were formed. Clusters, in the

main, form social networks, video resources and scientific and practical resources. The result is obtained for 20 clusters using the bigml.com resource. Most of all, the cluster under consideration consists of scientific and practical resources. The 2nd cluster in turn, these are social networks. The third cluster consists of calls to video resources. Appeal to other clusters is much less.

Network traffic; clustering; behavioral profile; anomalous traffic.

Введение. В стремительно глобализирующемся мире ускоренное получение любого ресурса или информации с помощью Интернета стало очень легко и доступно. Это очень позитивная и необходимая ситуация в условиях информационного общества. Но, как мы знаем, не вся генерируемая информация, является необходимой и полезной. Это, создаёт излишнюю нагрузку на компьютерную сеть, что в свою очередь, снижает доступность каналов связи. Это событие, является одним из тех событий, с которыми рано или поздно могут столкнуться корпоративные сети, неадаптированные к правилу поведения [1].

Согласно отчёту фирмы Symantec, представленному в 2014 году [2], число предотвратимых нападений на веб-ресурсы в течение одного дня составляет 586700. Принимая это во внимание, для того, чтобы пользователи сети могли избежать столкновений с угрозами, эффективно использоват корпоративные ресурсы, с ограниченними возможностями и для повыше-ния пропускной способности информационных каналов, предлагается формирование профиля поведения в трафике сети (в дальнейшем профиль поведения) на основе метода кластеризации сетевого трафика. Анализируя данные, полученные с помощью сетевого мониторинга трафика на основе оценки кластеризации, могут быть получены кластеры поведения определенного трафика, и реализация этого процесса осуществляется через алгоритм кластеризации. K-средних.

1. Анализ опубликованных работ. Одним из ключевых элементов управления сетью являются идентифи-кация сетевого трафика и категоризация [3]. В качестве примера можно привести приоритезацию потока формирования трафика, транспортной политики и диагностику мониторинга. Во всем мире с помощью IP сетей передается и принимается огромное количество информации [4]. Специалисты держат под контролем весь этот процесс и благодаря чему, выявляются и ликвидируются угрозы. Функции и параметры, включая заголовки пакета IP, позволяют получить большую информацию о сети и пользователях [5]. Кроме того, результаты анализа заголовков IP пакетов могут быть использованы для управления сетью и оптимизации, устранения угрозы и создания новых услуг. В [6], используя заголовки IP-пакетов, предлагается способ многоуровневой кластеризации в расширенной форме, объясняющий течение процесса в сети и профиль поведения пользователя. Кроме того, необходимо сказать, что проведенный процесс анализа используя заголовок IP-пакетов, обеспечивает неприкосновенность личной информации пользователей. Сетевой трафик или журналы файлов, собранные из трафика сети могут быть использованы для обнаружения аномалий и угроз. Для этого процесса используются различные методы и средства. Например, в [7], используя алгоритм кластеризации K-средних, предложен метод обнаружения аномалий в потоке трафика. Немаркированные данные сетевого трафика разделяются на два кластера, т.е. на нормальный и аномальный. В основе обнаружения аномалий в данных нового мониторинга лежит использование центра тяжести для выбора эффективного расстояния в определенных кластерах. Самоорганизующийся без центрального управления и без процесса контроля метод кластеризации является одним из самых новых подходов. Для этого, в [8] используется, основанный на взаимосвязи, метод поведе-ния муравьев. Преимущество данного метода заключается в том, что нет

необходимости в первичных данных и предварительного определения количества кластеров. Каждый из виртуальных муравьёв в отдельности и самостоятельно, исследуя сеть, выполняет процесс кластеризации. Но, поскольку этот метод является новым, коэффициент точности выполненного процесса вызывает сомнение.

В трафике сети подход “Machine learning” широко используется для определения аномальных потоков, основываясь на их уникальных статистических характеристиках. По сравнению с традиционной кластеризацией, нечёткая кластеризация является более гибкой, а для обнаружения вторжений и естественной обработки данных более целесообразной [9].

Многие методы кластеризации для обнаружения вторжений предусматривают разделение трафика на нормальный и аномальный. Методы кластеризации применяются для обнаружения разницы и схожих особенностей сессии трафика и для классификации каждого из них разделением на соответствующие группы [10]. Эти группы представляют присвоенные им знаки. В дальнейшем эти знаки используются для прогнозирования типов входящих сетевых трафиков.

Быстрая и точная идентификация сетевого трафика является одной из самых важных задач функции управления – QoS, мониторинга безопасности сети и т.д. Однако, в последнее время, количество узлов, использую-щих P2P увеличилось, и они, используя различные порты, скрываются под различными устройствами, необходимыми потоками сообщений или коди-рованными потоками сообщений, генерируя ненужные информационные потоки. В этом случае использование, считаемы классическими “port mapping” или “payload analysis” подходов, не эффективно. Альтернативным подходом является классификация сетевого TCP трафика исследованием поведения трафика внутри нескольких первичных пакетов. Это в будущем, кластеризируя всю информацию, позволяет облегчить процесс идентификации.

2. Лог-файлы обращений в интернет. Данные собраны в сетевой среде AzScienceNet состоящей из более чем 5000 адресов и эта сеть, также разделяется на несколько маленьких подсетей. С целью обеспечения ненарушения конфиденциальности пользователей, AzScienceNet основана на пользовательской политике и дополнительно огра-ничены данные о личности пользователей. Эти данные состоят из 10 переменных [11], приведённых в табл. 1.

Приведенные в табл. 1 семь переменных можно объяснить следующим образом:

1. Штамп времени. В целом, в области информационных технологий – символ или последовательность кодированной информации для регистрации даты появления, ликвидации, отправки или приема любого типа информации [12].

Таблица 1

Описание переменных кластеризации

Индекс	Объяснение переменных
1	Штамп времени
2	Время процесса
3	IP адрес
4	Результирующие коды
5	Объем контента
6	Метод запроса
7	URL
8	Код иерархии
9	IP отвечающего
10	Содержание

2. Время процесса. Регистрирует время процесса проведенное в кэше. То есть промежуток времени между началом и концом передачи пакетов HTTP [13].

3. IP адрес. Здесь регистрируются адреса обращений за информацией и к ресурсам.

4. Результирующие коды собирают информацию об ответе, отказе на запросы и т.д [14].

5. Объем контента важно для определения объема общего трафика с регистрацией объема контентов всех отправляемых и принимаемых пакетов.

6. Метод запроса, как правило, пишутся заглавными буквами, состоят из коротких GET, HEAD и т.д. английских слов. На основе этих методов определяется для чего был отправлен запрос от пользователя веб ресурса [15].

7. URL (Uniform Resource Locator) регистрирует имена доменов первого уровня и ссылки обращающихся пользователей сети.

8. Код иерархии предоставляет информацию о форме обработки запросов. Например, запрос был отправлен на прямую или через партнер-ский сервер и т.д [16].

9. IP отвечающего – IP адрес отвечающего на запросы

10. Содержание находится в заголовке HTTP ответа и показывает тип содержимого в объекте [17].

Все эти данные собираются с помощью прокси-сервера Squid. Прокси-сервер Squid [18] используется для реализации процесса накопления и управления лог-файлов сетевого трафика. Прокси-сервер Squid является программным обеспечением, с открытым кодом и его использование целесообразно в крупных сетях, где суточное число пользователей превышает 2000. Преимущество прокси-сервера Squid в том что, он является кэшируемым прокси-сервером, а в этом случае обращаемые ресурсы накапливаются в кэше и при повторном обращении процесс обработки завершается более ускоренно. Это в свою очередь положительно влияет на доступность сети. Лог-файлы, с помощью прокси-сервера Squid, накапливаются на специальной базе данных и используются в процессе анализа (таб. 2).

Таблица 2

Пример данных, собранных прокси-сервером Squid

Штамп времени UNIX	Время процесса (мсек)	IP адрес	Результирующие коды	Объем контента (байт)	Метод запроса	URL	Код иерархии	IP отвечающего	Содержание
1444780867.298	39	10.100.80.51	TCP_MISS/200	10946	GET	http://pagead2.googlesyndication.com	HIER_DIRECT	216.58.208.98	application/x-shockwave-flash
1444795608.042	3598	10.100.80.23	TCP_MISS/301	567	POST	http://v.icecentury.com/	HIER_DIRECT	54.169.165.185	text/html
1444795738.177	222	10.100.80.14	TCP_MISS/304	318	GET	http://code.createjs.com	HIER_DIRECT	23.77.228.124	application/x-javascript
1444799392.183	38	10.100.80.61	TCP_MISS/200	345	HEAD	http://ds.download.windowsupdate.com	HIER_DIREC	188.43.72.35	application/octet-stream

3. Очистка информации в лог-файле. Лог-файлы, накапливаемые с помощью прокси-сервера Squid, создают широкие возможности для интерпретации. Это в свою очередь создаёт условие для использования лог-файлов для различных целей. Пример данных, накопленных прокси-сервером Squid приведен в табл. 2. Однако, в рамках данной статьи нет необходимости конкретного рассмотрения всех 10-ти переменных представленных прокси-сервером Squid. При подходе со стороны информационной безопасности для идентификации профиля пользователя нет необходимости рассмотрения содержания обращения, IP назначения, http

иерархического кода, способа опроса и кодов результата. Поэтому, во время анализа лог-файлов для облегчения и ускорения процесса обработки необходимо учитывать указанные переменные.

4. Идентификация профиля пользователя. Когда мы говорим о профиле идентификации, имеем ввиду вектор интересов и тематические выборы построенные на основе обращаемых веб ресурсов. Сбор тематических профилей пользователей создаёт матрицу. В этой матрице на каждой строке указывается пользователь, а в каждом столбике показаны признаки. В зависимости от частоты обращения ресурсов входящих в категории поведения пользователей и объёма входящего трафика, вычисляется значимость признаков. Для повышения качества модели проводится процесс нормализации свойств в интервале [0;1].

После завершения процесса проектирования признаков, для построения модели выбираются более информативные и достоверные признаки. Это уменьшает объем обрабатываемой информации, создаёт условие для предотвращения повторения процесса обучения, а также, в целом, повышает качество модели. В рассматриваемом случае ресурсы группируются согласно тематической категории. Понятно, что ресурсы, которые относятся к одной тематической категории, могут быть размещены в различных источниках.

Первым этапом решения проблемы Data mining является проектирование признаков (feature engineering) [19]. Это является ответственным и трудоёмким этапом и наряду с этим, непосредственно, влияет на результаты процесса. В рассматриваемом случае объектами являются пользователи сети, а в качестве признаков рассматриваются веб ресурсы, к которым обращаются пользователи. В результате полученного изображения признаков, формируется тематический профиль пользователей и получается матрица пользователь/категории, состоящая из информативных признаков. Полученная матрица имеет большие размеры (табл. 3), но по форме соответствует разреженной матрице (sparse matrix).

5. Постановка задачи. Мы будем использовать алгоритм К-средних для кластеризации трафика сети [20]. Причиной является то, что для решения задачи кластеризации алгоритм К-средних оказывается очень быстрым и простым. Если $X = \{x_1, \ldots, x_n\}$, то множество данных состоит из ***n*** сессий трафика. x_i представляет собой каждую трафик-сессию в *d*-мерной Евклидовой среде. $x_i = (f_1, \ldots, f_d)$, когда ***i*** трафик-сессия имеет значения $f_1, \ldots, f_d$, ***d*** – значение свойств. Это является основной целью разделения трафик-сессии по кластерам. Во время этого процесса ставится условие, что бы расстояние между n данными и соответствующими центроидами К кластеров было минимально. У каждого кластера имеется центр μ_k известный как сентроид, и он может считаться представителем этой группы.

Таблица 3

Матрица информативных признаков пользователи / категории

	Кат1/ Объем (Гб)	Кат2/ Объем (Гб)	Кат3/ Объем (Мб)	Кат1/ Время (мин)	Кат2/ Время (мин)	Кат3/ Время (мин)	Кат1/ Запрос (количество)	Кат2/ Запрос (количество)	Кат3/ Запрос (количество)
Полз. 1	12	6	800	126	98	22	5355	4742	1586
Полз. 2	14	4,8	350	148	71	18	10163	3102	1475
Полз. 3	3,1	2,7	787	78	38	28	608	1554	3217

Таким образом, nxd матрица данных является входом алгоритма К-средних, К – количество кластеров, а сентроиды являются первичными данными:

1. Сначала необходимо определить К точки, представляющие центроидные группы.

2. Для расчета Евклидова расстояния между каждым данным и самым близким центроидом используется уравнение:

$$dist(x,y) = (\sum_{i=1}^{n}(x_i - y_i)^2)^{\frac{1}{2}}.$$

После определения всех точек, позиции К центроидов заново вычис-ляются и это означает, что середина всех точек определенной группы μ_k должна также заново вычисляться.

2- и 3-й пункты должны повторяться до тех пор, пока не из менится позиция центроидов.

6. Выбор количества кластеров. В этом разделе до применения алгоритм К-средних, будет показано, как выбирается количество кластеров [21]. Первым измеряется внутрикластерное расстояние, определяющее расстояние между точкой и центроидом. После этого определяется усредненное значение всех этих расстояний:

$$intra = \frac{1}{N}\sum_{i=1}^{K}\sum_{x \in C_i}\|x - x_i\|^2.$$

где N – количество сессий (точек); К – количество кластеров, а z_i является центроидом кластера C_i. Далее, необходимо измерить межкластерное расстояние и при этом необходимо учитывать, что чем больше это расстояние, тем лучше. Для этого используется приведенная ниже формула:

$$\text{inter} = \min(\|z_i - z_j\|^2), \; i = 1, 2, \ldots, K-1;$$
$$j = i+1, \ldots, K.$$

Для определения количества К кластеров в алгоритме К-средних необходимо использовать следующую формулы:

$$validity = \frac{\text{intra}}{\text{inter}}.$$

7. Результаты экспериментов. В результате применения модели кластеризации были сформированы определённые кластеры. Кластеры, в основном, формируют социальные сети, видео-ресурсы и научно-практические ресурсы (рис. 1). Результат показанный на рис. 1 получен для 20 кластеров с помощью bigml.com ресурса [22]. Больше всех обращаемый кластер А состоит из научно-практических ресурсов. 2-ой по порядку обращаемый кластер Б, это социальные сети. Кластер С состоит из обращений к видео-ресурсам. Обращение к другим кластерам значительно меньше. Это связано с тем, что пользователи основную часть необходимой информации получают от социальных сетей и видео-ресурсов.

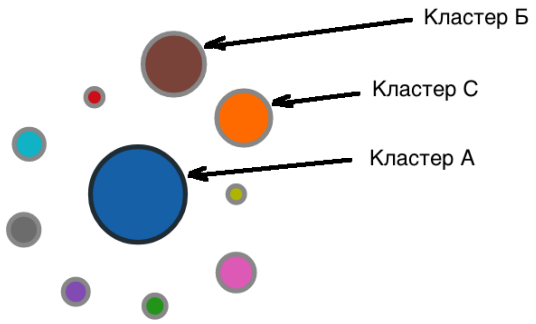

Рис. 1. Результаты применения кластеризации

Заключение. Данная статья посвящена проблеме определения профилей пользова-телей AzScienceNet на основе кластеризации. Для этого выбрана самая высокоскоростная и простая модель кластеризации на основе К-средних. В результате проведённых исследований были обеспечены: целесообразное распределе-ние сетевых ресурсов, оптимизация сетевого трафика, определение источников аномальной активности и обеспечение своевременной ликвидации угроз.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Соколов А.С.* Моделирование сегмента вычислительной сети и выявление проблемных участков в процессе мониторинга // Прикладная информатика. – 2011. – № 3. – С. 116-120.
2. http://www.itu.int/en/ITUD/Cybersecurity/Documents/Symantec_annual_internet_threat_report_ITU2014.pdf.
3. *Callado A., Kamienski C., Szabo G., Gero B., Kelner J., Fernandes S., Sadok D.* A Survey on Internet Traffic Identification // IEEE Communications Surveys & Tutorials. – 2009. – Vol. 11, Issue 3. – P. 37-52.
4. *Mingbo L., Wenjie S., Qianhong Z., Zhaoping T.* Design and implementation of IP network traffic monitoring system // 15th International Conference on Optical Communications and Networks (ICOCN). – 2016. – P. 23-35.
5. *Howlett T.* Open Source Security Tools: Practical Guide to Security Applications, 2004. – ed. 1. Prentice Hall. – 608 p.
6. *Kumpulainen P., Hätönen K., Knuuti O., Alapaholuoma T.* Internet traffic clustering using packet header information // Joint International IMEKO TC1+ TC7+ TC13 Symposium, Jena, Germany, 2011. – P. 13-20.
7. *Gerhard M., Sa L., Georg C.* Traffic Anomaly Detection Using K-Means Clustering // In Proceedings of performance, reliability and dependability evaluation of communication networks and distributed systems, 4GI/ITG-Workshop MMBnet, Hamburg, Germany, 2007. – P. 25-33.
8. *Ekola T., Laurikkala M., Lehto T., Koivisto H.* Network traffic analysis using clustering ants // Proceedings. World Automation Congress. – Seville, Spain, 2004. – Vol. 17. – P. 275-280.
9. *Duo Liu, Chung-Horng Lung, LambadanÏs I., Seddigh N.* Network traffic anomaly detection using clustering techniques and performance comparison // Proceedings the 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Canada, 2013. – P. 1-4.
10. *Shokri, R., Oroumchian F., Yazdani N.* CluSID: a clustering scheme for intrusion detection improved by information theory // Proceedings of the 7th IEEE Malaysia International Conference on Communications and IEEE International Conference in Networks, Kuala Lumpur, Malasia, 2005. – P. 553-558.
11. http://wiki.squid-cache.org/SquidFaq/SquidLogs.
12. https://ru.wikipedia.org/wiki/UNIX-время.
13. https://en.wikipedia.org/wiki/Network_packet.
14. https://ru.wikipedia.org/wiki/Список_кодов_состояния_HTTP.
15. https://ru.wikipedia.org/wiki/HTTP#Методы.
16. http://squid-handbuch.de/hb/node106_mn.html.
17. https://ru.wikipedia.org/wiki/Список_MIME-типов.
18. http://www.squid-cache.org/Intro/why.html.
19. *Han J., Kambe M., Pei J.* Data Mining: Concepts and Techniques, ed. 3. – Morgan Kaufmann Publishers is an imprint of Elsevier, 2012. – 740 p.
20. *Yang G., Zhou G., Yin Y., Yang X.* K-Means Based Fingerprint Segmentation with Sensor Interoperability // Journal on Advances in Signal Processing (EURASIP). – 2010. – Vol. 10, No. 54. – P. 1-12.
21. *Kodinariya M., Makwana R.* Review on determining number of Cluster in K-Means Clustering // International Journal of Advance Research in Computer Science and Management Studies. – 2013. – Vol. 1, Issue 6. – P. 90-95.
22. http://www.bigml.com.

REFERENCES

1. *Sokolov A.S.* Modelirovanie segmenta vychislitel'noy seti i vyyavlenie problemnykh uchastkov v protsesse monitoringa [The modeling segment of the computer network and identification of problem areas in the monitoring process], *Prikladnaya informatika* [Applied Informatics], 2011, No. 3, pp. 116-120.
2. Available at: http://www.itu.int/en/ITUD/Cybersecurity/Documents/Symantec_annual_internet_threat_report_ITU2014.pdf.
3. *Callado A., Kamienski C., Szabo G., Gero B., Kelner J., Fernandes S., Sadok D.* A Survey on Internet Traffic Identification, *IEEE Communications Surveys & Tutorials*, 2009, Vol. 11, Issue 3, pp. 37-52.
4. *Mingbo L., Wenjie S., Qianhong Z., Zhaoping T.* Design and implementation of IP network traffic monitoring system, 15*th International Conference on Optical Communications and Networks (ICOCN)*, 2016, pp. 23-35.
5. *Howlett T.* Open Source Security Tools: Practical Guide to Security Applications, 2004, ed. 1. Prentice Hall, 608 p.
6. *Kumpulainen P., Hätönen K., Knuuti O., Alapaholuoma T.* Internet traffic clustering using packet header information, *Joint International IMEKO TC1+ TC7+ TC13 Symposium, Jena, Germany, 2011,* pp. 13-20.
7. *Gerhard M., Sa L., Georg C.* Traffic Anomaly Detection Using K-Means Clustering, *In Proceedings of performance, reliability and dependability evaluation of communication networks and distributed systems, 4GI/ITG-Workshop MMBnet, Hamburg, Germany, 2007*, pp. 25-33.
8. *Ekola T., Laurikkala M., Lehto T., Koivisto H.* Network traffic analysis using clustering ants, *Proceedings. World Automation Congress*. Seville, Spain, 2004, Vol. 17, pp. 275-280.
9. *Duo Liu, Chung-Horng Lung, LambadanÏs I., Seddigh N.* Network traffic anomaly detection using clustering techniques and performance comparison, *Proceedings the 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Canada, 2013*, pp. 1-4.
10. *Shokri, R., Oroumchian F., Yazdani N.* CluSID: a clustering scheme for intrusion detection improved by information theory, *Proceedings of the 7th IEEE Malaysia International Conference on Communications and IEEE International Conference in Networks, Kuala Lumpur, Malasia, 2005*, pp. 553-558.
11. Available at: http://wiki.squid-cache.org/SquidFaq/SquidLogs.
12. Available at: https://ru.wikipedia.org/wiki/UNIX-время.
13. Available at: https://en.wikipedia.org/wiki/Network_packet.
14. Available at: https://ru.wikipedia.org/wiki/Список_кодов_состояния_HTTP.
15. Available at: https://ru.wikipedia.org/wiki/HTTP#Методы.
16. Available at: http://squid-handbuch.de/hb/node106_mn.html.
17. Available at: https://ru.wikipedia.org/wiki/Список_MIME-типов.
18. Available at: http://www.squid-cache.org/Intro/why.html.
19. *Han J., Kambe M., Pei J.* Data Mining: Concepts and Techniques, ed. 3. Morgan Kaufmann Publishers is an imprint of Elsevier, 2012, 740 p.
20. *Yang G., Zhou G., Yin Y., Yang X.* K-Means Based Fingerprint Segmentation with Sensor Interoperability, *Journal on Advances in Signal Processing (EURASIP)*, 2010, Vol. 10, No. 54, pp. 1-12.
21. *Kodinariya M., Makwana R.* Review on determining number of Cluster in K-Means Clustering, *International Journal of Advance Research in Computer Science and Management Studies*, 2013, Vol. 1, Issue 6, pp. 90-95.
22. Available at: http://www.bigml.com.

Статью рекомендовал к опубликованию д.т.н., профессор А.З. Меликов.

Алгулиев Расим Магамед оглы – Институт информационных технологий при НАНА; e-mail: rasim@science.az; AZ1141, Азербайджанская Республика, г. Баку, ул. Б. Вахабзаде 9; тел.: 994125390167; академик.

Имамвердиев Ядигяр Насиб оглы – e-mail: yadigar@iit.science.az; док. фил. по тех.; тел.: 994125390167.

Набиев Бабак Расим оглы – e-mail: babek@iit.science.az; AZ1141, Азербайджанская Республика, г. Баку, ул. Б. Вахабзаде 9; тел: 994125390167.

Alguliyev Rasim Mahammad – Institute of Information Technology of ANAS; rasim@science.az; AZ1141, B. Vahabzade street, 9A, Azerbaijan Republic, Baku; phone: 994125390167, Active member of ANAS, doctor of technical sciences, Professor.

Imamverdiyev Yadigar Nasib – e-mail: rasim@science.az; phone: 994125104253; dr. of tech. sc., professor.

Nabiyev Babak Rasim – e-mail: rasim@science.az; phone: 994125390167; postgraduate student.

УДК 004.056

Л.К. Бабенко, И.А. Писарев

АНАЛИЗ БЕЗОПАСНОСТИ ПРОТОКОЛА СИСТЕМЫ ЭЛЕКТРОННОГО ГОЛОСОВАНИЯ НА ОСНОВЕ СЛЕПЫХ ПОСРЕДНИКОВ С ПОМОЩЬЮ ИНСТРУМЕНТА AVISPA*

Разработка систем электронного голосования является важной проблемой в современном мире. Такие системы надежнее и удобнее традиционных способов голосования. Однако, их разработка является гораздо более сложной и доказать, что какая-либо система является надежной на достаточном уровне так же крайне сложно. В данной работе рассматривается анализ безопасности криптографического протокола, который используется в созданной авторами системе электронного голосования на основе слепых посредников. Анализируется протокол на самом ключевом этапе системы – голосования. Проведено описание протокола. Показан ход преобразования данных в процессе взаимодействия сторон во время этапа голосования. Указаны уточнения по поводу использования тех или иных техник для обеспечения защищенности информации на всем протяжении этапа голосования. Проверяется защищенность криптографического протокола на этом этапе. В качестве инструмента для верификации безопасности протоколов используется система Avispa. В статье приводится описание протокола на специальном языке CAS+, которое преобразуется в язык HLPSL (High−Level Protocol Specification Language) и анализируется данным инструментом. Поставлены цели анализа безопасности разработанного протокола такие как: аутентификация сторон, проверка секретности данных, защита от replay-атак. Приведены особенности описания протоколов с помощью инструмента Avispa. Произведена проверка безопасности протокола системы электронного голосования на основе слепых посредников, рассмотрена схема взаимодействия сторон, включая анализ сообщений, которые может перехватить злоумышленник. Показана эффективность защиты криптографического протокола от действий злоумышленника. Сделаны выводы по использованию инструмента Avispa для анализа безопасности протоколов.

Электронное голосование; криптографические протоколы; криптографическая защита; верификация безопасности криптографических протоколов.

L.K. Babenko, I.A. Pisarev

PROTOCOL SECURITY ANALYSIS OF ELECTRONIC VOTING SYSTEM BASED ON BLIND INTERMEDIARIES WITH THE AVISPA TOOL

The development of electronic voting systems is an important problem in the modern world. Such systems are more reliable and convenient than traditional methods of voting. However, their development is much more complicated and to prove that any system is reliable at a sufficient level is also extremely difficult. In this paper, we analyze the security of a cryptographic protocol,

* Работа поддержана грантом Министерства образования и науки РФ № 2.6264.2017/8.9.