Intellectual Technologies on Transport
No 1

Интеллектуальные технологии на транспорте
№ 1

Санкт-Петербург
St. Petersburg
2018

Интеллектуальные технологии на транспорте
№ 1, 2018

Сетевой электронный научный журнал, свободно распространяемый через Интернет.
Публикует статьи на русском и английском языках с результатами исследований и практических достижений в области интеллектуальных технологий и сопутствующих им научных исследований

Журнал основан в 2015 году

Учредитель и издатель

Федеральное государственное бюджетное образовательное учреждение высшего образования «Петербургский государственный университет путей сообщения Императора Александра I» (ФГБОУ ВО ПГУПС)

Адрес редакции
190031 Санкт-Петербург, Московский пр., 9, ПГУПС
email: itt-pgups@yandex.ru, сайт: http://itt-pgups.ru

ISSN 2413-2527

Журнал зарегистрирован Федеральной службой по надзору в сфере связи и массовых коммуникаций, свидетельство ЭЛ № ФС77-61707 от 07 мая 2015 г.

Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ)

Intellectual Technologies on Transport
Issue № 1, 2018

Network electronic scientific journal, open access. It publishes articles in Russian and English with the results of research and practical achievements in the field of intelligent technologies and associated research

Founded in 2015

Founder and Publisher
Federal State Educational Institution of Higher Education
«Emperor Alexander I Petersburg State Transport University»

Editorial adress
190031, St. Petersburg, Moskovskiy pr., 9, 2–108
email: itt-pgups@yandex.ru, http://itt-pgups.ru

ISSN 2413-2527

The journal is registered by the Federal Service for Supervision of Communications and Mass Media, EL no. FS77-61707 testimony from May 7, 2015

The journal is registered in the Russian Science Citation Index (RSCI)

Содержание

Демин А. В., Дмитриева С. П.
Метод построения прогнозной оценки поведения автономной многопараметрической динамической технической системы 5

Карпович С. Н.
Корпус текстов русского языка для тестирования алгоритмов тематического моделирования (на англ.). 11

Смагин В. А., Бубнов В. П.
Моделирование процессов на основе последовательного гиперфрактального распределения. 20

Молдовян А. А., Татчина Я. А.
Способы псевдовероятностного блочного шифрования 25

Молдовян Н. А., Вайчикаускас М. А.
Генерация степенных сравнений как способ открытого шифрования и протокол отрицаемого шифрования 32

Максимов В. А.
Алгоритм хранения информации на основе комбинирования способов размещения и кодирования неоднородных по важности данных (на англ.) 38

Шардаков К. С.
Сравнительный анализ популярных систем мониторинга сетевого оборудования, распространяемых по лицензии GPL. 44

Contents

Demin A. V., Dmitrieva S. P.
A Method of Constructing Estimates ofthe Future Behavior of a Dynamic Autonomous Multiparameter Technical Systems 5

Karpovich S. N.
The Russian Language Text Corpus for Testing Algorithms of Topic Model (English) 11

Smagin V. A., Bubnov V. P.
Modelling of Processes on the Basis of Consecutive Giperfractal Distributions 20

Moldovyan A. A., Tatchina Ya. A.
Deniable-encryption Methods Based on Block Ciphers 25

Moldovyan N. A., Vaichikauskas M. A.
Generation of Polynomial Equations as a Method for Public Key Encryption and Deniable Encryption Protocol 32

Maksimov V. A.
Algorithm of Data Storage Based on Combining of the Methods of Locating and Coding non-Homogeneous Data (English) 38

Shardakov K. S.
Comparative Analysis of the Popular Monitoring Systems for Network Equipment Distributed Under the GPL License 44

Метод построения прогнозной оценки поведения автономной многопараметрической динамической технической системы

Демин А. В., Дмитриева С. П.
Университет ИТМО
Санкт-Петербург, Россия
dav_60@mail.ru, spdmitrieva@corp.ifmo.ru

Аннотация. **Важнейшим требованием к характеристикам многопараметрических динамических технических систем (МДТС), в частности, к системам мониторинга окружающей среды, работающим в автоматическом режиме в условиях неоднозначности обстановки относительно внешних возмущений, является устойчивость функционирования весь период жизненного цикла. Повышение эффективности работы МДТС достигается за счет улучшения ее функционально-параметрических характеристик путем разработки и реализации алгоритмов прогнозирования состояния её компонентов. В зависимости от объёма априорных данных, достоверность и точность прогноза определяется выбранным методом построения и алгоритма реализации в виде прогнозной функции МДТС.**

Ключевые слова: **прогнозная оценка, поведение сложной технической системы, алгоритм, прогнозная аналитическая модель, аппроксимация априорных данных.**

Введение

Для повышения надёжности и возможности предотвращения нежелательных последствий функционирования многопараметрических динамических технических систем (МДТС), особенно работающих в автоматическом режиме в условиях неоднозначности обстановки относительно внешних возмущений, можно использовать разные методы и средства, позволяющие прогнозировать состояние техногенных систем (например, атомных электростанций, космических станций, нефте-, газопроводов и др.). Техническими средствами текущего контроля показателей качества МДТС являются сенсорные устройства: на основе получаемой от них информации можно определить прогнозную функцию. В связи с этим в зависимости от объёма априорных данных достоверность и точность прогноза определяется методом построения и алгоритма реализации в виде прогнозной функции для МДТС [1, 2].

Для обеспечения функционально-эксплуатационной устойчивости МДТС необходимо периодическое тестирование, поэтому нужно решить следующие теоретические и практические задачи:

- развитие методов построения и алгоритмов реализации долгосрочного прогнозирования в автономных МДТС в соответствии с апостериорными данными при неопределенности воздействия внешних возмущений (изменении давления, температуры, влажности);
- получение своевременной информации в реальном времени о метеоусловиях и дальнейшее прогнозирование метеообстановки, которое позволило бы обеспечить предсказуемость результатов народно-хозяйственной деятельности и безопасность жизнедеятельности.

Современные формализованные подходы к анализу и построению прогнозной функции состояния МДТС базируются на математических методах оптимизации и решения вариационных задач для систем с использованием прогнозирующих моделей – Model Predictive Control (MPC) [3–6].

Набор инструментальных средств исследования и проектирования алгоритмов управления в дискретных и непрерывных системах на основе предсказаний динамики их поведения – Model Predictive Control Toolbox (MPC Tools) – предоставляет инструменты для систематического анализа, разработки и настройки регуляторов с моделью предсказания. Можно проектировать регуляторы и выполнять симуляцию регуляторов с моделью предсказания при помощи функций MATLAB® или блоков Simulink®. Можно задавать и изменять модель предсказания, горизонты управления и прогнозирования, ограничения на входе, на выходе и весовые значения. Инструментарий позволяет диагностировать проблемы, которые могут привести к сбоям во время работы и даёт советы по изменению значений весовых коэффициентов и ограничений для улучшения работы и повышения надежности. При помощи запуска разных сценариев в линейных и нелинейных симуляциях можно оценить работу регулятора.

Структурное описание МДТС автономной мобильной метеостанции

Концептуальная модель МДТС отображает все тактико-технические требования технического задания на проектируемую систему, что фактически является модельным отображением её целевой функции (ЦФ). На рис. 1 дано структурно-функциональное изображение МДТС, которая является предметом исследований, на рис. 2 – её модельное отображение. В качестве практически исследуемых МДТС в работе использована автономная мобильная метеостанция, содержащая датчик измерения оптической дальности видимости (лидар) и метеодатчики, которую можно отнести к классу распределенных многоагентных измерительных систем. Структурная схема взаимодействия метеостанции включает в себя центр сбора и аналитической обработки техногенных наблюдений, датчик измерения оптической дальности видимости (лидар), мобильные сенсорные датчики, фиксирующие метеорологические данные, и встроенный радиомодем, позволяющий принимать и обрабатывать радиосигнал.

Рис. 1. Структурно-функциональная схема взаимодействия автономной мобильной метеостанции

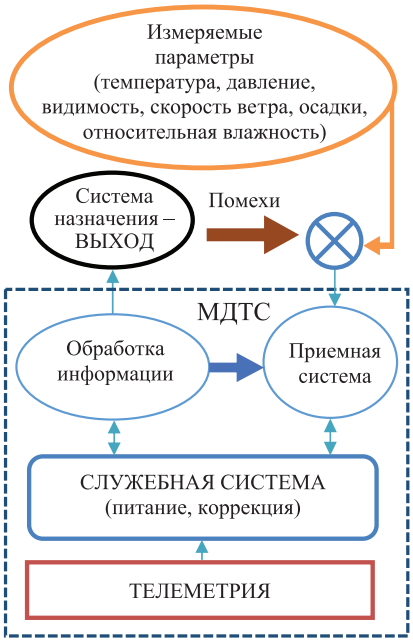

Рис. 2. Модельное отображение МДТС автономной мобильной метеостанции

Численный метод построения прогнозной оценки МДТС

Процедура параметрического моделирования МДТС является центральной процедурой разработки прогнозной модели. Допустим, что моделирование технических процессов в сочетании с различными условиями, для которых проводились эксперименты, образует некую систему (рис. 3), которую можно описать через параметрические свойства её элементарных единиц: множества входных параметров МДТС – $\{V_i\}_1^{N_{Вх}}$; множества внешних возмущений окружающей среды – $\{W_i\}_1^{N_{Возм}}$; множества внутрисистемных

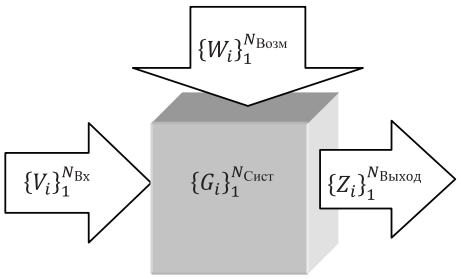

Рис. 3. Общая схема функционирования МДТС в реальных условиях эксплуатации

показателей – $\{G_i\}_1^{N_{Сист}}$; множества выходных системных параметров – $\{Z_i\}_1^{N_{Выход}}$.

В процессе аналитического моделирования выдвинуто предположение, что один из выходных системных параметров, например параметр z_1, может быть искомым, а остальные выходные параметры z_2, …, z_n в процессе разработки прогнозной модели МДТС будут учитываться только при наличии обратной связи от них (телеметрии системы), что и показано в выражении

$$z_1 = f_1(g_1,...,g_k) \circ f_1(v_1,...,v_l) \circ f_1(w_1,...,w_m).$$

Численный метод построения прогнозной функции реализуется путем представления аналитической функции в виде композиции трёх основных компонент: первая составляющая характеризует влияние внутренних параметров и представляет собою некую общую часть искомой модели, вторая (уточняющая) описывает воздействие внешних параметров системы, третья (регулирующая) учитывает величину обратной связи (например, телеметрии):

$$\begin{aligned} &Z(x_1,x_2,...,x_{N_{Выход}}) = V(x_1,x_2,...,x_{N_{Вход}});\\ &W(x_1,x_2,...,x_{N_{Возм}}) \circ G(x_1,x_2,...,x_{N_{Сист}}). \end{aligned} \quad (1)$$

В зависимости от конкретных целей и условий применения прогнозных функций при их построении можно принимать во внимание или, напротив, исключать из рассмотрения отдельные компоненты в структуре функции, можно управлять точностью будущих прогнозов и получать оценки для физического явления с различной степенью детализации, учитывая или игнорируя, таким образом, отдельные факторы влияния.

Наибольший интерес представляет построение прогнозной модели для открытых МДТС, т. е. когда учитываются внешние факторы, прогнозную модель МДТС можно описать как мультифакторную, что следует из выражения (1), и мультипараметрическую, что следует из выражения

$$f(x) = f_1(x) \circ f_2(x),$$

где $f_1(x)$ фиксирует воздействие на объект постоянных факторов окружающей среды; $f_2(x)$ – влияние переменных величин.

Обобщенная прогнозная модель МДТС может быть построена по её модельному представлению [7]. Если модель $\{M_{OB}(g_i)_1^N\}$ является образом реальной системы $\{OB_{real}(g_i)_1^N\}$, то она адекватна при соотношении

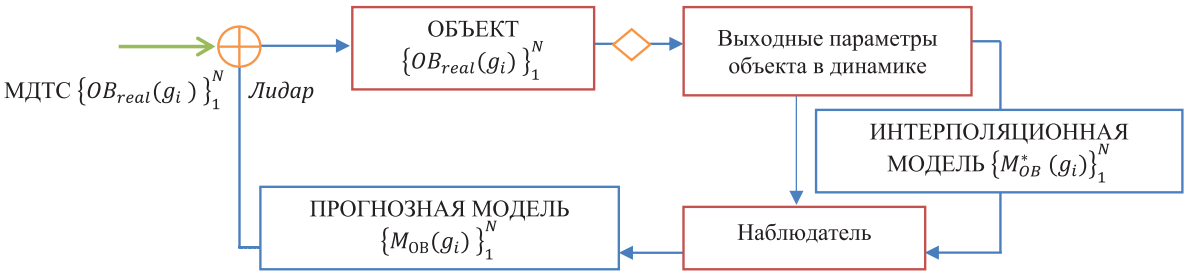

Рис. 4. Структурно-функциональная схема прогнозирования состояния МДТС с функцией прогнозного управления

$$\{M_{OB}(g_i)_1^N\} \cap \{OB_{real}(g_i)_1^N\} = \max.$$

Процессы, протекающие в каждом типе подсистем (блоков), основаны на уравнениях математической физики, при этом важно заметить, что математические модели $\{OB_{real}(g_i)_1^N\}$ в общем случае могут быть представлены на макро-, микро- и метауровнях. Новый подход прогнозирования заключается в построении прогнозной модели МДТС на основании интерполяционной функции по априорным исходным данным, а модели $\{M^*_{OB}(g_i)_1^N\}$ – как образа реальной МДТС [8] (рис. 4). Полнота и достоверность априорных исходных данных определяет точность прогнозирования.

Этапы реализации метода построения прогнозной функции для МДТС:

1) построение интерполяционной функции в соответствии с априорными исходными данными реальной МДТС $\{OB_{real}(g_i)_1^N\}$;

2) создание прогнозной модели $\{M^*_{OB}(g_i)_1^N\}$ на основе операции экстраполяции интерполяционной функции;

3) выполнение операции компьютерного моделирования для $\{M^*_{OB}(g_i)_1^N\}$;

4) сравнение результатов моделирования $\{M_{OB}(g_i)_1^N\}$ с исходными данными в контрольных точках;

5) выполнение условия: если невязка результатов моделирования с исходными данными лежит в поле допуска, то модель $\{M_{OB}(g_i)_1^N\}$ принимается за прогнозную функцию $\{OB_{real}(g_i)_1^N\}$;

6) выполнение условия: если невязка результатов моделирования с исходными данными не лежит в поле допуска, то процедуру повторяют с необходимой корректировкой алгоритма;

7) корректировка полученной прогнозной модели $\{M_{OB}(g_i)_1^N\}$ путем прогона её через объект $\{OB_{real}(g_i)_1^N\}$, при котором выполняется следующее условие: если выходные данные с объекта совпадают в пределах поля допуска данных со значениями исследуемой прогнозной модели, то можно считать, что прогнозная модель искомой системы сформирована; если выходные данные с объекта не соответствуют допустимым значениям, то корректировка прогнозной модели должна быть продолжена.

Алгоритм построения прогнозной оценки МДТС

При разработке прогнозной модели следует первоначально изучить реально существующую систему: исследовать её структурно-функциональную схему, эксплуатационные ограничения, взаимосвязи элементов и влияние внешних факторов, таким образом, лишь после предварительного исследования МДТС можно приступать к ее моделированию. Построение прогнозной модели МДТС, начинающееся с описания её целевой функции (ЦФ) и заканчивающееся достижением конечного результата, включает в себя четыре этапа (рис. 5).

Апробация метода построения прогнозной оценки поведения МДТС

Полный пакет для прогнозирования МДТС образует совокупность метода и модели. При сопоставлении подходов моделирования МДТС существенные преимущества имеет компьютерная обработка данных, которая позволяет манипулировать большим числом изменяемых величин и прогнозировать динамику процессов нелинейного характера с появлением эффектов синергизма [9, 10].

Для решения исследовательской проблемы был создан программный продукт на языке программирования Си++, в задачи которого вошли сбор, анализ и форматирование экспериментальных данных, выявление зависимости между температурой и внешними факторами среды, влияющими на построение прогнозной функции, проверка адекватности спрогнозированных величин их реальным величинам.

Для апробирования предложенного численного метода и разработанных алгоритмов построения прогнозной функции в качестве изучаемого параметра МТДС ввиду доступности, охвата и полноты БД решено выбрать метеорологическую величину (интернет-ресурс с БД статистических метеонаблюдений). На рис. 6 изображена процедура выбора оптимальной аппроксимации для статистики температуры воздуха, на рис. 7 – аппроксимация средней температуры воздуха в доверительных интервалах. Важно, что по методу и алгоритму построения аналитических прогнозных моделей МТДС для реальных условий эксплуатации на основании апостериорной информации можно спрогнозировать любую другую функциональную характеристику системы.

В результате прогнозирования по данным четырех объектов (изменениям температуры воздуха в 6.00, 12.00, 18.00, 24.00 ч) построены четыре прогнозные сезонные модели с выделенной базовой уточняющей и регулирующей компонентой, позволяющие с точностью ≈70–80 % предсказывать функционально-параметрическое состояние автономной метеостанции. На рис. 8 дана оценка достоверности прогнозной модели, среднее отклонение прогнозных данных от реальных составляет ≈5-6 °C.

Рис. 5. Алгоритм построения прогнозной модели по ЦФ

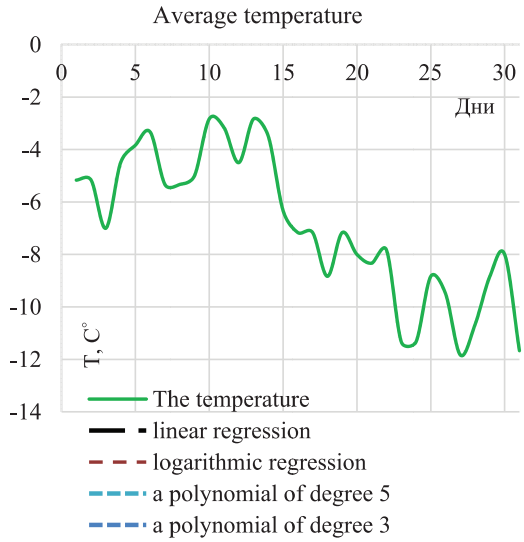

Рис. 6. Процедура выбора оптимальной аппроксимации для статистики температуры воздуха (°C)

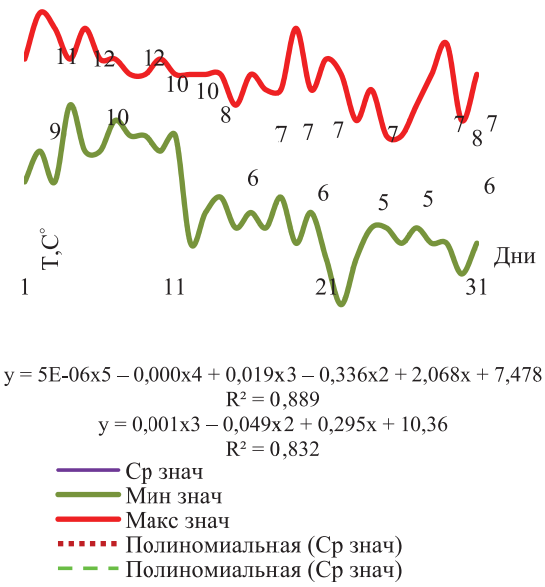

Рис. 7. Аппроксимация средней температуры воздуха (°C) в доверительных интервалах

Таким образом, на примере автономной мобильной метеостанции на основании экспериментальных данных был практически применен разработанный численный метод, алгоритмы прогнозирования состояния МДТС и ПО, позволяющее посредством прогнозного управления экономить ресурс автономной МДТС, гарантируя устойчивость её функционирования на протяжении всего жизненного цикла.

Заключение

Использование результатов исследования позволяет при неопределенности воздействия внешних возмущений обеспечить стабильность функционирования автономных МДТС за счет прогнозного управления внутренними техническими процессами системы. Результаты работы были использованы ООО «ЛОМО-МЕТЕО», «ОКБ Тест» и «АвтоВизус».

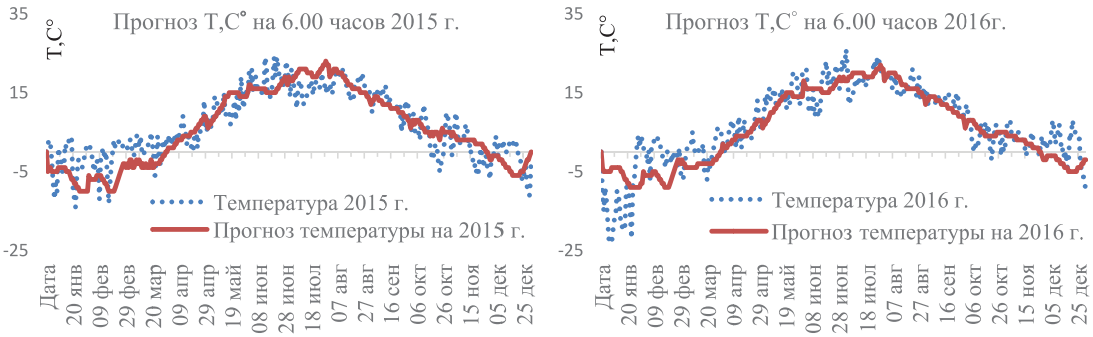

Рис. 8. Оценка достоверности прогноза температуры воздуха

В статье выполнено обоснованное построение прогнозной модели применительно к датчику температуры. Это позволяет управлять режимами работы станции путем включения-выключения указанного датчика для обеспечения экономичности и повышения длительности автономной работы. В перспективе целесообразно реализовать прогнозные модели метеостанции для всех видов датчиков. При этом следует отметить, что разработанный метод распространяется на все факторы.

Литература

1. Торшина И. П., Якушенков Ю. Г. Некоторые особенности системного подхода при проектировании оптико-электронных систем 3-го поколения // Контенант. 2017. Т. 16, № 1. С. 22-27.

2. Торшина И. П., Якушенков Ю. Г. Некоторые особенности моделирования оптико-электронных систем визуализации 3-го поколения // Сб. материалов XIII междунар. науч.-техн. конф. «Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации „Распознавание – 2017“».– Курск: ЮЗГУ, 2017. С. 347-348.

3. Захаров А. И., Загайнов А. И. Реализация программного комплекса для вычисления фрактальных параметров сложных систем // Интеллектуальные технологии на транспорте. 2015. № 2. С. 47-53.

4. Красновидов А. В. Подход к построению алгоритмов статистического анализа потоков ошибок в дискретных каналах связи // Интеллектуальные технологии на транспорте. 2015. № 2. С. 20-26.

5. Fu Z., Zhou X., Chen Y., Gong J., Peng F., Yan Z., Zhang T., Yang L. The influence of random slowdown process and lock-step effect on the fundamental diagram of the nonlinear pedestrian dynamics: An estimating-correction cellular automaton // Commun. Nonlinear Sci. Numer. Simul. 2015. Vol. 20 (3). P. 832-845.

6. Bonilla J., Dormido S., Cellier F. E. Switching moving boundary models for two-phase flow evaporators and condensers // Commun. Nonlinear Sci. Numer. Simul. 2015. Vol. 20 (3). P. 743-768.

7. Demin A. V., Dmitrieva S. P. Algorithm of real-time developing a forecast model of engineering systems // ICUMT 7.– Brno, 2015. P. 14-19. http://www.icumt.info/2015.

8. Demin A. V., Dmitrieva S. P. The algorithm of state estimation for dynamic multivariable technical system // CSNT 2016.– Chandigarth City, India: CSNT, 2016. P. 23-29.

9. Akimov S. V., Verkhova G. V. The four-level integrative model, methodology of structural and parametric synthesis of system objects // Collection: Proc. XIX Int. Conf. Soft Computing and Measurements SCM`2016. 2016. P. 321-323.

10. Akimov S. V., Verkhova G. V. The linguistic support of morphological modeling set // Collection: Proc. XIX Int. Conf. Soft Computing and Measurements SCM`2016. 2016. P. 337-340.

A Method of Constructing Estimates of the Future Behavior of a Dynamic Autonomous Multiparameter Technical Systems

Demin A. V., Dmitrieva S. P.
ITMO University
St. Petersburg, Russia
dav_60@mail.ru, spdmitrieva@corp.ifmo.ru

Abstract. **Stability of functioning for the whole period of the life cycle – this is the most important requirement is that the characteristics of multivariable dynamic technical systems (MDTS) and automatic systems for environmental monitoring, working in conditions of uncertainty of external disturbances. The efficiency MDTS is achieved by improving its functional and parametric performance through the development and implementation of algorithms for forecasting of its components. Depending on the amount of prior data, the reliability and accuracy of the forecast depends on the selected build method and algorithm implementation in the form of the forecast function MDTS.**

Keywords: **predictive assessment of the behavior of complex technical systems, algorithm, predictive analytic model, an approximation of a priori data.**

References

1. Torshina I. P., Yakushenkov Yu. G. Some features of the system approach in the design of third generation optoelectronic systems. *Contenant*, 2017, vol. 16, no. 1, pp. 22-27. (In Russ.)

2. Torshina I. P., Yakushenkov Yu. G. Some features of simulation of optic-electronic systems of visualization of the third generation. *Proc. XIII Int. Sci. and Tech. Conf. "OPTICS-EXPO – 2017"*. Kursk, SSU, 2017. Pp. 347-348. (In Russ.)

3. Zakharov A. I., Zagainov A. I. Implementation of Software for Calculating the Fractal Parameters of Complex Systems. *Intellectual Technologies on Transport*, 2015, no. 2, pp. 47-53. (In Russ.)

4. Krasnovidov A. V. An Approach to the Construction of Algorithms for the Statistical Analysis of Error Flows in Digital Communications Channels. *Intellectual Technologies on Transport*, 2015, no. 2, pp. 20-26. (In Russ.)

5. Fu Z., Zhou X., Chen Y., Gong J., Peng F., Yan Z., Zhang T., Yang L. The influence of random slowdown process and lock-step effect on the fundamental diagram of the nonlinear pedestrian dynamics: An estimating-correction cellular automaton. *Commun. Nonlinear Sci. Numer. Simul.*, 2015, Vol. 20 (3), pp. 832-845.

6. Bonilla J., Dormido S., Cellier F. E. Switching moving boundary models for two-phase flow evaporators and condensers. *Commun. Nonlinear Sci. Numer. Simul.*, 2015, Vol. 20 (3), pp. 743-768.

7. Demin A. V., Dmitrieva S. P. Algorithm of real-time developing a forecast model of engineering systems. *ICUMT 7.* Brno, 2015. Pp. 14-19. http://www.icumt.info/2015.

8. Demin A. V., Dmitrieva S. P. The algorithm of state estimation for dynamic multivariable technical system. *CSNT 2016.* Chandigarth City, India: CSNT, 2016. Pp. 23-29.

9. Akimov S. V., Verkhova G. V. The four-level integrative model, methodology of structural and parametric synthesis of system objects. *Collection: Proc. XIX Int. Conf. Soft Computing and measurements SCM`2016.* 2016. Pp. 321-323.

10. Akimov S. V., Verkhova G. V. The linguistic support of morphological modeling set. *Collection: Proc. XIX Int. Conf. Soft Computing and measurements SCM`2016.* 2016. Pp. 337-340.

The Russian Language Text Corpus for Testing Algorithms of Topic Model

Karpovich S. N.
JSC "Olimp"
Moscow, Russia
cims@yandex.ru

Abstract. **This paper proposes a special corpus for testing algorithms Topic model SCTM-ru. In the conditions of the prompt growth of quantity of data, the problem of development of tools and systems for their automatic processing. To create systems and testing algorithms should be suitable datasets. Existence of free collections of documents, text corpora in Russian, is necessary for researches methods of natural language processing, considering linguistic features of language. Designated special housing requirements: must be distributed under a free license, the number of documents should be sufficient for the study, must include the text of documents in natural language should contain demanded algorithms Topic model information. The comparative analysis of corpus in Russian and foreign languages is carried out, discrepancy of characteristics of the existing corpus with the designated requirements is revealed.**

Keywords: **text corpora, topic model, natural language processing, Russian language.**

Introduction

Information is becoming the main product and commodity of the contemporary society. The following areas are in the process of active development: science, economics, politics, and manufacturing; digital data is being generated and accumulated in many fields. To successfully retrieve and process information from data, one shall have proper tools, systems and algorithms. A demand for Natural Language Processing systems is growing. Natural Language Processing is already used in common software and services. For instance, software for reading news feeds is capable of grouping news by topics, search engines find documents with information of value for the user and mailing services filter spam messages automatically. Various algorithms are used for clustering and classification of text data; most popular are k-means, SVM, neural networks. An upcoming trend in automatic processing of texts is development of probabilistic topic modelling algorithms.

Topic modelling is a method of building of a topic model of a text documents collection. The topic model sets the ratio between topics and documents in the corpus of texts. For the first-time topic modelling was described in the paper by C. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala in 1998 [1]. Thomas Hoffman in 1999 proposed probabilistic latent semantic indexing (PLSI) [2]. One of most popular topic models is Latent Dirichlet Allocation (LDA), this model is the generalization of probabilistic semantic indexing and was developed by David Blei, Andrew Ng and Michael I. Jordan in 2002 [3]. Other topic models are usually an extension to LDA. Fig. 1 is an example of building a topic model of a document.

Algorithms of topic modelling are oriented at the work with a natural language text. Initial solutions were based on a suggestion that text is a "bag of words", i. e. word order in the text is of no value. Further models have successfully implemented algorithms that consider dependencies between words with the help of latent Markovian models. The review [4] considers five primary

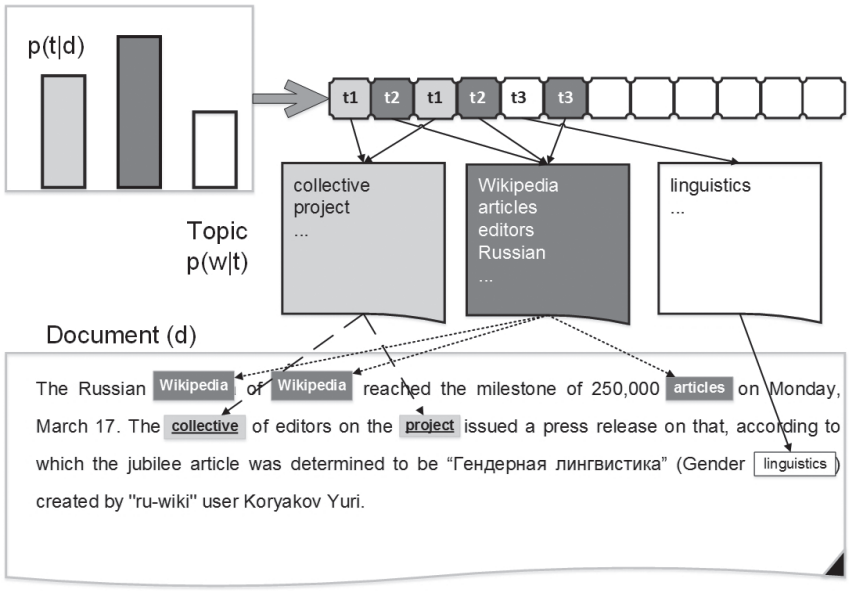

Fig. 1. Building a topic model of a document: p(w|t) – matrix of sought-for conditional distributions of words by topics; p(t|d) – matrix of sought-for conditional distributions of topics by documents; d — document; w — word; d, w — observable variables; t — topic (latent variable)

classes of probabilistic topic models: basic, taking into account relations between documents, taking into account relations between words, temporary, teacher-taught.

Availability of text corpora will make it possible to develop systems for automatic Natural Language Processing, as well as topic modelling algorithms. When developing a topic model, one shall consider language features of texts. A Russian text corpus, which is distributed through a free license, is necessary for development of topic modelling methods operating with the Russian language.

Corpus linguistics is a complicated linguistic discipline, which has formed in the last decades based on computer equipment. It studies the construction of linguistic corpora, methods of data processing in them and their generation and usage technology proper. "A corpus is a reference system based on an electronic collection of texts composed in a certain language" – such definition of the text corpus is available on the website of the Russian National Corpus [5]. The study [6] notes the following: "Corpora, as a rule, are designed for repeated use by many users; therefore, their markup and their linguistic support shall be unified in a certain manner". Feasibility of establishment and sense behind using the corpus depend on the following premises:

1) sufficient (representative) volume of the corpus;

2) data of different type is contained within the corpus in their natural context form;

3) once developed and prepared, data store may be used repeatedly.

First-order corpora are a collection of texts with a common feature, for instance, source, author, place of publication. A special text corpus is a balanced corpus, representative, as a rule, small, identifiable to a certain research task and designed for use mainly for purposes that are in line with composer's ideas. Text corpora or corpus, large collection of documents, dataset, as specified in the paper [4], are synonymic concepts.

The purpose of this paper is to create a special Russian text corpus SCTM-ru, suitable for study of algorithms of probabilistic topic modelling. Let us specify requirements to the created corpus. The corpus shall be distributed through a free license, the quantity of documents shall be sufficient for research, and it shall contain the following:

- original text of documents;
- dates of described events;
- authorship information;
- topics.

Let us consider the opportunity of using existing text corpora for purposes of testing topic modelling algorithms.

Review of text corpora and datasets

The Russian National Corpus (RNC) [5] contains more than 335 K documents in Russian, which divided into sub corpora. It includes 180 K texts of the Newspaper Corpus. A license agreement shall be signed to use the offline version of the main corpus (1 M tokens).

The OpenCorpora [7, 8] contains around 3 K documents in Russian, 93 K marked sentences, 10 data sources, some documents contain information about the author and date of described events. Linguistic information, such as morphological, semantic and syntactic, is assigned to text parts. The corpus is not suitable for objectives of building temporal and author-based models, since not all documents of the corpus contain information about the author and the date of the events.

The Associated Press [9] corpus contains 2 K documents in English. Corpus documents are not labelled with a date of a described event, publication author, and document category. The corpus is applicable to research a limited quantity of topic modelling algorithms.

The New York Times Annotated Corpus [10] is a large English text corpus of newspaper articles and news distributed through a closed license.

20 Newsgroups [11] is a collection of news in English, prepared to research algorithms of automatic text processing. 20 Newsgroups contains around 20 K documents. Data about authors and date of publication that is important for building topic models is not marked up. Preliminary processing of news text is required for use in topic modelling.

Reuters Corpora [12] is a large English news corpus. Three datasets contain more than 3 M news. It is distributed through a limited license only for scientific research and is provided only upon signature of the license agreement. There is an earlier version of the corpus named Reuters-21578 [13], which is popular for testing algorithms of automatic text processing and is distributed through a limited license, being available for offline analysis.

The computer corpus of texts from Russian newspapers of the end of the 20th century [14, 15] was established in 1999; it is currently developed and researched using grants of the Russian Foundation for Basic Research. The corpus is designed to analyses linguistic features (vocabulary, morphology, syntax, phraseology, stylistics) of the contemporary newspaper language. The corpus contains 23 K texts of full issues of 13 different Russian newspapers in Russian. The corpus includes 11 M tokens.

The corpus of the Russian literary language [16, 17] is represented in the form of an array of morphologically annotated texts in the Russian literary language. The corpus includes more than 1 M tokens with a balanced genre composition.

The Helsinki Annotated Corpus of Russian texts HANCO [18] – the corpus contains morphological, syntactic and functional information about texts with total volume of 100 K texts retrieved from the magazine "Itogi". Rights to full texts of magazine articles belong to owners.

Table represents comparison of algorithms of topic modelling of corpus characteristics that are important for research: corpus language, distribution license, availability for download and research on computers with no access to Internet, data on author, data on date of described events, text topic. Considered text corpora do not fully meet the requirements specified in this paper. The developed special corpus for topic modelling SCTM-ru is distributed through a free license, the language of the corpus is Russian, it contains data on authorship, the date of the events, topic affiliation of documents, is available for download and research on computers with no access to Internet.

Technology of SCTM-ru corpus development

The technological process of corpus development includes the following steps:

1) source detection;

2) preliminary processing of document texts;

3) markup of parameters of each document in the corpus;

4) provision of access to the corpus.

Table. Comparative table of text corpus characteristics

Corpus	Language	Open license	Available for download	Data on author	Data on date	Topics
RNC	Russian	–	–	+	–	–
Open Corpus	Russian	+	+	+	+	–
Associated Press	English	+	+	–	–	–
The New York Times Annotated Corpus	English	–	–	+	+	+
20 News-groups	English	+	+	–	–	–
Reuters Corpora	English	–	–	+	+	+
Russian Literary Language Corpus	Russian	–	–	–	–	–
HANCO	Russian	–	–	–	–	–
SCTM-ru	Russian	+	+	+	+	+

In accordance with the indicated requirement to availability of corpus data, texts used as content shall be distributed through a free license, available for download and free use.

As result of preliminary processing of texts and markup of parameters of each document, the corpus shall store and mark up in a certain way the information necessary to construct topic models. Information that was unclaimed in topic modelling shall be excluded from the corpus, as useless.

Various objectives of topic modelling may require a certain procedure of data arrival into the topic modelling system, from successive for temporal models, to one-off for regular topic models. Therefore, to provide access to the corpus, it is sufficient to provide an opportunity for its download and subsequent use in accordance with specific objectives of the researcher.

Source of data for SCTM-ru corpus

We suggest using the international news website "Russian Wikinews" (Wikinews) as a data source, where texts of articles are distributed through a free license of Creative Commons Attribution 2.5 Generic, are available for download and analysis on any computers, including computers with no access to Internet. Papers [19, 20] specify advantages of wiki-resources, such as Wikidictionary and Wikipedia, for use as a source of data for research purposes. Wiki-resources are websites of the second generation of Internet characterized by the fact that many ordinary users were involved to develop their content, and those users help to expand them and update information. Large volume, continuous expansion, neutrality of opinions, and availability are among the advantages of all wiki-resources, including Wikinews.

Wikinews is a brother project of large Wikipedia designed to write news articles. The example of a Wikinews article is shown in fig. 2. A signature feature of the Wikinews website compared to any other news website is the fact that any person may take part in creating a piece of news. Rules of Wikinews require writing news from a neutral point of view, in unbiased form, selecting material and relevant topics, using valid sources.

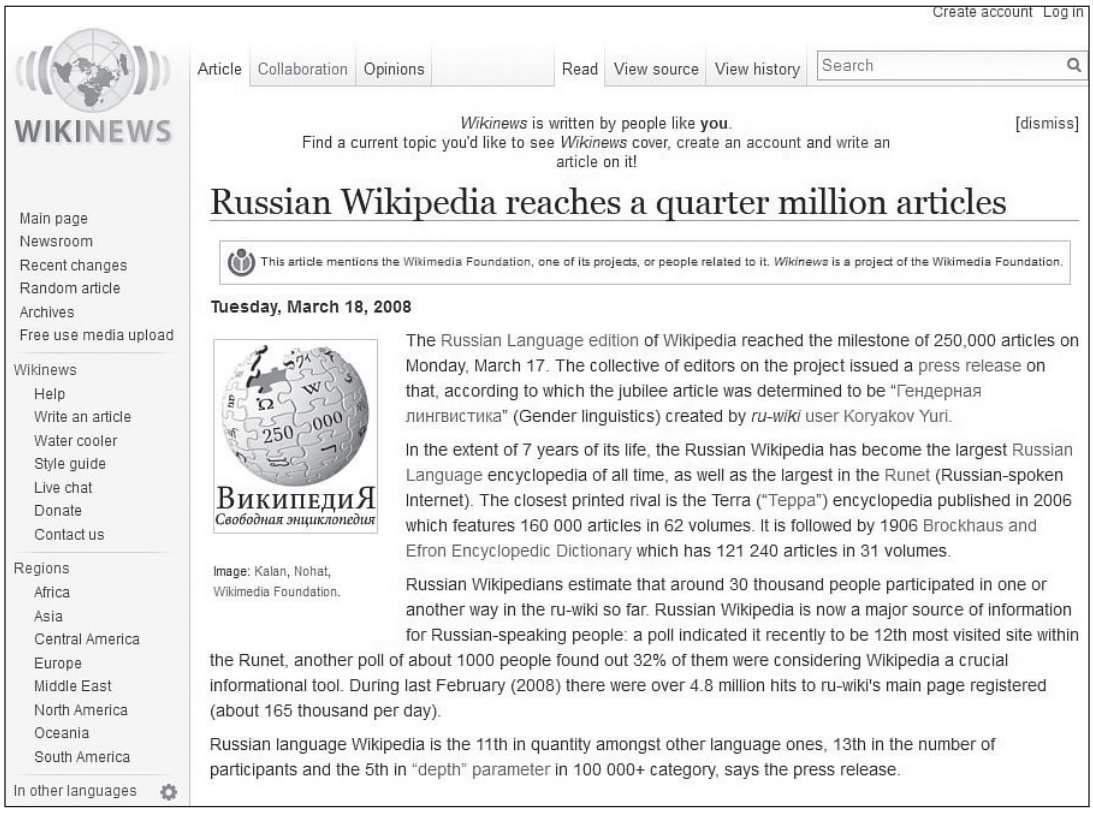

Fig. 2. Article „50,000 Articles in Russian Wikipedia" on website of Russian Wikinews

XML-file of Wikinews data base export includes the following XML-elements:

+ <page> – group of news article elements;
+ <title> – article title;
– <ns> – identifier or name of a namespace, the element is intended to separate primary articles from internal ones, zero corresponds to the primary namespace;
+ <id> – unique article ID;
+ <revision> – group of elements of the relevant article version;
– <id> – primary revision key used to monitor article changes;
– <parented> – parent article ID;
– <timestamp> – date and time of article revision creation;
+ <contributor> – group of article authorship elements;
– <username> – article author name;
+ <id> – unique article author ID;
+ <text> – article text with elements of wiki-markup;
– <sha1> – article hash code produced by Secure Hash Algorithm SHA-1, used to monitor versions;
– <model> – model of article content, in this case wikitext;
– <format> – format of article data, in this case text/x-wiki.

For topic modelling objectives the information, which is contained in elements marked with (+), is necessary. Elements that contain information, which is not used in topic modelling algorithms are marked with (–). Example of a part of XML-tree of Wikinews data base export file is shown in fig. 3.

Preliminary processing of Wikinews data

In the export file of Wikinews the articles are sorted according to the revision creation date <timestamp>, this date is not related to the date of the described events. Authors are recommended to specify the date of the events in the article text, using wiki-markup. The example of wiki-markup of the date {{: Date | December 24, 2005}} is shown in fig. 4 inside the element <text>. Some articles in the export file of Wikinews does not contain the date of the events in the wiki-markup, but at the same time it is specified in the text or in the category. To save maximum information required for topic modelling algorithms, the date of the events was, where possible, restored from the text and categories. In 455 articles it was not possible to restore the date of the events, these articles are selections of news that occurred on the same day, in different years, and present no value for constructing topic models, and they were excluded from the corpus. Documents of the SCTM-ru corpus are sorted by the date of the events, from old to new ones.

The export file of the Wikinews database contains information about the author of the last article revision. We use such information as authorship ID for constructing author-topic models. Since 58 pieces of Wikinews do not contain the author data, and articles are valuable, the technical decision was made to assign a unique author ID – 2 – to these articles and include the latter into the SCTM-ru corpus.

The text of the Wikinews article contains references that are arranged in a certain way. References are divided into three groups: internal – a tool to relate pages inside the language section of Wikipedia, interlingual links (interwiki) – means to organize connections between various wiki-systems in Internet and references to pages of brotherly wiki-projects (for instance, to Wikipedia). The article text enclosed within double square brackets is an internal reference. If the case of the referring word or token does not match with the nominative case, there is a line within double square brackets, to the left of which there is the nominative case of the reference text, and to the right – the text that corresponds to the sentence grammar. Topic modelling algorithms consider the number of each word lemma entries into the text, in internal references each word has two entries in different wordforms and will be taken into account twice in the topic model, thus having perverted frequency characteristics of the model. Documents of the SCTM-ru corpus contain

```
<?xml version="1.0" encoding="utf-8"?>
<page>
  <title>Russian Wikipedia reaches a quarter million articles</title>
  <ns>0</ns><id>102340</id>
  <revision><id>625027</id><parentid>625026</parentid><timestamp>2008-04-26T10:47:46Z</timestamp>
    <contributor><username>Cirt</username><id>17538</id></contributor>
    <comment>{{archived}}</comment><model>wikitext</model><format>text/x-wiki</format>
    <text xml:space="preserve">{{WikimediaMention}}{{date|March 18, 2008}}
...
The [[w:Russian Wikipedia|Russian Language edition]] of [[Wikipedia]] reached the milestone
of 250,000 articles on Monday, March 17. The collective of editors on the project issued a
[[w:ru:Википедия:Пресс-релиз/250K|press release]] on that, according to which the jubilee
article was determined to be “[[w:ru:Гендерная лингвистика|Гендерная лингвистика]]” (Gender
linguistics) created by ''ru-wiki'' [[w:ru:User:Koryakov Yuri|user Koryakov Yuri]].
...
== Sources ==
...
[[Category:Russia]]
[[Category:Internet]]
[[Category:Wikipedia]]
[[Category:Wikimedia Foundation]]

[[es:La Wikipedia en ruso llega al cuarto de millón de artículos]]
[[ru:Русская Википедия набрала четверть миллиона статей]]</text>
    <sha1>88kc9g4e6yos4ju6508o4lhmt9z4iux</sha1>
  </revision>
</page>
```

Fig. 3. Example of XML-article „50,000 Articles in Russian Wikipedia“ on website of Russian Wikinews

```
<?xml version="1.0" encoding="utf-8"?>
<page>
  <title>Russian Wikipedia reaches a quarter million articles</title>
  <id>102340</id>
  <userid>17538</userid>
  <category>Wikimedia Foundation</category>
  <category>Wikipedia</category>
  <category>Russia</category>
  <category>Internet</category>
  <data>18 March 2008</data>
  <text>
    ...
    The Russian Wikipedia of Wikipedia reached the milestone of 250,000 articles on
    Monday, March 17. The collective of editors on the project issued a press release
    on that, according to which the jubilee article was determined to be “Гендерная
    лингвистика” (Gender linguistics) created by ''ru-wiki'' user Koryakov Yuri.
    ...
  </text>
</page>
```

Fig. 4. Example of XML-document „50,000 Articles in Russian Wikipedia“ in SCTM-ru corpus

only that part of the reference that corresponds to the sentence grammar.

News shall be accompanied with references to a documentary source. They are usually divided into four types: other articles of Wikinews, external links to online-sources, citations of printed media and websites with reference or related information. For the section of the article “Sources” they use wiki-markup == Sources == (see example in fig. 4). For purposes of topic modelling, links to sources are of no big value, therefore the decision was made to exclude them from the SCTM-ru corpus.

The important element of Wikinews markup and important data for constructing topic models is information on categories, to which the article is related. Categories of the article are defined by its author.

Software was developed in C# language, the development environment is Visual Studio Express 2013, for preliminary processing of texts. Regular expressions were used to search within the export file of Wikinews. The software is multi-modular, each module performs one certain operation. The software receives the initial XML-file as an input, specially prepared regular expressions sequentially search through the file looking for a match by the template, and an XML-file is created with changes made within one iteration at the output. To preserve integrity of initial data, each search through the initial XML-file makes only a few changes that are thoroughly checked by the system administrator, afterwards the software is started with another processing module.

Multi-modular software was developed to calculate statistics of the SCTM-ru corpus. The document count module analyses the XML-tree of the corpus, retrieves unique IDs for each document and counts their total. The author count module retrieves a list of unique IDs of Wikinews article authors and counts their total. The category count module retrieves unique categories from the XML-tree of the corpus and counts their total. The module for processing of the dates of the events described in articles analyses the XML-tree of the corpus, retrieves information on the date of the event of each document, counts unique values, finds the earliest and latest dates of the document.

To count vocabulary of the SCTM-ru corpus, a module was developed using regular expressions and MyStem software. The module takes the text from specified elements of the XML-tree (title, text), regular expressions from the text retrieve all sequences of Russian alphabet letters. In the process of word count the sequence of Russian alphabet letters separated from other letters with something other than letters (punctuation marks, blanks) is a word. MyStem software was used to determine the word lemmas. MyStem software performs morphological analysis of the text in Russian. Hypotheses are generated for the words that are absent in the dictionary [21].

SCTM-ru corpus markup

As SCTM-ru corpus storage format, XML (eXtensible Markup Language) was chosen, as one of most convenient formats for use in software environment and conversion of data into other formats. XML features make it possible to save the text of the initial Wikinews article and highlight additional parameters of the document.

XML-file of the corpus (SCTM-ru) consists of the following elements:

- <page> – group of document elements;
- <title> – document title;
- <id> – unique document ID;
- <userid> – unique author ID;
- <category> – document category;
- <date> – date of document events;
- <text> – document text;

Example of one document markup in SCTM-ru corpus is shown in fig. 4.

The document title (title) is separated from the document text, since title words may be given higher priority in construction of a topic model.

The unique article author ID (userid) is a parameter necessary in author-topic models. The Author-Topic over Time model [22] is an extension to LDA, where distribution of authors, topics and documents in time is evaluated in process of model construction.

Document categories (category) are categories specified by the article author. For instance, in fig. 4 in the article „50,000 Articles in Russian Wikipedia“ the „Russian Wikipedia“ category is specified. Information on the category is important for topic modelling, therefore saved in the SCTM-ru corpus, see fig. 4.

Availability of information on documents belonging to categories will make it possible to automatically check accuracy, completeness, exactness of tested topic modelling algorithms. Information on the document categories may be used in Labeled LDA models described in [23].

The date of the events described in the article (date) is used to build temporal topic models. Example of the model using the date under the title „Topic over Time – TOT" is shown in the paper [24]. When a temporal model is constructed, apart from standard distributions of words among topics and topics by documents, one shall assess distribution of each topic over time, which makes it possible to track and display dynamics of topics variation over time.

The document text (text) corresponds to the initial article text. We purposefully leave the initial text without any change, without its conversion into a model of a „bag of words", and without linguistic processing, to make it possible to study unique features of the Russian language. Information on the sequence of words in the document text is used in the models that consider mutual occurrence of words. For instance, the model titled Hidden Topic Markov's Model – HTMM described in the paper [25] is based on suggestions that words in the sentence structure, as well as sentences themselves are related to one common topic, and the topics of words in the document produce a Markov's chain. As a result of work the HTMM reduces ambiguity of words, widens topic understanding.

Conclusion

As a result of the work done, a special Russian corpus of texts was prepared (SCTM-ru), which is suitable to test various algorithms of probabilistic topic modelling. The objectives set for the work were achieved: SCTM-ru corpus contains original texts of documents in Russian, information on date of events described in the document, information about author and categories, to which the document is related, is available for download and use on devices with no access to Internet.

The source of corpus data is the international news website "Russian Wikinews". The SCTM-ru corpus contains 7 K documents, 185 authors, almost 12 K unique categories. Events described in the documents are distributed among more than 2 K unique dates, from November 2005 to June 2014. The SCTM-ru corpus contains 2.4 M tokens that consist of Russian letters only. The corpus vocabulary includes 150.6 K unique wordforms, 59 K unique lemmas.

The volume of the developed corpus gives ground to suggest its representativeness for various tasks of automatic processing of natural language texts. As noted in the paper [26], "It is not reasonable to wait until someone balances the corpus scientifically before using it, and it would not be prudent to assess results of corpus analysis as "not well-founded" or "irrelevant" just because one cannot prove that the used corpus is "balanced". Variety of events described in the SCTM-ru corpus and large team of article authors (21 K members) justify the suggestion on its balance. One may be convinced of corpus balance after analysis of internal features and construction of topic models.

The suggested technology of text corpus development for objectives of topic modelling makes it possible to expand the SCTM-ru corpus due to new articles. Similarly, language corpora may be established in any language from 33 languages presented in Wikinews. Within the proposed format collections and corpora may be created from various sources of data, at the same time only information required for topic modelling algorithms shall be kept.

Then on the basis of the developed corpus the features of existing variations of topic modelling algorithms will be studied, new algorithms will be developed, which take into account linguistic features of the Russian language. The SCTM-ru corpus is distributed through an open license and is available for download at <www.cims.ru>.

References

1. Papadimitriou Ch. H., Raghavan P., Hisao Tamaki, Vempala S. Latent semantic indexing: A probabilistic analysis. – 1998.

2. Hoffman Th. Probabilistic Latent Semantic Indexing. *Proc. 22 Annual Int. SIGIR Conf. Res. Dev. Inform. Retrieval*, 1999.

3. Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 2003.

4. Daud A., Li J., Zhou L., Muhammad F. Knowledge discovery through directed probabilistic topic models: a survey. *Proc. Front. Comput. Sci. Chin.*, 2010, pp. 280-301.

5. Russian National Corpus. Available at: www.ruscorpora.ru (accessed 12.01.2015). (In Russ.)

6. Zakharov V. P. International standards in corpora linguistics. *Struct. Appl. Ling. [Strukturnaya i prikldnaya lingvistika]*, 2012, no. 9, pp. 201-221. (In Russ.)

7. Granovsky D. V., Bocharov V. V., Bichineva S. V. Opencorpora: how it work and perspectives [Otkrytyy korpus: printsipy raboty i perspektivy]. *Computer linguistics and development of semantic search on Internet: Proc. 13th All-Russian integrated conf. "Internet and Modern Society" [Kompyuternaya lingvistika i razvitie semanticheskogo poiska v internete: Trudy nauchnogo seminara XIII vserossiyskoy obedinennoy konferencii "Internet i sovremennoe obschestvo]*. St. Petersburg, Oct. 19-22, 2010 / ed. V. S. Rubashkin. St. Petersburg, 2010. 94 p. (In Russ.)

8. OpenCorpora. Available at: opencorpora.org (accessed 15.01.2015). (In Russ.)

9. Small corpus of Associated Press. Available at: www.cs.princeton.edu/~blei/lda-c (accessed 6.01.2015).

10. The New York Times Annotated Corpus. Available at: catalog.ldc.upenn.edu/LDC2008T19 (accessed 14.01.2015).

11. The 20 Newsgroups data set. Available at: qwone.com/~jason/20Newsgroups (accessed 24.01.2015).

12. Reuters Corpora. Available at: trec.nist.gov/data/reuters/reuters.html (accessed 24.01.2015).

13. Reuters-21578 Text Categorization Collection Data Set. Available at: archive.ics.uci.edu/ml/datasets/Reuters21578-+Text+Categorization+Collection (accessed 24.01.2015).

14. Vinogradova V. B., Kukushkina O. V., Polikarpov A. A., Savchuk S. O. The computer corpus of Russian newspapers of the XX th century end: the creation, categorization, automated analysis of linguistic features. *Russian Language: its Historical Destiny and Present State: Int. Congress of Rus. Language Res. Philological Faculty of the Lomonosov Moscow State Univ. (MSU) [Russkiy yazyk: istoricheskie sudby i sovremennost: Mezhdunarodnyy congress rusistov issledovateley. Moskva, filologicheskiy f t MGU im. M. V. Lomonosova]* 13-16 March 2001. Moscow: Moscow State Univ. Press, 2001. P. 398. (In Russ.)

15. The computer corpus of Russian newspapers of the XX century end. Available at: www.philol.msu.ru/~lex/corpus/corp_descr.html (accessed 24.01.2015). (In Russ.)

16. Ventsov A. V., Grudeva E. V. About Corpus of Standard Written Russian (narusco.ru). *Rus. Ling.*, 2009, Vol. 33, no. 2, pp. 195-209. (In Russ.)

17. Corpus of Standard Written Russian. Available at: www.narusco.ru (accessed 24.01.2015). (In Russ.)

18. HANCO Corpus. Available at: www.helsinki.fi/venaja/russian/e-material/hanco/index.htm (accessed 24.01.2015).

19. Krizhanovsky A. A., Smirnov A. V. An approach to automated construction of a general-purpose lexical ontology based on Wiktionary.*J. Comput. Syst. Sci. Int.*, 2013, Vol. 52, no. 2, pp. 215-225.

20. Smirnov A. V., Kruglov V. M., Krizhanovsky A. A., Lugovaya N. B., Karpov A. A., Kipyatkova I. S. A quantitative analysis of the lexicon in Russian WordNet and Wiktionaries. *SPIIRAS Proc*. [*Trudy SPIIRAN*], 2012, Is. 23, pp. 231-253. (In Russ.)

21. System for automatic morphological analysis of Russian MyStem. Available at: api.yandex.ru/mystem (accessed 12.12.2014). (In Russ.)

22. Xu S., Shi Q., Qiao X., et al. Author-Topic over Time (AToT): a dynamic users' interest model, in Mobile, Ubiquitous, and Intelligent Computing. Berlin, Germany; Springer, 2014. Pp. 239-245.

23. Ramage D., Hall D., Nallapati R., Manning C. D. Labeled LDA. A supervised topic model for credit attribution in multi-labeled corpora. *Empirical Methods in Nat. Lang. Proc.*, 2009. Pp. 248-256.

24. Xuerui Wang, McCallum A. Topics over Time: A Non-Markov ContinuousTime Model of Topical Trends. *Proc. 12th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Philadelphia, USA, Aug. 20-23, 2006.

25. Gruber A., Rosen-Zvi M., Weiss Ya. Hidden Topic Markov Models. *Proc. Artificial Intel. Statistics (AISTATS)*, San Juan, Puerto Rico, USA, March 21-24, 2007.

26. Zakharov V. P., Azarova I. V. Special text corpora parametrization. *Structural and Applied Linguistics: Interuniv. collection* [*Strukturnaya i prikladnaya lingvistika: mezhvuzovskiy sbornik*]. Vol. 9. St. Petersburg; St. Petersburg State Univ., 2012. Pp. 176-184. (In Russ.)

Корпус текстов русского языка для тестирования алгоритмов тематического моделирования

Карпович С. Н.
АО «Олимп»
Москва, Россия
cims@yandex.ru

Аннотация. **Предложен специальный корпус текстов SCTM-ru для тестирования алгоритмов тематического моделирования. В условиях стремительного роста количества информационных данных остро проявляется проблема разработки инструментов и систем для их автоматической обработки. Для создания систем и тестирования алгоритмов должны существовать подходящие наборы данных. Необходимо наличие свободных коллекций документов, текстовых корпусов на русском языке для исследований методов автоматической обработки текстов на естественном языке с учетом лингвистических особенностей языка. Обозначены требования к специальному корпусу: он должен распространяться по свободной лицензии, количество документов должно быть достаточным для исследования, должен содержать тексты документов на естественном языке, а также востребованную в алгоритмах тематического моделирования информацию. Проведен сравнительный анализ корпусов на русском и иностранных языках, выявлено несоответствие характеристик существующих корпусов обозначенным требованиям.**

Ключевые слова: **текстовый корпус, тематическая модель, обработка естественного языка, русский язык.**

Литература

1. Papadimitriou Ch. H., Raghavan P., Hisao Tamaki, Vempala S. Latent semantic indexing: A probabilistic analysis. – 1998.

2. Hoffman Th. Probabilistic Latent Semantic Indexing // Proc. 22 Annual Int. SIGIR Conf. Res. Dev. Inform. Retrieval, 1999.

3. Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet Allocation // J. Mach. Learn. Res. 2003.

4. Daud A., Li J., Zhou L., Muhammad F. Knowledge discovery through directed probabilistic topic models: a survey // Proc. Front. Comput. Sci. Chin. 2010. P. 280-301.

5. Национальный корпус русского языка НКРЯ. URL: www.ruscorpora.ru (дата обращения 12.01.2015).

6. Захаров В. П. Международные стандарты в области корпусной лингвистики // Структурная и прикладная лингвистика. 2012. № 9. С. 201-221.

7. Грановский Д. В., Бочаров В. В., Бичинева С. В. Открытый корпус: принципы работы и перспективы // Компьютерная лингвистика и развитие семантического поиска в Интернете: тр. науч. семинара XIII Всерос. Объединен. конф. «Интернет и современное общество». Санкт-Петербург, 19-22 окт. 2010 г. /под ред. В. Ш. Рубашкина. – СПб., 2010. 94 с.

8. Открытый корпус. URL: opencorpora.org (дата обращения 10.01.2015).

9. Small corpus of Associated Press. URL: www.cs.princeton.edu/~blei/lda-c (дата обращения 6.01.2015).

10. The New York Times Annotated Corpus. URL: catalog.ldc.upenn.edu/LDC2008T19 (дата обращения 14.01.2015).

11. The 20 Newsgroups data set. URL: qwone.com/~jason/20Newsgroups (дата обращения 24.01.2015).

12. Reuters Corpora. URL: trec.nist.gov/data/reuters/reuters.html (дата обращения 24.01.2015).

13. Reuters-21578 Text Categorization Collection Data Set. URL: archive.ics.uci.edu/ml/datasets/Reuters-21578+Text+Categorization+Collection (дата обращения 24.01.2015).

14. Виноградова В. Б., Кукушкина О. В., Поликарпов А. А., Савчук С. О. Компьютерный корпус текстов русских газет конца 20-го века: создание, категоризация, автоматизированный анализ языковых особенностей // Русский язык: исторические судьбы и современность: Междунар. конгресс русистов-исследователей. Москва, филологический ф-т МГУ им. М. В. Ломоносова 13-16 марта 2001 г. Труды и материалы. – М.: Изд-во Москов. ун-та, 2001. С. 398.

15. Компьютерный корпус текстов русских газет конца XX века. URL: www.philol.msu.ru/~lex/corpus/corp_descr.html (дата обращения 24.01.2015)

16. Венцов А. В., Грудева Е. В. О корпусе русского литературного языка (narusco.ru) // Рус. лингвистика. 2009. Т. 33, № 2. С. 195-209.

17. Корпус русского литературного языка. URL: www.narusco.ru (дата обращения 24.01.2015).

18. Хельсинкский аннотированный корпус русских текстов ХАНКО. URL: www.helsinki.fi/venaja/russian/e-material/hanco/index.htm (дата обращения 24.01.2015).

19. Krizhanovsky A. A., Smirnov A. V. An approach to automated construction of a general-purpose lexical ontology based on Wiktionary // J. Comput. Syst. Sci. Int. 2013. Vol. 52, № 2. P. 215-225.

20. Смирнов А. В., Круглов В. М., Крижановский А. А., Луговая Н. Б., Карпов А. А., Кипяткова И. С. Количественный анализ лексики русского WordNet и викисловарей // Тр. СПИИРАН. 2012. Вып. 23. С. 231-253.

21. Программа морфологического анализа текстов на русском языке MyStem. URL: api.yandex.ru/mystem (дата обращения 12.12.2014).

22. Xu S., Shi Q., Qiao X. et al. Author-Topic over Time (AToT): a dynamic users' interest model, in Mobile, Ubiquitous,

and Intelligent Computing. – Berlin (Germany): Springer, 2014. P. 239-245.

23. Ramage D., Hall D., Nallapati R., Manning C. D. Labeled LDA. A supervised topic model for credit attribution in multi-labeled corpora // Empirical Methods Nat. Lang. Proc. 2009. P. 248-256.

24. Xuerui Wang, McCallum A. Topics over Time: A Non-Markov ContinuousTime Model of Topical Trends // Proc. 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Philadelphia, USA, Aug. 20-23, 2006.

25. Gruber A., Rosen-Zvi M., Weiss Ya. Hidden Topic Markov Models // Proc. Artificial Intel. Statistics (AISTATS), San Juan, Puerto Rico, USA, March 21-24, 2007.

26. Захаров В. П., Азарова И. В. Параметризация специальных корпусов текстов // Структурная и прикладная лингвистика: межвуз. сб. Вып. 9. – СПб.: СПбГУ, 2012. С. 176-184.

Моделирование процессов на основе последовательного гиперфрактального распределения

Смагин В. А.
Военно-космическая академия им. А. Ф. Можайского
Санкт-Петербург, Россия
va_smagin@mail.ru

Бубнов В. П.
Петербургский государственный университет
путей сообщения Императора Александра I
Санкт-Петербург, РФ
bubnov1950@yandex.ru

Аннотация. **Фракталы, теория фракталов применяются при описании различных явлений – от биологических до квантово-механических. Предложена математическая модель представления процессов в виде последовательного гиперфрактального распределения. Она базируется на модели квантования информации и гипердельтном распределении вероятностей, ранее предложенном автором. Для формирования последовательности предложено нелинейное интегральное уравнение с целочисленным ядром. По нему находятся базовый фрактал и субфракталы (кластеры). Рассмотрен пример для равномерного распределения. Оценены вероятностные и энтропийные свойства компонентов разложения. Рекомендовано использовать подход в метрологии, теории информации и теории эффективности.**

Ключевые слова: **последовательности фракталов, субфрактал, вероятностные свойства, энтропийные свойства, детерминированные и случайные процессы.**

Введение

В книге известного норвежского физика дается ясное и простое изложение математических свойств фракталов и описываются приложения теории фракталов в гидродинамике, океанологии, гидрологии, в исследовании перколяционных процессов и пр. [1]. Кроме того, приводятся методы компьютерной графики.

Указанный источник рекомендуется для аспирантов и студентов, желающих ознакомиться с теорией фракталов и применять ее при описании различных явлений – от биологических до квантовомеханических.

Фрактал (от лат. fractus – дроблёный, сломанный, разбитый) – это множество, обладающее свойством самоподобия (объект, в точности или приближённо совпадающий с частью самого себя, т. е. целое имеет ту же форму). Создание теории фракталов и методов её применения принадлежит известному учёному – автору большого цикла работ Б. Мандельброту [2].

Целью данной статьи является попытка предложить один из возможных математических методов моделирования детерминированных и случайных процессов, идея которого восходит как к работе автора [3], так и к идее применения фракталов, изложенной в [1, 2]. Она состоит в том, чтобы разработать приближённый метод исследования детерминированных и случайных процессов, с которыми тесно связаны прикладные методы современной науки технического и информационного профиля. В более простом понимании – связать науку о фракталах с теорией вероятностей и теорией информации. При этом автор не претендует на создание научной теории, а излагает своё понимание на простых прикладных примерах.

Суть метода

В статье [3], посвящённой разработке одной из моделей исследования немарковских процессов в теории надёжности и массовом обслуживании, была предложена Марковская модель гипердельтного распределения (см. также [4]). Она основана на методе равенства начальных моментов теоретического и аппроксимирующего распределений теории вероятностей. Здесь эту идею предлагается применить к построению приближённого последовательного распределения, построенного на выделении базового, основного фрактала и совокупности субфракталов более низкого ранга по сравнению с базовым фракталом. Затем на основе дельта-функций Дирака строится плотность вероятности фрактального распределения. С её помощью оцениваются вероятностно интересующие исследователя показатели объекта. Кроме того, при необходимости предлагается применять функцию распределения энтропии для дополнительного оценивания качества объекта (см. [5]).

Математическая модель

Основывается на идее оптимального квантования информации [6]. Точнее, она использует запись авторами величины математического ожидания квантованной случайной величины:

$$M(x) = (x + c)\int_0^\infty (E(\frac{z}{x}) + 1) f(z)dz,$$

где c – установленный постоянный пробел между квантами; x – величина кванта; E – наибольшая целая часть числа с недостатком; $f(x)$ – плотность вероятности случайного количества квантуемой информации Z.

Пусть задана плотность вероятности $f(x)$ на полуинтервале $[0, \infty)$, требуется представить её в виде убывающей фрактальной последовательности, составленной базовым, основным фракталом Φ_0 и множеством субфракталов Φ_i, $i = 1, 2, 3, \ldots$ Тогда можно принять следующую формулу для производства фрактализации:

$$\int_0^\infty E(\frac{z}{x}) f(z)dz - 1 = 0. \qquad (1)$$

Содержательный смысл (1) состоит в том, что математическое ожидание E не должно превышать единичного значения, но и не быть отрицательным. Нужно принять равенство $x = \Phi_0$ и решить полученное нелинейное уравнение относительно неизвестной величины Φ_0. При этом нужно как можно точнее вычислить эту величину, а затем проверить правильность достаточно строгого решения уравнения (1). Эта численная величина и будет представлять базовый, основной факториал. Можно убедиться, что точное решение уравнения (1) достигается при $\max\Phi_0$ и оно будет единственным.

Следующим шагом будет вычисление значения величины первого субфрактала Φ_1. Этот процесс будет аналогичен предыдущему процессу – вычисления Φ_0. Но нелинейное уравнение необходимо изменить таким образом, чтобы оно приняло форму

$$\int_{\Phi_0}^{\infty} E\left(\frac{z-\Phi_0}{\Phi_1}\right)\frac{f(z)}{\int_{\Phi_0}^{\infty} f(u)du}dz - 1 = 0. \qquad (2)$$

При известном определённом значении Φ_0 уравнение (2) нужно решить относительно Φ_1 с достаточной строгостью и проверить точность полученного решения. Сомножитель с внутренним интегралом в (2) определяет условную вероятность того, что предшествующий интервал базового фрактала был успешно завершён.

Следующий шаг – вычисление значения второго субфрактала Φ_2. Для этого необходимо использовать нелинейное интегральное уравнение

$$\int_{\Phi_0+\Phi_1}^{\infty} E\left(\frac{z-\Phi_0-\Phi_1}{\hat{O}_2}\right)\frac{f(z)}{\int_{\Phi_0+\Phi_1}^{\infty} f(u)du}dz - 1 = 0. \qquad (3)$$

Сомножитель с внутренним интегралом в (3) определяет условную вероятность того, что предшествующий интервал до Φ_2, включающий базовый фрактал и первый субфрактал, был успешно завершён.

После процедуры следует также обеспечить и проверить точность полученного решения. Далее следует вычислить следующие субфракталы по аналогичным, но видоизменённым уравнениям до тех пор, пока величина последнего фрактала не будет пренебрежимо малой.

Автор статьи не нашел метода решения рассматриваемого нелинейного уравнения в замкнутом аналитическом виде, поэтому применил способ непосредственной замены искомой величины таким образом, чтобы высокая точность решения строго соответствовала максимальному значению искомого фрактала или субфрактала.

Пример вычислений для заданного распределения вероятностей

Задано равномерное распределение, его плотность вероятности записывается в виде

$$f(x) = dunif(x, a, b); \quad a = 0; \quad b = 100 \text{ ед.}$$

Для определения значения величины базового фрактала требуется решить уравнение

$$\int_{0}^{\infty} E\left(\frac{z}{\Phi_0}\right) f(z)dz - 1 = 0.$$

В результате решения путём простого подбора определяем $\Phi_0 = 33{,}33$ ед. с точностью $1{,}819\times10^{-4}$. Число фракталов равно единице, а интервал временной занятости составит 33,33 ед.

Определяем значение первого субфрактала путём решения уравнения относительно Φ_1:

$$\int_{\Phi_0}^{\infty} E\left(\frac{z-\Phi_0}{\Phi_1}\right)\frac{f(z)}{\int_{\Phi_0}^{\infty} f(u)du}dz - 1 = 0.$$

Значение

$$\int_{\Phi_0}^{\infty} f(u)du = 0{,}667.$$

Величина Φ_1 будет 22,21 ед. с точностью $7{,}2\times10^{-4}$. Число субфракталов Φ_1 равно единице, поэтому интервал временной занятости составит 22,21 ед.

Определяем значение второго субфрактала путём решения уравнения относительно Φ_2:

$$\int_{\Phi_0+\Phi_1}^{\infty} E\left(\frac{z-\Phi_0-\Phi_1)}{\Phi_2}\right)\frac{f(z)}{\int_{\Phi_0+\Phi_1}^{\infty} f(u)du}dz - 1 = 0.$$

Значение

$$\int_{\Phi_0+\Phi_1}^{\infty} f(u)du = 0{,}445.$$

Величина Φ_2 будет 8,63 ед. с точностью $5{,}185\times10^{-7}$. Число Φ_2 субфракталов равно единице, поэтому временной интервал занятости составит 8,63 ед.

Определяем значение третьего субфрактала путём решения уравнения относительно Φ_3:

$$\int_{\Phi_0+\Phi_1+\Phi_2}^{\infty} E\left(\frac{z-\Phi_0-\Phi_1-\Phi_2)}{\Phi_3}\right)\frac{f(z)}{\int_{\Phi_0+\Phi_1+\Phi_2}^{\infty} f(u)du}dz - 1 = 0.$$

Значение

$$\int_{\Phi_0+\Phi_1}^{\infty} f(u)du = 0{,}338.$$

Величина Φ_3 будет 11,95 ед. с точностью $4{,}889\times10^{-5}$. Число Φ_3 субфракталов равно единице, поэтому временной интервал занятости составит 11,95 ед.

На этом ограничимся, считая, что величина четвёртого субфрактала будет близка к нулю.

Анализ результатов

В примере получены следующие значения продолжительности времени занятости фрактала и субфракталов:

$t_0 = 33{,}33$ ед.; $t_1 = 22{,}21$ ед.; $t_2 = 8{,}63$ ед.; $t_3 = 0$ ед.

Обозначим соответствующие вероятности нахождения процесса в периодах времени занятости:

$$P_0(t_0) = 0{,}667;\ P_1(t_1) = 0{,}519;\ P_2(t_2) = 0{,}316;\ P_3(t_3) = 0.$$

На рис. 1 показано изменение вероятности процесса занятости по времени. Каждый отрезок прямой на рис. 1 соответствует вычисленной вероятности на конец отрезка. Так что если мы хотим сгладить контур кривой непрерывной линией вероятности, то должны соединить плавной линией точку 1 на оси ординат с правыми концами всех отрезков, исключая самый нижний отрезок.

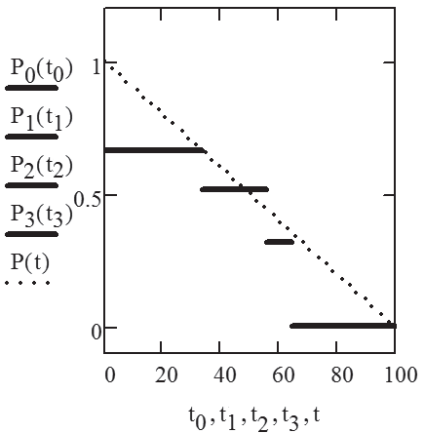

Рис. 1. Вероятности процесса занятости

При достаточном количестве отрезков эта линия должна плавно переходить в нулевую линию – линию абсцисс. На рис. 1 при учёте только двух субфракталов невозможно плавно соединить точки с осью абсцисс, кривую приходится обрывать, не достигая оси.

На рис. 1 прямой пунктирной линией показан график вероятности того, что событие, определяемое заданным равномерным распределением в примере с плотностью вероятности

$$f(t) = dunif(x, a, b),\ a = 0,\ b = 100 \text{ ед.}$$

и определяемое формулой

$$P(t) = 1 - \int_0^t f(z)dz,$$

не может произойти. Эта линия хорошо аппроксимируется концами отрезков вероятностей базового фрактала и второго субфрактала. Это подтверждает корректность фрактальной аппроксимации заданного распределения вероятностей на основе последовательного гиперфрактального распределения.

Оценка эффективности фрактального прогноза по ресурсу Н. М. Седякина

Здесь величина ресурса – это критерий важности, или веса фрактала. Ресурс за время t определяется по формуле

$$r(t) = \int_0^t \lambda(z)dz,\ \lambda(t) = \frac{f(t)}{P(t)}.$$

Временные интервалы занятости $t_0 = 0$; $t_1 = 33{,}33$; $t_2 = 55{,}5$; $t_3 = 64{,}2$. Остаточный временной интервал $t > 64{,}2$. Фрактальные (прогнозируемые) ресурсы

$$r_0(t_0) = 0;\ r_1(t_1) = 0{,}334;\ r_2(t_2) = 0{,}555;\ r_3(t_3) = 0{,}642.$$

Суммарный фрактальный ресурс 1,531. Остаточный ресурс $r(t > 64{,}2) = 0{,}358$. Полный ресурс 1,889. Доля фрактальных ресурсов 1,532 / 1,889 = 81,048 %. Доля остаточного ресурса 0,358 / 1,889 = 18,952 %. Таким образом, определение базового фрактала и двух субфрактлов для рассмотренного распределения вероятности оценивается 81,048 % – уровнем эффективности. Это, на наш взгляд, хорошая оценка.

Оценим важность фракталов примера по критерию плотности вероятности случайной величины энтропии [5]. Для этого по вычисленным вероятностям для фракталов найдём средние значения энтропии H, её второй начальный момент α и среднеквадратическое отклонение σ. Значения указанных величин приведены в таблице.

Характеристики фракталов

P	*H*	α	σ
0,667	0,636	0,512	0,328
0,519	0,692	0,481	0,046
0,316	0,624	0,518	0,484
0	0	0,008	0,092

Выполним аппроксимацию плотности энтропии нормальным распределением

$$g(x) = \frac{C}{\sqrt{2\pi}\sigma} e^{\frac{(x-H)^2}{2\sigma^2}},$$

где C – константа нормирования плотности. На рис. 2 приведены графики трёх плотностей вероятностей для трёх первых строк таблицы.

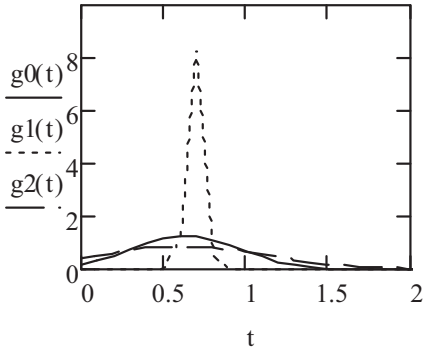

Рис. 2. Плотность вероятностей строк таблицы

Из рис. 2 следует, что первый субфрактал примера имеет самую большую значимость, или вес. Действительно, грубое оценивание значений концентрации величины энтропии (информации) фракталов I по данным таблицы и рис. 2 приводит к следующим результатам: фрактал Ф (первая строка таблицы и нулевой график плотности рис. 2 – I = 5,21; первый субфрактал Φ_1 (вторая строка таблицы и первый график плот-

ности рис. 2 – I = 9,031; второй субфрактал $Ф_2$ (третья строка таблицы и второй график плотности рис. 2 – I = 4,502.

О прикладной направленности исследования

Математическая модель, предложенная в статье, может найти применение в различных прикладных научных исследованиях. Например, гиперфрактальное распределение позволяет представлять современные системы счисления в виде совокупности фракталов, вводить новые системы счисления для моделирования процессов и применять их в изучении проблем теории информации. Фактически оно приложимо к моделированию процессов любой физической природы – материальной или информационной.

Оно позволяет изучать случайные процессы с целью увеличения точности их познания. Является достаточно привлекательным для метрологии [7], теории эффективности [8].

Заключение

Фракталы, теория фракталов в настоящее время применяются при описании различных явлений – от биологических до квантовомеханических [9–14]. На основании знакомства с отдельными трудами этой теории в статье предложена математическая модель предоставления процессов на основе последовательного гиперфрактального распределения. Она опирается на элементы теории квантования информации и на гипердельтное распределение вероятностей, ранее предложенное автором. Для моделирования последовательности фракталов распределений вероятностей предложено нелинейное интегральное уравнение. Ядро этого уравнения представлено в целочисленном виде. В результате последовательного решения уравнения находятся базовый фрактал и производные от него субфракталы (кластеры). Величину и содержание каждого фрактала можно изучать для оценивания его важности, информативности и т. д.

Модель может применяться в метрологии, теории информации, теории эффективности и в решении конкретных прикладных задач детерминированной или случайной направленности. Также на её основе можно получить новые результаты в теории фракталов.

Литература

1. Feder J. Fractals. – NY: Springer, 1988. 254 p.

2. Mandelbrot B. B. Les Objets Fractals: Forme, Hasard et Dimension. – Paris: Flammarion, 1975. 190 p.

3. Смагин В. А., Филимонихин Г. В. Моделирование случайных процессов на основе гипердельтного распределения // АВТ. 1990. № 5. С. 25-31.

4. Смагин В. А. Коррекция гипердельтного распределения в теории случайных процессов // Интеллектуальные технологии на транспорте. 2015. № 2. С. 27-31.

5. Смагин В. А. Техническая синергетика. Вероятностные модели сложных систем. – СПб., 2004. 171 с.

6. Андронов А. М., Бокоев Т. Н. Оптимальное в смысле заполнение квантование информации // Изв. АН СССР. Техническая кибернетика. 1979. № 3. С. 154-158.

7. Дорохов А. Н. Метрологическое обеспечение эксплуатации вооружения и военной техники: учеб. / под ред. А. Н. Миронова. – СПб., 2009. 755 с.

8. Петухов Г. Б. Основы теории эффективности целенаправленных процессов. Ч. 1. Методология, методы, модели. – СПб., 1989. 660 с.

9. Захаров А. И., Загайнов А. И. Реализация программного комплекса для вычисления фрактальных параметров сложных систем // Интеллектуальные технологии на транспорте. 2015. № 2. С. 47-53.

10. Falconer K. Fractal geometry. – UK: Univ. St. Andrews, 2003. 335 p.

11. Потапов А. А. Фракталы и дробные операторы в обработке информации фундаментальное направление синергетики // Изв. ЮФУ. Технические науки. 2011. № 6. С. 30-40.

12. Li H. Fractal analysis of side channels for breakdown structures in XLPE cable insulation // J. Mater. Sci.: Mater. Electron. Springer Sci. 2013. № 24. P. 1640-1643.

13. Martínez C. A. T., Fuentes C. Chapter 1. Applications of Radial Basis Function Schemes to Fractional Partial Differential Equations // Mathematics Fractal Analysis – Applications in Physics, Engineering and Technology / ed. F. Brambila. 2017.

14. Agboola O., Onyango M. S., Popoola P., Oyewo O. A. Chapter 10. Fractal Geometry and Porosity // Mathematics Fractal Analysis – Applications in Physics, Engineering and Technology / ed. F. Brambila. 2017.

Modelling of processes on the basis of consecutive giperfractal distributions

Smagin V.A.
A.F. Mozhaysky Military Aerospace Academy
St. Petersburg, Russia
va_smagin@mail.ru

Bubnov V. P.
Emperor Alexander I St. Petersburg
State Transport University
St. Petersburg, Russia
bubnov1950@yandex.ru

Abstract. **Fractals, the theory of fractals are applied at the description of various phenomena, from biological to quantum mechanical. The mathematical model of representation of processes in the form of consecutive hyper fractal distribution is offered. It is based on model of quantization of information and the giperdeltny distribution of probabilities which is earlier offered by the author. For formation of the sequence the nonlinear integrated equation with an integer kernel is offered. On him there are a basic fractal and subfractals (clusters). An example for uniform distribution is reviewed. Estimation of probabilistic and entropy properties of components of decomposition is made. Use in metrology is recommended, to the theory of information and the theory of efficiency.**

Keywords: **the sequences of fractals, subfraktal, the probabilistic properties, entropy properties determined and casual processes.**

References

1. Feder J. Fractals. NY, Springer, 1988. 254 p.

2. Mandelbrot B.B. Les Objets Fractals: Forme, Hasard et Dimension. Paris, Flammarion, 1975. 190 p.

3. Smagin V.A., Philimonikhin G.V. Modeling of casual processes on the basis of giperdeltny distribution [Modelirovanie sluchaynykh protsessov na osnove giperdel'tnogo raspredeleniya]. *AVT*, 1990, no. 5, pp. 25-31. (In Rus.)

4. Smagin V.A. Correction of the Hyperdelta Distribution in the Theory of Stochastic Processes [Korrektsiya giperdel'tnogo raspredeleniya v teorii sluchaynykh protsessov]. *Intellectual Technologies on Transport*, 2015, no. 2, pp. 27-31. (In Rus.)

5. Smagin V.A. Technical synergetrics. Probabilistic models of difficult systems [Tekhnicheskaya sinergetika. Veroyatnostnye modeli slozhnykh sistem]. St. Petersburg, 2004. 171 p. (In Rus.)

6. Andronov A.M., Bokoyev T.N. Optimum filling in sense quantization of information [Optimal'noe v smysle zapolnenie kvantovanie informatsii]. *News Academy of Sciences of the USSR. Technical cybernetics*, 1979, no. 3, pp. 154-158. (In Rus.)

7. Dorokhov A.N. Metrological support of operation of arms and military equipment [Metrologicheskoe obespechenie ekspluatatsii vooruzheniya i voennoy tekhnik]: The textbook; ed. A.N. Mironov. St. Petersburg, 2009. 755 p. (In Rus.)

8. Petukhov G.B. Bases of the theory of efficiency of purposeful processes. P. 1. Methodology, methods, models [Osnovy teorii effektivnosti tselenapravlennykh protsessov. Ch.1. Metodologiya, metody, modeli]. St. Petersburg, 1989. 660 p. (In Rus.)

9. Zakharov A.I., Zagaynov A.I. Realization of a program complex for calculation of fractal parameters of difficult systems [Realizatsiya programmnogo kompleksa dlya vychisleniya fraktal'nykh parametrov slozhnykh sistem]. *Intellectual Technologies on Transport*, 2015, no. 2, pp. 47-53. (In Rus.)

10. Falconer K. Fractal geometry. UK: Univ. St. Andrews, 2003. 335 p.

11. Potapov A.A. Fractals and fractional operators in information processing the fundamental direction of synergetrics [Fraktaly i drobnye operatory v obrabotke informatsii fundamental'noe napravlenie sinergetiki]. *News of SFU. Tech. sci.*, 2011, no. 6, pp. 30-40. (In Rus.)

12. Li H. Fractal analysis of side channels for breakdown structures in XLPE cable insulation. *J. Mater. Sci.: Mater. Electron. Springer Sci.*, 2013, no. 24, pp. 1640-1643.

13. Martínez C.A.T., Fuentes C. Chapter 1. Applications of Radial Basis Function Schemes to Fractional Partial Differential Equations. *Mathematics Fractal Analysis – Applications in Physics, Engineering and Technology*; ed. F. Brambila. 2017.

14. Agboola O., Onyango M.S., Popoola P., Oyewo O.A. Chapter 10. Fractal Geometry and Porosity. *Mathematics Fractal Analysis – Applications in Physics, Engineering and Technology*; ed. F. Brambila. 2017.

Способы псевдовероятностного блочного шифрования

Молдовян А. А.
Санкт-Петербургский институт информатики
и автоматизации РАН
Санкт-Петербург, Россия
maa1305@yandex.ru

Татчина Я. А.
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
Санкт-Петербург, Россия
iana.tatchina@gmail.com

Аннотация. **Псевдовероятностное шифрование представлено как новый алгоритмический механизм обеспечения информационной безопасности, реализующий защиту информации в случае атак с принуждением к раскрытию ключа шифрования. Базовым требованием к преобразованиям данного вида является вычислительная неразличимость по шифртексту от вероятностного шифрования. Рассматриваются способы и алгоритмы, выполняющие псевдовероятностное шифрование как одновременное криптографическое преобразование фиктивного и секретного сообщений по двум различным ключам, состоящее в формировании блоков промежуточных шифртекстов и их обратимом отображении в единый расширенный блок выходной криптограммы. Предложены алгоритмы, включающие задание процедуры объединяющего отображения в виде решения систем линейных уравнений и сравнений, в которых в качестве модуля используются числа и двоичные многочлены. Предложенные способы обладают высокой производительностью и представляют значительный интерес для практического применения в системах информационной безопасности.**

Ключевые слова: **криптография, отрицаемое шифрование, псевдовероятностное шифрование, симметричное шифрование, блочные шифры, вероятностное шифрование, криптограмма.**

Введение

Отрицаемое шифрование (ОШ) позволяет обеспечить защиту информации в ситуациях, когда участники сеанса защищенной связи подвергаются атаке с принуждением [1–3]. Такие атаки подразумевают, что у злоумышленника имеются некоторые ресурсы воздействия на получателя и/или отправителя, вынуждающие получателя и/или отправителя раскрыть секретные параметры процедуры зашифровывания или расшифровывания, например, исходное сообщение и секретный ключ.

Практическое значение ОШ определяется тем, что его применение позволяет решить ряд важных нестандартных задач, связанных с информационной безопасностью информационно-телекоммуникационных технологий. Этот вид шифрования применяется для защиты информации в протоколах распределенных вычислений [4], защиты от скупки голосов систем тайного голосования через Интернет [5, 6].

Специальным вариантом ОШ является псевдовероятностное (ПВ) шифрование, которое расширяет класс алгоритмических средств защиты информации, используемых в составе комплексных средств обеспечения информационной безопасности [7]. Стойкость ПВ-шифрования к принуждающим атакам обеспечивается тем, что криптограмма (шифртекст) формируется путем совместного криптографического преобразования двух сообщений – секретного и фиктивного – при выполнении требования вычислительной неразличимости от шифртекста, полученного в процессе вероятностного шифрования фиктивного сообщения по фиктивному ключу. Последние могут быть раскрыты атакующему в случае принуждающей атаки с сохранением требуемого уровня защищенности секретного сообщения. При этом пользователи имеют возможность убедительно утверждать, что они использовали алгоритм вероятностного шифрования для защиты информации. Ранее в рамках данного общего подхода были разработаны протоколы ПВ-шифрования, основанные на коммутативных шифрующих преобразованиях [8], и гибридные ПВ-шифры [9].

В схеме ПВ криптографического преобразования получатель и отправитель сообщения обмениваются двумя ключами шифрования – секретным и фиктивным, по которым выполняется процедура преобразования секретного и фиктивного сообщений, соответственно. При этом формируется единый шифртекст, из которого с использованием одного и того же алгоритма расшифровывания может быть восстановлено секретное или фиктивное сообщение в зависимости от используемого ключа. Такая возможность обеспечивается тем, что размер единого шифртекста больше, чем размер каждого из рассматриваемых двух входных сообщений. В случае вероятностного шифрования имеет место аналогичное увеличение размера шифртекста, что создает принципиальную возможность выполнить требование вычислительной неразличимости по шифртексту ПВ-шифрования от вероятностного [10]. Доказательное выполнение этого требования связано с представлением атакующему алгоритма вероятностного шифрования, которому соответствует такой алгоритм расшифровывания, с помощью которого по фиктивному ключу из шифртекста восстанавливается фиктивное сообщение.

Интерес к способам и алгоритмам ПВ-шифрования обусловливается возможностью реализации на их основе новых механизмов защиты информации, реализуемых в виде криптографически скрытых каналов (криптографических стегоканалов) и криптографических обманных ловушек [11, 12]. Данные защитные механизмы нового типа представляют интерес для практического применения в составе комплексных средств обеспечения информационной безопасности информационно-телекоммуникационных систем, используемых на железнодорожном транспорте.

Настоящая статья построена следующим образом. В первой части описываются основные требования к построению алгоритмов псевдовероятностного шифрования и модель потенциального нарушителя. Во второй части представлены способы блочного ПВ-шифрования с вычислением

блоков шифртекста как двоичных чисел путем решения систем сравнений. В третьей части описываются алгоритмы ПВ-шифрования с формированием блоков выходного шифртекста в виде решения систем линейных уравнений в конечном поле (простом и двоичном).

Требования к алгоритмам блочного псевдовероятностного шифрования

Алгоритмы блочного ПВ-шифрования как частные случаи алгоритмов ОШ с разделяемым ключом ориентированы на обеспечение стойкости к атакам с принуждением к раскрытию ключа шифрования. Принудительная атака представляет собой некоторое обобщение разных потенциальных атак, в ходе которых атакующий получает ключ шифрования. Наиболее характерным для модели принудительной атаки является то, что атакующий после перехвата шифртекста получает значение ключа шифрования, с помощью которого может быть восстановлено исходное сообщение. Принимая такую модель, можно абстрагироваться от конкретных способов действий атакующего, с помощью которых ему становится известным ключ. Такими действиями могут быть установка технических и программных закладок, подкуп, хищение ключевых носителей, перехват ключевой информации по побочным каналам, криптоанализ и др.

При разработке механизмов защиты информации, основанных на обманных криптографических, могут создаваться условия облегченного получения ключа шифрования, например, использование более короткого фиктивного ключа по сравнению с секретным ключом. Для практического использования защитных механизмов, основанных на ПВ-шифровании, следует разработать алгоритмы, обеспечивающие достаточно высокую скорость ПВ-шифрования и отсутствие признаков, по которым нарушитель мог бы определить, что шифрование не является вероятностным.

Для обеспечения последнего положения представляется важным выполнение следующих требований к построению ПВ-шифров, в том числе блочных алгоритмов ПВ-шифрования:

• криптограмма, полученная в процессе ПВ-шифрования, должна быть неотличима от криптограммы, получаемой в ходе процедуры вероятностного шифрования;

• восстановление из криптограммы фиктивного и секретного сообщений должно происходить независимо друг от друга;

• алгоритмы расшифровывания фиктивного и секретного сообщений должны совпадать.

Псевдовероятностное шифрование с вычислением блоков шифртекста из системы сравнений

Рассмотрим способ блочного ПВ-шифрования, реализуемый в виде совместного шифрования двух независимых сообщений – фиктивного M и секретного T, выполняемого по двум независимым ключам в единый шифртекст. Для обеспечения возможности независимого восстановления из последнего каждого из входных сообщений с использованием одного и того же алгоритма зададим разбиение M и T на блоки данных одинаковой длины, равной n бит, и формирование n-битовых блоков промежуточных шифртекстов с использованием некоторого апробированного алгоритма блочного шифрования E.

Полный секретный ключ ПВ-шифрования сформируем в виде секретного (K_1, p_1) и фиктивного (K_2, p_2) ключей, где элементы K_1 и K_2 – это ключи блочного n-битового шифра E, а p_1 и p_2 – пара взаимно простых чисел длины $n + 1$ бит. Совместное криптографическое преобразование сообщений $T = (T_1, T_2, \ldots, T_z)$ и $M = (M_1, M_2, \ldots, M_z)$, представленных в виде последовательностей n-битовых блоков данных, определим как последовательное преобразование соответствующих пар входных блоков T_i и M_i в единый расширенный блок шифр текста C_i (для $i = 1, 2, \ldots, z$) по следующему алгоритму:

1) используя функцию блочного шифрования E и ключ K_1, преобразовать входной блок данных T_i в блок промежуточного шифртекста $C_{Ti} = E_{K1}(T_i)$;

2) используя блочный шифр E и ключ K_1, преобразовать входной блок данных M_i в блок промежуточного шифртекста $C_{Mi} = E_{K2}(M_i)$;

3) используя ключевые элементы p_1 и p_2 и трактуя n-битовые блоки промежуточных шифртекстов C_{Ti} и C_{Mi} как числа, представленные в двоичном виде, вычислить единый блок выходного шифртекста C_i в виде решения системы сравнений

$$\begin{cases} C_i \equiv C_{T_i} \bmod p_1; \\ C_i \equiv C_{M_i} \bmod p_2. \end{cases} \qquad (1)$$

Для записи значения блока единого шифртекста C_i резервируется $2n + 2$ битов, что является достаточным, поскольку при любых значениях p_1 и p_2 имеет место условие $C_i < 2^{2n+2}$. Криптограмма C, содержащая в преобразованном виде оба входных сообщения, представляется в виде последовательности блоков шифртекста C_i: $C = (C_1, C_2, \ldots, C_z)$.

В соответствии с китайской теоремой об остатках решение системы линейных сравнений (1) описывается формулой

$$C_i = [C_{Ti} p_2 (p_2^{-1} \bmod p_1) + C_{Mi} p_1 (p_1^{-1} \bmod p_2)] \bmod p_1 p_2,$$

которая и задает вычислительную процедуру отображения пары блоков промежуточных шифртекстов в единый расширенный блок выходного шифртекста. Легко заметить, что значения $p_2 (p_2^{-1} \bmod p_1)$ и $p_1 (p_1^{-1} \bmod p_2)$ могут быть вычислены предварительно, что позволит повысить производительность описанного алгоритма блочного ПВ-шифрования.

Стойкость данного ПВ-шифра к принуждающим атакам обеспечивается возможностью правдоподобно утверждать, что шифртекст сформирован путем вероятностного шифрования фиктивного сообщения M по фиктивному ключу (K_2, p_2). Доводом в пользу такой интерпретации является представляемый ассоциированный алгоритм вероятностного шифрования и связанный с ним алгоритм процедуры расшифровывания шифртекста.

Ассоциируемый алгоритм вероятностного шифрования можно описать следующим образом:

1) разбить сообщение M на n-битовые блоки M_i:

$$M = (M_1, M_2, \ldots, M_z);$$

2) зашифровать каждый i-й ($i = 1, 2, \ldots, z$) блок входных данных M_i, выполняя следующие шаги:

2.1) используя n-битовый алгоритм блочного шифрования E, зашифровать блок данных M_i по ключу K_2

$$C_{Mi} = E_{K2}(M_i);$$

2.2) сгенерировать случайное число r, удовлетворяющее условию $2^n < r < 2^{n+1}$ и взаимно простое с p_2;

2.3) сгенерировать случайное число $R < 2^n$;

2.3) вычислить i-й блок криптограммы C_i как решение системы линейных сравнений

$$\begin{cases} C_i \equiv C_{M_i} \bmod p_2; \\ C_i \equiv R \bmod r. \end{cases} \quad (2)$$

Легко показать, что каждый блок C_i криптограммы C мог быть получен в результате вероятностного шифрования блока фиктивного сообщения M_i по фиктивному ключу при использовании представленного алгоритма вероятностного блочного шифрования. Действительно, пусть C_i и $C_{Mi} = E_{K2}(M_i)$ удовлетворяют первому сравнению системы (2). Тогда для любого значения r, которое является взаимно простым с p_2 и удовлетворяет условию $C_i < rp_2$, имеется значение $R = C_i \bmod r$, при котором система (2) имеет своим решением заданное значение C_i. Таким образом, зашифровывание фиктивного сообщения M по фиктивному ключу с использованием ассоциированного алгоритма вероятностного шифрования потенциально может привести к формированию шифртекста, полученного с помощью ПВ-шифрования сообщений M и T.

Алгоритму ПВ-шифрования и ассоциируемому с ним алгоритму вероятностного шифрования соответствует один и тот же алгоритм расшифровывания криптограммы $C = (C_1, C_2, \ldots, C_i, \ldots, C_z)$ по ключу (K_2, p_2), который описывается следующими шагами:

1) каждый i-й ($i = 1, 2, \ldots, z$) блок C_i расшифровать, выполнив следующие два шага:

1.1) вычислить блок промежуточного шифртекста $C_{Mi} = C_i \bmod p_2$;

1.2) расшифровать блок C_{Mi} по ключу K_2, используя функцию блочного расшифровывания $D = E^{-1}$

$$M_i = D_{K2}(C_{Mi});$$

2) объединить все восстановленные блоки M_i в сообщение $M = (M_1, M_2, \ldots, M_i, \ldots, M_z)$.

В случае принуждающей атаки отправитель и получатель криптограммы C предоставляют атакующему ключ (K_2, p_2) и сообщение M в качестве секретных значений. При этом они поясняют, что блочный шифр E использован в режиме вероятностного шифрования (и предоставляют атакующему ассоциированный алгоритм вероятностного шифрования). Применение режима вероятностного шифрования они поясняют желанием повышения стойкости шифрования и защиты от возможных непредвиденных слабостей блочного шифра E.

Секретное сообщение из криптограммы $C = (C_1, C_2, .., C_z)$ восстанавливается также путем выполнения последнего алгоритма, но при использовании секретного ключа (K_1, p_1), что определяет такую последовательность шагов:

1) каждый i-й ($i = 1, 2, \ldots, z$) блок C_i расшифровать, выполнив следующие два шага:

1.1) вычислить блок промежуточного шифртекста $C_{Ti} = C_i \bmod p_1$;

1.2) расшифровать блок C_{Ti} по ключу K_1, используя функцию блочного расшифровывания $D = E^{-1}$

$$T_i = D_{K1}(C_{Ti});$$

2) объединить все восстановленные блоки T_i в сообщение $T = (T_1, T_2, \ldots, T_i, \ldots, T_z)$.

Таким образом, рассмотренный блочный ПВ-шифр удовлетворяет принятым критериям построения, т. е. в нем совпадают алгоритмы восстановления из криптограммы C фиктивного и секретного сообщения. Также они совпадают с алгоритмом расшифровывания криптограммы, полученной с помощью ассоциированного алгоритма вероятностного шифрования.

Псевдовероятностные шифры с вычислением блоков шифртекста из системы уравнений

Рассмотрим построение скоростных алгоритмов псевдовероятностного шифрования с отображением пар промежуточных шифртекстов, реализуемое в виде процедуры решения системы линейных уравнений в простом конечном поле. Так же, как в предыдущем шифре, предполагается предварительное зашифровывание блоков входных сообщений M и T с помощью блочного шифра E на ключах $K = (K_1, K_2)$ и $Q = (Q_1, Q_2)$, соответственно, подключи которых используются также в качестве коэффициентов уравнений, входящих в систему. При этом при генерации подключей K_1, K_2, Q_1 и Q_2 следует обеспечить выполнимость условия $K_1Q_2 - K_2Q_1 \neq 0 \bmod p$ (условие существования единственного решения для системы двух линейных уравнений).

Пусть, например, промежуточное шифрование выполняется с помощью блочного шифра E с размером входного блока $n = 64$ бит и 128-битовым ключом K, представленным в виде пары 64-битовых подключей K_1 и K_2: $K = (K_1, K_2)$. Блочное ПВ-шифрование сообщений T и M, имеющих одинаковый размер, задается следующим алгоритмом:

1) разбить сообщения T и M на 64-битовые блоки T_i и M_i:

$$T = (T_1, T_2, \ldots, T_i, \ldots, T_z);$$

$$M = (M_1, M_2, \ldots, M_i, \ldots, M_z);$$

2) каждый i-й блок T_i и каждый i-й ($i = 1, 2, \ldots, z$) блок M_i зашифровать, выполнив следующие два шага:

2.1) зашифровать блок данных T_i по ключу Q

$$C_{Ti} = E_Q(T_i);$$

2.2) зашифровать блок данных M_i по ключу K

$$C_{Mi} = E_K(M_i);$$

3) для каждого значения $i = 1, 2, \ldots, z$ сформировать 130-битовый блок криптограммы C_i в виде конкатенации двух 65-битовых значений C_i' и C_i'': $C_i = (C_i', C_i'')$, являющихся решением системы линейных уравнений с неизвестными C_i' и C_i'':

$$\begin{cases} K_1C_i' + K_2C_i'' \equiv C_{M_i} \bmod p; \\ Q_1C_i' + Q_2C_i'' \equiv C_{T_i} \bmod p, \end{cases} \quad (3)$$

где p – некоторое специфицированное простое число длиной 65 бит, такое, что операция умножения по модулю p

может быть выполнена без операции арифметического деления на p (например, $p = 18446744073709551629$; $p = 18446744073709553681$).

Ассоциируемый алгоритм вероятностного шифрования описывается следующим образом:

1) разбить сообщение M на 64-битовые блоки M_i

$$M = (M_1, M_2, \ldots, M_z);$$

2) каждый i-й ($i = 1, 2, \ldots, z$) блок зашифровать, выполнив следующие три шага:

2.1) зашифровать блок данных M_i по ключу K с использованием n-битового блочного алгоритма шифрования E по формуле $C_{Mi} = E_K(M_i)$;

2.2) сгенерировать случайные числа $R < p$ и $r < p$;

2.3) вычислить i-й блок криптограммы C_i как решение системы сравнений

$$\begin{cases} K_1 C_i' + K_2 C_i'' \equiv C_{M_i} \bmod p; \\ C_i' + r_i C_i'' \equiv R \bmod p. \end{cases} \quad (4)$$

При фиксированном ключе K каждый i-й блок C_i криптограммы в общем случае может быть получен с помощью ассоциированного алгоритма вероятностного шифрования при различных парах значений R и r. Действительно, выбор произвольного числа r однозначно определяет значение R, при котором второе уравнение в (4) будет выполняться для фиксированного значения C_i.

Вместо (4) в ассоциируемом алгоритме вероятностного шифрования можно задать систему линейных уравнений, в которой второе уравнение имеет более простой вид:

$$\begin{cases} K_1 C_i' + K_2 C_i'' \equiv C_{M_i} \bmod p; \\ C_i' \equiv r_i C_i'' \bmod p. \end{cases}$$

Расшифровывание криптограммы $C = (C_1, C_2, \ldots, C_i, \ldots, C_z)$ по фиктивному ключу $K = (K_1, K_2)$ выполняется следующим образом:

1) каждый i-й ($i = 1, 2, \ldots, z$) 130-битовый блок C_i представить в виде конкатенации двух 65-битовых подблоков C_i' и C_i'' и расшифровать его, выполнив следующие два шага:

1.1) вычислить 64-битовый блок промежуточного шифртекста

$$C_{M_i} \equiv K_1 C_i' + K_2 C_i'' \bmod p;$$

1.2) расшифровать блок C_{Mi} по ключу K, используя функцию блочного расшифровывания $D = E^{-1}$:

$$M_i = D_K(C_{Mi});$$

2) объединить все восстановленные 64-битовые блоки данных M_i в сообщение

$$M = (M_1, M_2, \ldots, M_i, \ldots, M_z).$$

Раскрытие секретного сообщения из криптограммы $C = (C_1, C_2, \ldots, C_z)$ выполняется по этому же алгоритму при использовании секретного ключа $Q = (Q_1, Q_2)$:

1) каждый i-й ($i = 1, 2, \ldots, z$) 130-битовый блок C_i расшифровать, выполнив следующие два шага:

1.1) вычислить 64-битовый блок промежуточного шифртекста $C_{T_i} \equiv Q_1 C_i' + Q_2 C_i'' \bmod p$;

1.2) расшифровать блок C_{Ti} по ключу Q, используя функцию блочного расшифровывания $D = E^{-1}$:

$$T_i = D_Q(C_{Ti});$$

2) объединить все восстановленные 64-битовые блоки данных T_i в сообщение $T = (T_1, T_2, \ldots, T_z)$.

Подблоки C_i' и C_i'' имеют размер, равный 65 бит, а блок C_i – 130 бит. Это на 2 бита больше, чем суммарная длина блоков входных данных. Чтобы обеспечить равенство размера объединенного блока криптограммы сумме размеров входных блоков данных, следует воспользоваться заданием системы уравнений, аналогичной (3), в конечном поле двоичных многочленов.

Алгоритмы на основе решения систем линейных уравнений в конечных двоичных полях

Пусть требуется выполнить совместное шифрование двух сообщений $T = (T_1, T_2, \ldots, T_z)$ и $M = (M_1, M_2, \ldots, M_z)$, где T_i и M_i – 128-битовые блоки. Произведем промежуточное шифрование каждого блока данных M_i с помощью алгоритма блочного шифрования E с размером входного блока $n = 128$ бит и 256-битовым ключом K, представленным в виде пары 128-битовых подключей K_1 и K_2: $K = (K_1, K_2)$. Для промежуточного шифрования блоков T_i также воспользуемся блочным шифром E и другим 256-битовым ключом Q, представленным в виде пары 128-битовых подключей Q_1 и Q_2: $Q = (Q_1, Q_2)$. Генерацию ключей K и Q выполним как генерацию пар равновероятных случайных 128-битовых строк, рассматриваемых как двоичные многочлены и удовлетворяющих условию $K_1 Q_2 - K_2 Q_1 \neq 0 \bmod \eta(x)$, где $\eta(x)$ – неприводимый двоичный многочлен степени 128, являющийся специфицируемым параметром ПВ-шифра.

Блочное ПВ-шифрование сообщений M и T можно задать по следующему алгоритму:

1) каждый i-й 128-битовый блок данных T_i и каждый i-й ($i = 1, 2, \ldots, z$) блок M_i зашифровать, выполнив следующие два шага:

1.1) зашифровать 128-битовый блок данных M_i по ключу K

$$C_{Mi} = E_K(M_i);$$

1.2) зашифровать 128-битовый блок данных T_i по ключу Q

$$C_{Ti} = E_Q(T_i);$$

2) для каждого значения $i = 1, 2, \ldots, z$ сформировать 256-битовый блок криптограммы C_i в виде конкатенации двух 128-битовых двоичных многочленов C_i' и C_i'': $C_i = (C_i', C_i'')$, являющихся решением системы линейных уравнений с неизвестными C_i' и C_i'':

$$\begin{cases} K_1 C_i' + K_2 C_i'' \equiv C_{M_i} \bmod \eta(x); \\ Q_1 C_i' + Q_2 C_i'' \equiv C_{T_i} \bmod \eta(x). \end{cases} \quad (5)$$

Ассоциируемый алгоритм вероятностного шифрования описывается следующим образом:

1) разбить сообщение M на 128-битовые блоки M_i:

$$M = (M_1, M_2, \ldots, M_z);$$

2) каждый i-й ($i = 1, 2, \ldots, z$) блок зашифровать, выполнив следующие три шага:

2.1) зашифровать блок данных M_i по ключу Q с использованием 128-битового блочного алгоритма шифрования E по формуле $C_{Mi} = E_Q(M_i)$;

2.2) сгенерировать случайные двоичные многочлены $\lambda(x)$ и $\rho(x)$ степени 127;

2.3) вычислить i-й 256-битовый блок криптограммы $C_i = (C_i', C_i'')$ как решение системы уравнений

$$\begin{cases} K_1 C_i' + K_2 C_i'' \equiv C_{M_i} \bmod \eta(x); \\ C_i' + \lambda(x) C_i'' \equiv \rho(x) \bmod \eta(x). \end{cases} \quad (6)$$

Заданный блок криптограммы C_i может быть получен с помощью ассоциированного алгоритма вероятностного шифрования при фиксированных значениях фиктивного ключа K и блока промежуточного шифртекста C_{Mi} при различных парах значений многочленов $\lambda(x)$ и $\rho(x)$. При этом выбор произвольного многочлена $\lambda(x)$ однозначно определяет значение $\rho(x)$, для которого система уравнений (5) в качестве своего решения будет иметь пару многочленов C_i', C_i'', таких, что $C_i = (C_i', C_i'')$.

Расшифровывание криптограммы $C = (C_1, C_2, \ldots, C_z)$ по фиктивному ключу $K = (K_1, K_2)$ выполняется следующим образом:

1) каждый i-й ($i = 1, 2, \ldots, z$) 256-битовый блок C_i расшифровать, выполнив следующие два шага:

1.1) вычислить 128-битовый блок промежуточного шифртекста по формуле

$$C_{M_i} = K_1 C_i' + K_2 C_i'' \bmod \eta(x);$$

1.2) расшифровать 128-битовый блок C_{Mi} по ключу K, используя функцию блочного расшифровывания $D = E^{-1}$

$$M_i = D_K(C_{Mi});$$

2) объединить все восстановленные блоки M_i в сообщение $M = (M_1, M_2, \ldots, M_i, \ldots, M_z)$.

Раскрытие секретного сообщения из криптограммы $C = (C_1, C_2, \ldots, C_z)$ выполняется по ключу $Q = (Q_1, Q_2)$ следующим образом:

1) каждый i-й ($i = 1, 2, \ldots, z$) 256-битовый блок C_i расшифровать, выполнив следующие два шага:

1.1) вычислить 128-битовый блок промежуточного шифртекста

$$C_{T_i} \equiv Q_1 C_i' + Q_2 C_i'' \bmod \eta(x);$$

1.2) расшифровать 128-битовый блок C_{Ti} по ключу Q, используя функцию блочного расшифровывания $D = E^{-1}$,

$$T_i = D_K(C_{Ti});$$

2) объединить все восстановленные блоки T_i в сообщение $T = (T_1, T_2, \ldots, T_z)$.

В описанном ПВ-шифре в качестве функции блочного шифрования E можно использовать 128-битовый алгоритм блочного шифрования, рекомендуемый стандартом ГОСТ Р 34.12-2015. В общем случае для шифрования сообщений T и M можно использовать блочные шифры с разными размерами входного блока. Например, сообщение T можно разбивать на 64-битовые блоки данных T_i и шифровать последние с помощью 64-битового шифра, а сообщение M – на 128-битовые блоки M_i с последующим их шифрованием с использованием 128-битового блочного шифра. Однако блоки промежуточного шифртекста потребуется объединять путем совместного решения двух линейных уравнений (5), записанных по модулю одного и того же неприводимого двоичного многочлена $\eta(x)$ степени 128. Это приведет к тому, что объединенный блок C_i криптограммы будет иметь размер 256 бит, что превышает сумму размеров блоков T_i и M_i.

Для разбиения сообщений T и M на блоки данных разного размера предпочтительно использовать построение алгоритма ПВ-шифрования с объединением блоков промежуточных шифртекстов путем решения системы из двух уравнений, в которой в качестве модулей можно использовать двоичные многочлены разной степени (например, 128 и 64 бит), за счет чего можно обеспечить равенство размера объединенного блока (192 бит) сумме размеров блоков промежуточных шифртекстов, имеющих различную длину.

Заключение

В настоящей статье показана перспективность применения ПВ-шифрования для построения новых механизмов защиты информации, сформулированы основные требования к ПВ-шифрам, рассмотрены способы построения и предложены конкретные алгоритмы блочного ПВ-шифрования с разделяемым ключом. Выполнимость критерия вычислительной неразличимости по шифртексту блочного ПВ-шифра от блочного вероятностного шифра доказывается наличием ассоциированного алгоритма вероятностного шифрования.

Предложенные способы блочного ПВ-шифрования задают процедуру совместного зашифровывания двух сообщений, однако они легко расширяются для одновременного шифрования трех и более сообщений.

Дальнейшее развитие тематики ПВ-шифрования можно связать с разработкой рандомизированных ПВ-шифров, с поиском новых способов задания взаимно однозначного отображения блоков промежуточных шифртекстов в единый блок выходной криптограммы и ПВ-шифров, включающих процедуру предварительного сжатия входных сообщений.

Литература

1. Dachman-Soled D. On minimal assumptions for sender-deniable public key encryption // Public-Key Cryptography-PKC 2014: 17th Int. Conf. Practice and Theory in Public-Key Cryptography. Lecture Notes Comp. Sci. 2014. Vol. 8383. P. 574-591.

2. Ibrahim M. H. A Method for Obtaining Deniable Public-Key Encryption // Int. J. Network Security. 2009. Vol. 8, № 1. P. 1-9.

3. Canetti R., Dwork C., Naor M., Ostrovsky R. Deniable Encryption // Proc. Advances in Cryptology – CRYPTO 1997. Lectute Notes in Computer Science. Springer – Verlag. 1997. Vol. 1294. P. 90-104.

4. Ishai Yu., Kushilevits E., Ostrovsky R. Efficient non-interactive secure computation // Advances in Cryptology – EUROCRYPT 2011. Lectute Notes in Computer Science. Springer – Verlag. 2011. Vol. 6632. P. 406-425.

5. Meng B. A secure Internet voting protocol based on non-interactive deniable authentication protocol and proof protocol that two ciphertexts are encryption of the same plaintext // J. Networks. 2009. Vol. 4. P. 370-377.

6. Barakat T. M. A New Sender-Side Public-Key Deniable Encryption Scheme with Fast Decryption // KSII Transactions on Internet and Information Systems. 2014. Vol. 8, №. 9. P. 3231-3249.

7. Молдовян Н. А., Биричевский А. Р., Мондикова Я. А. Отрицаемое шифрование на основе блочных шифров // Информационно-управляющие системы. 2014. № 5. С. 80-86.

8. Moldovyan N.A., Shcherbacov A.V., Eremeev M.A. Deniable-encryption protocols based on commutative ciphers // Quasigroups and related systems. 2017. Vol. 25, № 1. P. 95-108.

9. Moldovyan N.A. Berezin A.N., Kornienko A.A., Moldovyan A.A., Bi-deniable Public-Encryption Protocols Based on Standard PKI // Proc. 18th FRUCT & ISPIT Conf., 18-22 Apr. 2016. St. Petersburg. P. 212-219.

10. Moldovyan N.A., Moldovyan A.A., Moldovyan D.N., Shcherbacov V.A. Stream Deniable-Encryption Algorithms // Comput. Sci. J. Moldova. 2016. Vol. 24, № 1 (70). P. 68-82.

11. Морозова Е.В, Мондикова Я. А., Молдовян Н. А. Способы отрицаемого шифрования с разделяемым ключом // Информационно-управляющие системы. 2013. № 6. С. 73-78.

12. Березин А. Н., Биричевский А. Р., Молдовян Н. А., Рыжков А. В. Способ отрицаемого шифрования // Вопр. защиты информации. 2013. № 2. С. 18-21.

Deniable-encryption Methods Based on Block Ciphers

Moldovyan A. A.
Saint-Petersburg Institute for Informatics and Automation of RAS
St. Petersburg, Russia
maa1305@yandex.ru

Tatchina Ya. A.
Saint-Petersburg State Electrotechnical University "LETI"
St. Petersburg, Russia
iana.tatchina@gmail.com

Abstract. **Pseudo-probabilistic encryption is presented as a new algorithmic mechanism for ensuring information security, which implements information protection in the event of attacks with compelling disclosure of the encryption key. The basic requirement for transformations of this type is the computational indistinguishability of the ciphertext from probabilistic encryption. We consider methods and algorithms that implement pseudo-probability encryption as a simultaneous cryptographic transformation of fictitious and secret messages in two different keys, consisting in the formation of blocks of intermediate ciphertexts and their reversible mapping into a single extended block of the output cryptogram. Algorithms are proposed that include the task of the unifying mapping procedure in the form of solutions of systems of linear equations and comparisons in which numbers and binary polynomials are used as a module. The proposed methods have a sufficiently high productivity and are of considerable interest for practical application in information security systems.**

Keywords: **cryptography, denied encryption, pseudo-probabilistic encryption, symmetric encryption, block ciphers, probabilistic encryption, cryptogram.**

References

1. Dachman-Soled D. On minimal assumptions for sender-deniable public key encryption, *Public-Key Cryptography-PKC 2014: 17th Int. Con. Practice and Theory in Public-Key Cryptography. Lecture Notes Comp. Sci.*, 2014, Vol. 8383, pp. 574-591.

2. Ibrahim M. H. A Method for Obtaining Deniable Public-Key Encryption, *Int. J. of Network Security*, 2009, Vol. 8, no. 1, pp. 1-9.

3. Canetti R., Dwork C., Naor M., Ostrovsky R. Deniable Encryption, *Proc. Advances in Cryptology – CRYPTO 1997. Lectute Notes in Computer Science. Springer – Verlag*, 1997, Vol. 1294, pp. 90-104.

4. Ishai Yu., Kushilevits E., Ostrovsky R. Efficient non-interactive secure computation, *Advances in Cryptology – EUROCRYPT 2011. Lectute Notes in Computer Science. Springer – Verlag*, 2011, Vol. 6632, pp. 406-425.

5. Meng B. A secure Internet voting protocol based on non-interactive deniable authentication protocol and proof protocol that two ciphertexts are encryption of the same plaintext*, J. Networks*, 2009, Vol. 4, pp. 370-377.

6. Barakat T. M. A New Sender-Side Public-Key Deniable Encryption Scheme with Fast Decryption, *KSII Transactions on Internet and Information Systems.* 2014, Vol. 8, no. 9, pp. 3231-3249.

7. Moldovyan N. A., Birichevskiy A. R., Mondikova Ya. A. Deniable Encryption Based on Block Ciphers [Otritsaemoe shyfrovanie na osnove blochnyh shyfrov], *Information-control systems [Informatsionno-upravlyayuschchie sistemy]*, 2014, no. 5, pp. 80-86. (In Russ.)

8. Moldovyan N. A., Shcherbacov A. V., Eremeev M. A. Deniable-encryption protocols based on commutative ciphers, *Quasigroups and related systems*. 2017. Vol. 25, no. 1, pp. 95-108.

9. Moldovyan N. A. Berezin A. N., Kornienko A. A., Moldovyan A. A. Bi-deniable Public-Encryption Protocols Based on Standard PKI. *Proc. 18th FRUCT & ISPIT Conf.*, 18-22 Apr. 2016. St. Petersburg. P. 212-219.

10. Moldovyan N. A., Moldovyan A. A., Moldovyan D. N., Shcherbacov V. A. Stream Deniable-Encryption Algorithms, *Comput. Sci. J. Moldova*, 2016, Vol. 24, no. 1 (70), pp. 68-82.

11. Morozova E. V., Mondikova Ya. A., Moldovyan N. A. Methods for Deniable Encryption with Shared Key [Sposoby otritsaemogo shifrovaniya s razdelyaemym klyuchom]. *Information-control systems [Informatsionno-upravlyayuschchie sistemy]*, 2013, no. 6, pp. 73-78. (In Russ.)

12. Berezin A. N., Birichevskiy A. R., Moldovyan N. A., Ryzgkov A. V. Method for Deniable Encryption [Sposob otritsaemogo shifrovaniya]. *Items of Information Protection [Voprosy zashchity informatsii]*, 2013, no. 2, pp. 18-21. (In Russ.)

Генерация степенных сравнений как способ открытого шифрования и протокол отрицаемого шифрования

Молдовян Н. А.
Санкт-Петербургский институт
информатики и автоматизации РАН
Санкт-Петербург, Россия
nmold@mail.ru

Вайчикаускас М. А.
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
Цюрих, Швейцария
m.vaichikauskas@gmail.com

Аннотация. **Представлен новый способ открытого шифрования, в котором процесс зашифровывания выполняется путем генерации коэффициентов кубичного уравнения, а процесс расшифровывания заключается в решении данного уравнения. Безопасность данного метода основывается на сложности задачи факторизации, а именно на сложности факторизации составного модуля, который служит открытым ключом. Секретный ключ представляет собой пару чисел p и q, таких что $n = pq$. Процесс расшифровывания выполняется путем решения кубичного сравнения по модулю n. Первым шагом данного процесса является нахождение корней уравнения в полях $GF(p)$ и $GF(q)$. В работе предлагается метод решения кубичных уравнений в простых конечных полях. Предложенный способ открытого шифрования применен для построения протокола отрицаемого шифрования, стойкого к двусторонним принуждающим атакам.**

Ключевые слова: **криптография, шифрование, открытое шифрование, отрицаемое шифрование, открытый ключ, вероятностное шифрование, задача факторизации, кубичные уравнения, простое конечное поле.**

Введение

Новый способ открытого шифрования состоит в генерации шифртекста в виде набора коэффициентов уравнений второй, третьей и четвертой степеней [1, 2]. Он позволяет одновременно шифровать два, три и четыре сообщения в одну криптограмму, благодаря чему может быть реализовано отрицаемое шифрование (ОШ), которое представляет собой криптографический механизм, обеспечивающий защиту информации в случае принуждающих атак. Подробно концепция ОШ и некоторые области его применения рассмотрены в работах [3, 4]. В [5] обоснована состоятельность ОШ с разделяемым секретным ключом как нового самостоятельного механизма защиты информации от несанкционированного доступа, перспективного для применения в средствах обеспечения компьютерной безопасности, использующихся в том числе для защиты телекоммуникаций в информационно-вычислительных системах на железнодорожном транспорте.

Способ открытого шифрования, основанный на генерации степенных уравнений и включающий одновременное шифрование двух и более сообщений, предполагает неоднозначность процедуры расшифровывания. В данной статье рассматривается способ открытого шифрования с использованием кубичных уравнений, свободный от неоднозначности расшифровывания шифртекста, что является его существенным преимуществом по сравнению с аналогичными способами. Также предложен способ отрицаемого шифрования, в основу которого положен разработанный способ открытого шифрования.

Новый способ открытого шифрования

Так же, как в алгоритме шифрования [1], личным секретным ключом владельца открытого ключа n является пара сильных простых чисел p и q [6], таких, что $n = pq$, $p^2 = 7 \bmod 9$, $q^2 = 7 \bmod 9$, и число 3 не делит ни одно из чисел $(p-1)$ и $(q-1)$.

Идея обеспечения однозначности процедуры расшифровывания, состоящей в решении кубичного уравнения над конечным кольцом классов по модулю n, состоит в генерации кубичного выражения, которое представляется в виде произведения многочлена первой степени $(x - M)$, где M – шифруемое сообщение $(M < n)$, и многочлена второй степени $(x^2 + Zx + Y)$, не имеющего корней в указанном кольце. При этом случайные значения $Z < n$ и $Y < n$ генерируются так, что дискриминант многочлена второй степени является невычетом. Однако шифрование выполняется лицами, которым неизвестны делители трудно факторизуемого числа n, поэтому требуется расширить открытый ключ, включив в него некоторый квадратичный невычет $N < n$, который генерируется владельцем открытого ключа после генерации личного секретного ключа. Для этого владелец открытого ключа генерирует случайное значение, для которого выполняются следующие два условия:

$$N^{\frac{p-1}{2}} \equiv -1 \bmod p;$$
$$N^{\frac{q-1}{2}} \equiv -1 \bmod q.$$

Открытым ключом является пара чисел N и n.

Алгоритм шифрования сообщения M по открытому ключу описывается следующим образом:

1) сформировать случайное число $Z < n$ и вычислить значение по формуле

$$Y = Z^2/4 - N \bmod n;$$

2) вычислить коэффициенты A, B и D кубичного уравнения

$$x^3 + Ax^2 + Bx + D = 0 \bmod n, \qquad (1)$$

используя формулы $A = Z - M \bmod n$, $B = Y - MZ \bmod n$ и $D = -MY \bmod n$, которые вытекают из условия $x^3 + Ax^2 + Bx + D = (x - M)(x^2 + Zx + Y)$.

Таким образом, сложность процедуры формирования шифртекста $C = (A, B, D)$ примерно равна трем операциям умножения по модулю n. Криптограмма C направляется по открытому каналу владельцу открытого ключа (N, n), который расшифровывает ее с использованием своего личного секретного ключа. Процедура расшифровывания состоит в решении кубичного уравнения (1), заданного коэффициентами A, B и D, т. е. в нахождении единственного его корня, равного значению M.

Решение уравнения (1) выполняется следующим образом. Решаются следующие два кубичных уравнения:

$$x^3 + Ax^2 + Bx + D \equiv 0 \bmod p; \quad (2)$$

$$x^3 + Ax^2 + Bx + D \equiv 0 \bmod q \quad (3)$$

в простых конечных полях $GF(p)$ и $GF(q)$, соответственно. Каждое из этих уравнений имеет единственный корень в простом конечном поле. Пусть корень уравнения (2) равен M_p, а уравнения (3) – M_q. Тогда корень уравнения (1) находится как решение системы линейных сравнений

$$\begin{cases} M \equiv M_p \bmod p \\ M \equiv M_q \bmod q \end{cases}. \quad (4)$$

В соответствии с греко-китайской теоремой об остатках решением системы (4) является значение

$$M = \left[M_p q \left(q^{-1} \bmod p \right) + M_q p \left(p^{-1} \bmod q \right) \right] \bmod pq. \quad (5)$$

Трудоемкость расшифровыванию задает решение уравнений (2) и (3). Рассмотрим, как можно решить кубичное уравнение в конечном поле, например уравнения (2).

По аналогии со способом решения кубичных уравнений в поле действительных чисел [7] путем замены переменной по формуле $x = z - A/3 \bmod p$ перейдем к решению следующего уравнения, свободного от квадрата неизвестного:

$$z^3 + Pz + Q = 0 \bmod p, \quad (6)$$

где

$$P = B - \frac{A^2}{3} \bmod p;$$

$$Q = \frac{2A^3}{27} - \frac{AB}{3} + D \bmod p.$$

Поскольку заведомо известно, что есть решение уравнения (6), вывод формулы для корней кубичного уравнения, приведенный в [7, с. 234–238], можно применить и в рассматриваемом случае. Такой подход дает следующую формулу для корней (6):

$$z = \alpha + \beta, \quad (7)$$

где

$$\alpha = \sqrt[3]{-\frac{Q}{2} + \sqrt{\frac{Q^2}{4} + \frac{P^3}{27}}} \bmod p;$$

$$\beta = \sqrt[3]{-\frac{Q}{2} - \sqrt{\frac{Q^2}{4} + \frac{P^3}{27}}} \bmod p. \quad (8)$$

При этом при наличии нескольких значений кубичного корня в формулах (8) следует выбирать такие пары, которые удовлетворяют соотношению

$$\alpha\beta = -\frac{P}{3}. \quad (9)$$

Поскольку при открытом шифровании используются случайные значения, т. е. этот процесс является вероятностным, под квадратным корнем в выражениях (8) может оказаться значение, которое является как квадратичным вычетом, так и квадратичным невычетом. В первом случае значение квадратного корня имеется, и его можно легко вычислить (см., например, [8]). При этом из-за условия выбора простых чисел в качестве секретного ключа (число 3 не делит ни одно из чисел $p - 1$ и $q - 1$) есть единственный кубичный корень из любого значения, поэтому существует единственный корень уравнения (6), который вычисляется непосредственно по формулам (7) и (8).

Во втором случае, чтобы найти решение уравнения (6), требуется перейти в поле $GF(p^2)$, выполнить вычисления по формулам (7) и (8) и взять в качестве искомых значений те, которые лежат в поле $GF(p)$, вложенном в поле $GF(p^2)$. В качестве поля $GF(p^2)$ удобно выбрать поле двоичных векторов, заданных над полем $GF(p)$ [8, 9] при определении операций сложения и умножения векторов $\boldsymbol{V} = (a, b) \in GF(p^2)$ и $\boldsymbol{U} = (c, d) \in GF(p^2)$ по формулам

$$\boldsymbol{V} + \boldsymbol{U} = (a, b) + (c, d) = (a + c \bmod p, b + d \bmod p); \quad (10)$$

$$\boldsymbol{VU} = (a, b)(c, d) = (ac + kbd \bmod p, bc + ad \bmod p), \quad (11)$$

где $k \in GF(p)$ – некоторый заданный квадратичный невычет.

Вместо уравнения (6) будем рассматривать уравнение над конечным полем двоичных векторов вида:

$$\boldsymbol{Z}^3 + \boldsymbol{PZ} + \boldsymbol{Q} = (0, 0), \quad (12)$$

где

$$P = (P, 0) = (B, 0) - \frac{(A, 0)^2}{3};$$

$$Q = (Q, 0) = \frac{2}{27}(A, 0)^3 - \frac{(A, 0)(B, 0)}{3} + (D, 0).$$

Если заведомо известно, что есть решение уравнения (6), то есть и решения уравнения (12), тогда имеем формулу для корней (12)

$$\boldsymbol{Z} = \boldsymbol{A} + \boldsymbol{B}, \quad (13)$$

где

$$A = \sqrt[3]{-\frac{Q}{2} + \sqrt{\frac{Q^2}{4} + \frac{P^3}{27}}}; \; B = \sqrt[3]{-\frac{Q}{2} - \sqrt{\frac{Q^2}{4} + \frac{P^3}{27}}}. \quad (14)$$

При этом при наличии нескольких значений кубичного корня в формулах (8) следует выбирать такие пары, которые удовлетворяют соотношению

$$AB = -\frac{P}{3}. \quad (15)$$

Заметим, что имеются следующие положения.

Утверждение 1. Пусть $a \in GF(p)$ есть квадратичный невычет, тогда для $(a, 0) \in GF(p^2)$ выполняется

$$\sqrt{(a,0)} = (0, \pm\sqrt{k^{-1}a}),$$

где k – квадратичный невычет, заданный в формуле (11), определяющей правило выполнения операции умножения в поле двоичных векторов.

Доказательство. В соответствии с формулой (12) получаем

$$(0, \pm\sqrt{k^{-1}a})^2 = \sqrt{(a,0)}. \quad \Diamond$$

Утверждение 2. В поле $GF(p^2)$, где $p > 3$, из произвольного кубичного вычета $\boldsymbol{A}$ есть три разных кубичных корня.

Доказательство. Любое простое число $p > 3$ представимо в виде $p = 6t \pm 1$ при некотором натуральном числе t, поэтому порядок мультипликативной группы поля $GF(p^2)$ представим в виде $p^2 - 1 = 36 \pm 12t$, т. е. число 3 – порядок этой конечной группы. Последняя является циклической, поэтому в ней есть ровно два элемента, имеющих порядок, равный 3. Пусть это будут элементы $\boldsymbol{E}$ и $\boldsymbol{E}^2$. Последние являются нетривиальными корнями из единичного элемента $(1,0) \in GF(p^2)$. Если $\boldsymbol{B}$ является кубичным корнем из $\boldsymbol{A}$, то $\boldsymbol{EB}$ и $\boldsymbol{E}^2\boldsymbol{B}$ – также кубичные корни из $\boldsymbol{A}$: $(\boldsymbol{EB})^3 = \boldsymbol{E}^3\boldsymbol{B}^3 = \boldsymbol{B}^3 = \boldsymbol{A}$ и $(\boldsymbol{E}^2\boldsymbol{B})^3 = \boldsymbol{E}^6\boldsymbol{B}^3 = \boldsymbol{B}^3 = \boldsymbol{A}$.

Предположение о существовании четвертого корня $B' = \sqrt[3]{A}$ приводит к наличию в конечной циклической группе более двух элементов, имеющих порядок 3. Это противоречие доказывает, что в поле $GF(p^2)$ имеются ровно три разных кубичных корня из каждого кубичного вычета.

Утверждение 3. В поле $GF(p^2)$ при $p^2 \equiv 7 \bmod 9$ один из кубичных корней $\boldsymbol{B}$ из кубичного вычета $\boldsymbol{A}$ может быть найден по формуле

$$B = A^{\frac{p^2+2}{9}}. \quad (16)$$

Доказательство. Поскольку $\boldsymbol{A}$ – кубичный вычет, есть некоторый элемент $\boldsymbol{X} \in GF(p^2)$, для которого выполняется $\boldsymbol{X}^3 = \boldsymbol{A}$, поэтому имеем

$$A^{\frac{p^2-1}{3}} = X^{p^2-1} = (1,0).$$

Следовательно,

$$B^3 = A^{\frac{p^2+2}{3}} = AA^{\frac{p^2-1}{3}} = A.$$

В силу того, что уравнение (12) заведомо имеет решения, вычисляемые по формулам (13) и (14), под кубичным корнем в (14) будут присутствовать кубичные вычеты. Вычисление кубичных корней в выражениях (14) может быть выполнено по формуле (16).

В целом в сравнении с процедурой зашифровывания расшифровывание имеет существенно более высокую вычислительную сложность, однако при использовании быстрого алгоритма возведения в большую степень в конечных полях $GF(p)$ и $GF(p^2)$ она выполняется достаточно быстро – в полтора-два раза медленнее процедуры расшифровывания в протоколе открытого шифрования Эль-Гамаля [10]. Однако в последнем процедура зашифровывания имеет вычислительную сложность в сотни раз более высокую, чем сложность зашифровывания в описанном способе открытого шифрования.

Протокол отрицаемого шифрования

Предложенный способ шифрования по открытому ключу может быть использован в протоколе ОШ, стойком к двухсторонним принуждающим атакам. Алгоритм ОШ строится следующим образом. Фиктивное сообщение $M < n$ зашифровывается по способу ОШ, описанному ранее, а секретное сообщение встраивается в значение параметра Z. Это встраивание выполняется с помощью разового разделяемого секретного ключа U, который согласовывается на этапе взаимной аутентификации пользователей путем умножения секретного сообщения T на ключ U по модулю n: $Z = TU \bmod n$. При восстановлении фиктивного сообщения одновременно восстанавливается и значение псевдослучайного параметра Z. Затем вычисляется секретное сообщение по формуле $T = U^{-1}Z \bmod n$.

Для взаимной аутентификации пользователей воспользуемся генерацией случайного запроса одним пользователем и вычислением ответа на этот запрос другим пользователем. В качестве ответа зададим значение цифровой подписи пользователя к полученному случайному запросу. В качестве специфицируемой схемы цифровой подписи выберем криптосистему RSA, поскольку трудноразложимый модуль уже имеется в качестве одного из элементов открытого ключа. Чтобы выполнять процедуры формирования и проверки подлинности цифровой подписи в рамках криптосхемы RSA, открытый ключ должен быть дополнен третьим элементом – числом e, таким, что $e < \varphi(n)$ и НОД $(e, \varphi(n)) = 1$. При этом личный секретный ключ также дополняется третьим элементом – числом $d = e^{-1} \bmod (p-1)(q-1)$.

Такая криптосхема реализована в следующем протоколе ОШ, в котором предполагается, что отправитель сообщения владеет открытым ключом (n_1, e_1, N_1), а получатель – открытым ключом (n_2, e_2, N_2).

1. Отправитель генерирует случайное число k_1, удовлетворяющее условию $0 < k_1 < N_1 - 1$, и вычисляет значение

$$R_1 = N_2^{k_1} \bmod n_2.$$

Затем формирует свою подпись

$$S_1 = S_1(R_1) = R_1^{d_1} \bmod n_1$$

к значению R_1 и направляет значения R_1 и $S_1(R_1)$ получателю.

2. Получатель, используя открытый ключ (n_1, e_1, N_1), проверяет подлинность подписи $S_1(R_1)$. Если подпись подлинная, то он генерирует случайное число k_2, удовлетворяющее условию $0 < k_2 < p - 1$, и вычисляет значение

$$R_2 = N_2^{k_2} \bmod n_2.$$

Затем формирует свою подпись $S_2(R_1)$ к значению R_1 и свою подпись $S_2(R_2)$ к значению R_2 и направляет значения R_2, $S_2(R_1)$ и $S_2(R_2)$ отправителю.

3. Отправитель, используя открытый ключ (n_2, e_2, N_2), проверяет подлинность подписи $S_2(R_1)$ к случайному значению,

которое он направлял получателю, и подлинность подписи S_2 (R_2) к разовому открытому ключу R_2 получателя. Если подпись подлинная, то он зашифровывает и передает секретное сообщение T ($T < n_2$) получателю, выполняя следующие шаги (в противном случае он прерывает сеанс связи):

3.1) генерирует фиктивное сообщение $M < n_2$;

3.2) вычисляет значение

$$U = R_2^{k_1} \bmod n_2,$$

легко заметить, что

$$U = N_2^{k_2 k_1} \bmod n_2,$$

и, используя значение U в качестве разового секретного ключа, вычисляет маскируемый под случайное число шифртекст $Z = UT \bmod n_2$;

3.3) вычисляет значение

$$Y = Z^2/4 - N_2 \bmod n_2$$

и криптограмму $C = (A, B, D)$, используя формулы $A = (Z - M) \bmod n$, $B = (Y - MZ) \bmod n$ и $D = -MY \bmod n$;

3.4) вычисляет свою подпись $S_1(C)$ к криптограмме $C = (A, B, D)$ и направляет значения $S_1(C)$ и C получателю;

4. Получатель проверяет подлинность подписи $S_1(C)$. Если подпись ложная, он игнорирует шифртекст C и прерывает сеанс связи. Если подпись подлинная, то он восстанавливает секретное сообщение T следующим путем:

4.1) подставляет в уравнение (1) значения коэффициентов A, B и D и, используя известные ему значения p и q, вычисляет целочисленный корень (1), который равен фиктивному сообщению M;

4.2) вычисляет значение Z по формуле $Z = (A + M) \bmod n$;

4.3) вычисляет значение разового секретного ключа

$$U' = R_1^{k_2} \bmod n_2,$$

легко заметить, что

$$U' = N_2^{k_1 k_2} \bmod n_2 = U;$$

4.4) Вычисляет секретное сообщение $T = U'^{-1}Z \bmod n_2$.

5. Подвергаясь двухсторонней принуждающей атаке, отправитель раскрывает фиктивное сообщение M, а получатель раскрывает свой личный секретный ключ (p, q, d_2).

Атакующий расшифровывает криптограмму по ключу (p, q, d_2) и убеждается, что значение M раскрыто правильно. Он может также вычислить и псевдослучайное значение Z, однако раскрыть секретное сообщение он не сможет, так как для этого надо знать разовый общий секретный ключ, который не зависит от личных секретных ключей участников сеанса секретной связи.

Чтобы уличить пользователей в обмане, атакующий должен решить задачу дискретного логарифмирования по трудноразложимому модулю n_2. Последняя задача не проще, чем факторизация числа n_2. На этом основано предположение: атакующий не сможет доказать, что случайные значения R_1 и R_2 были использованы как разовые открытые ключи в процедуре согласования разового общего секрета U, т. е. то, что пользователи скрытно использовали криптосхему Диффи – Хеллмана.

В случае активных принуждающих атак нарушитель, выдающий себя за получателя сообщения, обнаруживается на шаге 3 описанного протокола, а нарушитель, выдающий себя за отправителя сообщения, – на шаге 4, где проверяется подлинность подписи отправителя к криптограмме C.

Односторонний алгоритм отрицаемого шифрования

Если в модели потенциального нарушителя не предусматривается принуждающая атака на получателя сообщения и рассматриваются только пассивные атаки, то алгоритм ОШ может быть построен без скрытного использования схемы открытого согласования общего секретного ключа. Построение такого алгоритма может быть выполнено с использованием базового кубичного уравнения (1) и открытого ключа (n, N) получателя на основе включения в протокол следующих шагов.

1. Отправитель секретного сообщения $T < n$ генерирует фиктивное сообщение $M < n$ и вычисляет криптограмму $C = (A, B, D)$ следующим путем:

1.1) вычисляет псевдослучайное значение $Z = (2^{-1} - T)^2 \bmod n$ и значение

$$Y = Z^2/4 - N \bmod n;$$

1.2) вычисляет коэффициенты A, B и D, при которых уравнение (1) имеет единственный корень, по следующим формулам:

$$A = (Z - M) \bmod n;$$

$$B = (Y - MZ) \bmod n;$$

$$D = -MY \bmod n;$$

1.3) отправляет криптограмму $C = (A, B, D)$ получателю.

2. Получатель, используя свой личный секретный ключ, решает кубичное уравнение (1) и находит корень M – фиктивное сообщение. Затем он вычисляет секретное сообщение T, выполняя следующие шаги:

2.1) вычисляет значение Z по формуле $Z = (A + M) \bmod n$;

2.2) вычисляет значения

$$T_i = 2^{-1} - \sqrt{Z} \bmod n,$$

где $i = 1, 2, 3, 4$;

2.3) отбрасывает три случайных значения T_i, а оставшееся осмысленное сообщение берет в качестве восстановленного секретного сообщения T.

В случае принуждения отправителя сообщения к раскрытию использованных параметров шифрования он предоставляет атакующему фиктивное сообщение M и число Z, ссылаясь на последнее как на случайное значение. Выполнив шифрование M по открытому ключу (N, n) при использовании предоставленного «случайного» значения Z, атакующий получит криптограмму $C = (A, B, D)$.

Если она совпадает с шифртекстом, переданным по открытому каналу и по предположению известным атакующему, то последний вынужден признать, что ему честно раскрыли переданное сообщение. Чтобы раскрыть обман, атакующему нужно восстановить секретное сообщение по известному значению Z, однако это требует вычисления квадратного корня по модулю n, что не менее сложно, чем решение задачи факторизации числа n [11, 12]. На вычислительной трудности последней задачи построено достаточно большое число раз-

личных криптосхем, например, RSA [13], открытый шифр Рабина [12], схемы цифровой подписи [1, 14].

Заключение

Предложен способ открытого шифрования, заключающийся в формировании шифртекста в виде набора коэффициентов кубичного уравнения. В отличие от известного способа шифрования с формированием шифртекста в виде набора коэффициентов квадратного уравнения [1], в предложенном способе решена проблема неоднозначности текста, восстанавливаемого при выполнении процедуры расшифровывания. Основным шагом процедуры расшифровывания в предложенном способе является решение кубичных уравнений в простом конечном поле $GF(p)$. Чтобы найти корни таких уравнений в общем случае, предложено решать их в поле $GF(p^2)$, заданном в виде конечного поля двоичных векторов.

Описанный в статье способ открытого шифрования может быть использован для построения протокола ОШ, включающего одновременное шифрование секретного и фиктивного сообщений. При этом фиктивное сообщение восстанавливается как корень кубичного уравнения, а секретное – путем преобразования одного из коэффициентов неразложимого в поле $GF(p)$ квадратного трехчлена.

Литература

1. Молдовян Н. А., Вайчикаускас М. А. Расширение криптосхемы Рабина: алгоритм отрицаемого шифрования по открытому ключу // Вопросы защиты информации. 2014. № 2. С. 12-16.

2. Moldovyan N. A., Moldovyan A. A., Shcherbacov V. A. Provably Sender-Deniable Encryption Scheme // Proc. «The Third Conference of Mathematical Society of the Republic of Moldova» (IMCS-50). Chisinau, 19-23 Aug. 2014, Inst. Mathematics and Computer Science, Academy of Sciences of Moldova. 2014, P. 134-141.

3. Canetti R., Dwork C., Naor M., Ostrovsky R. Deniable Encryption // Advances in Cryptology – CRYPTO 1997. Proc. P. 90-104.

4. Ibrahim M. H. Receiver-Deniable Public-Key Encryption // Int. J. Network Security. 2009. Vol. 8, № 2. P. 159-165.

5. Березин А. Н., Биричевский А. Р., Молдовян Н. А., Рыжков А. В. Способ отрицаемого шифрования // Вопр. защиты информации. 2013. № 2. С. 18-21.

6. Gordon J. Strong primes are easy to find // Advances in cryptology – EUROCRYPT'84. Springer-Verlag LNCS. 1985. Vol. 209. P. 216-223.

7. Курош А. Г. Курс высшей алгебры. – М.: Наука, 1971. 431 с.

8. Молдовян Н. А. Теоретический минимум и алгоритмы цифровой подписи. – СПб.: Петербург-БХВ, 2010. 304 с.

9. Moldovyan N. A., Moldovyanu P. A. Vector form of the finite fields GF (pm) // Bull. Acad. de Stiinte a Republicii Moldova. Matematica. 2009. № 3. P. 57-63.

10. ElGamal T. A public key cryptosystem and a signature scheme based on discrete logarithms // IEEE Transactions on Information Theory. 1985. Vol. IT-31, № 4. P. 469-472.

11. Moldovyan N. A., Moldovyan A. A. Class of Provably Secure Information Authentication Systems // Springer Verlag CCIS 4th Int. Workshop MMM-ANCS'07 Proc. 13-15 Sept. 2007. 2007. Vol. 1. P. 147-152.

12. Коутинхо С. Введение в теорию чисел. Алгоритм RSA. – М.: Постмаркет, 2001. – 323 с.

13. Rabin M. O. Digitalized signatures and public key functions as intractable as factorization // Technical report MIT/LCS/TR-212, MIT Laboratory for Computer Sci. 1979.

14. Moldovyan A. A., Moldovyan N. A., Shcherbakov V. A. Short signatures from difficulty of the factoring problem // Bull. Acad. de Stiinte a Republicii Moldova. Matematica. 2013. № 2-3. P. 27-36.

Generation of Polynomial Equations as a Method for Public Key Encryption and Deniable Encryption Protocol

Moldovyan N.A.
Saint-Petersburg Institute for Informatics and Automation of RAS
St. Petersburg, Russia
nmold@mail.ru

Vaichikauskas M.A.
Saint-Petersburg State Electrotechnical University "LETI"
Zurich, Switzerland.
m.vaichikauskas@gmail.com

Abstract. **The paper introduces a new method for public encryption in which the enciphering process is performed as a generation of the coefficients of some cubic equation and the deciphering process is solving the equation. Security of the method is based on a difficulty of the factoring problem, namely, difficulty of factoring a composite number n that serves as a public key. The private key is the pair of primes p and q such that $n = pq$. The deciphering process is performed as solving cubic congruence modulo n. Finding roots of cubic equations in the fields $GF(p)$ and $GF(q)$ is the first step of the decryption. The paper also describes a method for solving cubic equations defined over prime finite fields. Introduced method of public encryption is applied for development of deniable encryption protocol, which is resistant against two-sided coercive attacks.**

Keywords: **cryptography, encryption, public key encryption, deniable encryption, public key, probabilistic encryption, factorization problem, cubic equation, prime finite field.**

References

1. Moldovan N.A., Vaichikauskas M.A. Rabin Cryptoscheme Expansion: Public-key Deniable Encryption Algorithm [Rasshirenie Kriptoskhemy Rabina: Algoritm otritsaemogo shifrovaniya po otrkrytomu klyuchu]. *Information Security Questions [Voprosy Zashchity Informatsii]*, 2014, no. 2, pp. 12-16. (In Russ.)

2. Moldovyan N.A., Moldovyan A.A., Shcherbacov V.A. Provably Sender-Deniable Encryption Scheme. *Proc. "The Third Conference of Mathematical Society of the Republic of Moldova" (IMCS-50)*. Chisinau, 19-23 Aug. 2014, Institute of Mathematics and Computer Science, Academy of Sciences of Moldova. 2014. P. 134-141.

3. Canetti R., Dwork C., Naor M., Ostrovsky R. Deniable Encryption. *Advances in Cryptology – CRYPTO 1997.* Proc. Pp. 90-104.

4. Ibrahim M.H. Receiver-Deniable Public-Key Encryption. *Int. J. Network Security*, 2009, Vol. 8, no. 2, pp. 159-165.

5. Berezin A.N., Birichevskiy A.R., Moldovyan N.A., Ryzgkov A.V. Method for Deniable Encryption [Sposob otritsaemogo shifrovaniya]. *Items of Information Protection [Voprosy zashchity informatsii]*, 2013, no. 2, pp. 18-21. (In Russ.)

6. Gordon J. Strong primes are easy to find. *Advances in cryptology – EUROCRYPT'84. Springer-Verlag LNCS*, 1985, Vol. 209, pp. 216-223.

7. Kurosh A.G. High Algebra Course [Kurs Vysshey Algebry]. Moscow, Nauka, 1971. 431 p. (In Russ.)

8. Moldovyan N.A. Theoretical Minimum and Algorithms of Digital Signature [Teoreticheskiy Minimum i Algoritmy Tsifrovoy Podpisi]. St. Petersburg, Peterburg-BHV, 2010. 304 p. (In Russ.)

9. Moldovyan N.A., Moldovyanu P.A. Vector form of the finite fields GF (pm). *Bulletinul Academiei de Stiinte a Republicii Moldova. Matematica*, 2009, no. 3, pp. 57-63.

10. ElGamal T. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, 1985, Vol. IT-31, no. 4, pp. 469-472.

11. Moldovyan N.A., Moldovyan A.A. Class of Provably Secure Information Authentication Systems. *Springer Verlag CCIS.* 4th Int. Workshop MMM-ANCS'07 Proc. September 13-15, 2007. 2007, vol. 1. P. 147-152

12. Koutinho S. Introduction to Number Theory. RSA Algorithm [Vvedenie v Teoriyu Chisel. Algoritm RSA]. Moscow, Postmarket, 2001. 323 p. (In Russ.)

13. Rabin M.O. Digitalized signatures and public key functions as intractable as factorization. *Technical report MIT/LCS/TR-212, MIT Laboratory for Computer Sci.*, 1979.

14. Moldovyan A.A., Moldovyan N.A., Shcherbakov V.A. Short signatures from difficulty of the factoring problem. *Bull. Acad. de Stiinte a Republicii Moldova. Matematica*, 2013, no. 2-3, pp. 27-36.

Algorithm of Data Storage Based on Combining of the Methods of Locating and Coding non-Homogeneous Data

Maksimov V.A.
A. F. Mozhaisky Military Aerospace academy
St. Petersburg, Russia
falcon225@yandex.ru

Abstract. **Proposed an approach to the designing and application of advanced spacecraft data storage systems based on heterogeneous storage devices. This approach serving for for the purpose of increasing fault tolerance of information processing systems. Using of heterogeneous storage devices and storage methods allows flexible system management depend on its condition. It allows reduce all kinds of redundancy at storage systems. The novelty of this work lies in taking into account difference of the importance of stored data.**

Keywords: **storage systems, heterogeneity, distribution, fault tolerance, correctness, remote sensing of the Earth, regenerating coding.**

Introduction

The current trends at the development of space satellites are characterized by increasing volume of data collected during spacecraft lifecycle and by a shift at the place of processing these data from ground-based processing units to on-board processing units. This information can be divided by type (telemetry, special, navigation, etc.), or by importance for the consumer (ground control system, onboard systems, etc.).

The largest amounts of data are generated by the earth remote sensing satellites. Its characteristics are improved with every generation and as a result amount of data accumulated on board is significantly growing. However, the problem of fault tolerance onboard data storage is not specific only for remote sensing satellites. Errors occurred during the storage process at commands, software or navigation data may lead to the disasters consequences [1].

Long-term operation of the remote sensing satellites allows to reveal a number of features of the onboard equipment operation:

1) due the high performance of all kinds of spacecraft equipment, it generates a large amounts of data, which requires special approaches to their processing and storage. At the same time, parameters of FLASH-memory modules, used at the storage systems are not full-suitable for fault-tolerant storage.

2) one of the mainstream tends for last decades at Earth remote-sensing is decreasing of data delivering time. One of the solutions for this can be full or partial onboard data processing. But organization of such onboard data processing demands errors check and corrections of this operations. Besides it demands fault-tolerant storage of processing results [2].

3) the demands placed on existing and prospective Earth remote sensing satellites (ERSS) with regard of the lifecycle extension are constantly increasing. However, the methods of reliable space information systems design for long-term effects of space factors are insufficiently developed [3].

4) base data attributes for fault-tolerant maintenance are integrity and errorless. Both of this attributes suffer from corruption and data loss due storage at the data storage system and transferring throw high speed radio links. Increasing the level of integrity and errorless of data, as an integral part of the spacecraft fault-tolerance in general, requires the introduction of various types of redundancy (hardware, software, time redundancy, etc.). But it is not always acceptable in terms of mass-dimensional or functional restrictions for the satellites systems.

One of the new approach to the onboard data storage systems (ODSS) design is construction of data storage systems from modules based on different physical principles [4, 5]. These memory modules (MRAM, FRAM, SONOM, etc.) have some differences at the specifications, which makes them attractive for use at on board satellite systems. It should be noted that the choice of one specific type of memory is difficult because of substantial differences in their characteristics (e. g., having a higher radiation tolerance, they have less available information capacity).

At the same time, it should be noted that the data stored at ODSS have different properties.

For example, system data (data required for the correct operation of the system) have noticeably much importance than results of remote sensing data. Besides the structure of remote sensing data is also heterogeneous.

In most Russian ERS satellites, after the first step of data processing it has stored encoded in a JPG file format. In accordance with international standards (ISO/IEC 10918) in the file container is JFIF, we can identify several data regions.

Thus, the damage or data blocks loss from regions with SOS, APPn, COM markers are not critical for main ERS functions. At worth case this will lead to minor decline of ERS quality results (collision of a few pixels or blocks of pixels at result image). However, even slight corruption DQT, DHI, DRI regions blocks can lead to a significant decline of ERS quality results or to impossibility of remote sensing results perception.

Now exist a significant theoretical and practical groundwork in the field of distributed data storage with the using a regenerating codes [6–8] that allow significantly improve the reliability of data storage compared to the replication method.

All of the above leads to the necessity of development methods for the designing and application of heterogeneous storage systems.

In order to resolve the emerging contradiction between necessity of reliable storage of large amounts of data on board the spacecraft on the one side and the lack of currently developed methods of designing and application of such storage at the other side, proposed an algorithm of data storage based on combining of the methods of locating and coding non-homogeneous data.

Concept of heterogeneous data storage system

Concept of distributed data storage at heterogeneous data storage system (HDSS) composed in designing and application such system at ERS onboard systems. HDSS consist from set of storage units (SU). Each storage unit consists of memory controller (MC) and set of memory modules (MM). HDSS means aggregation of hardware and software parts. Hardware part means structure of HDSS. Software part means a set of processing procedures on data (data encoding, data placement). Introducing of data storage method similar to Big Data at data centers is offered instead of double or triple replication. It means that initial data divides to several data regions according importance index. Each data regions is encoding with regenerating code of certain construction. Next step is determination of data placement (assignment data block with certain importance index to a certain type of memory modules).

Data placement depends on data type and HDSS storage units condition. Before data blocks will write at the MM they should be encoding with error-correcting codes.

As writing above, at the current time exist a couple of works at this field, but there is no one system method for designing and application of fault tolerance HDSS.

Introducing in HDSS different types of memory allows to flexible data placement management and ensure fault tolerance.

Besides that development of the inter satellite communication and evolution of nano-satellite conceptions allows to considerate inter satellite storage network [9, 10].

In this case heterogeneous nature of such system can provide flexible management of data storage process witch is necessary because of low-available nodes at network.

As further development of concept it is possible to introduce at the importance index value of semantic importance of data (importance for the final consumer).

Application of HDSS for ERSS can lead to some drawbacks:

- increasing system complexity;
- necessity of development special management system.

Main target of HDSS organization is reliable and errorless data storage.

Fault tolerance of HDSS at this work means probability of data block will not loss or contains not corrected errors.

Main advantages of distributed storage concept are:

- high technical redness of HDSS;
- maintaining of HDSS fault tolerance;
- maintaining of data availability;
- flexible management of system;
- maintaining demand efficiency of HDSS;
- technical and technology reserve for future research.

Considering ERS as information system it is obvious that it main function determinates as issuance of reliable information about the objects of observation to the end user.

So, reliable and errorless data storage at the heterogeneous HDSS of ESR is carried out through:

1) Dividing data objects to the regions according to the importance index.

2) Application different constructions of regenerating codes for different data.

3) Data placement according to the importance index.

4) Application of error corrections codes depends of data placement and importance.

Using such complex multi-level approach can ensure high HDSS fault-tolerance.

Math formulation of problem

Technical analysis of modern storage systems and technologies condition shows that storage process characterized by couple of properties witch borders reliability of data storage at current ERS concept.

First, designing HDSS from homogeneous memory modules is hardly to realized for all kinds of memory (except NAND-FLASH) because of huge mass of result system. Besides that it is hardly to make a difference for different importance index data.

Second, storage with replication of data leads to significant increasing of data redundancy and respectively to decreasing efficient storage capacity.

Third, evaluation fault tolerance of HDSS performed for systems “at average” and not takes in to account importance index of data.

Fourth, for onboard data processing needed large volumes of reliable memory with limitless (or very large) count cycles of write and read before failure.

Current approaches to HDSS design based on replication of data and using non-reliable NAND FLASH can not ensure such opportunities regardless of the level of scaling and improvement of technological processes due to constructive limitations.

Five, using of non NAND FLASH MU will allow save energy resources and use them for data transformation for more reliable storage and reducing the level of information redundancy.

Task of HDSS design is form from array of available memory modules types and controllers such structure of HDSS that met the fitness criteria. Besides that it is necessary to choose constructions of regenerating and error correct codes and data placement for each importance index of data. Forming storage methods should met criteria of minimal information redundancy. Parameters of HDSS with such storage methods too should met the fitness criteria.

Storage methods should be formed for maximum payload of HDSS (initial storage methods). Storage methods (combination of construction of regenerating and error correction codes and data placement) should be corrected according with system condition. Storage methods may be form for each data obtaining and should lead to maximum of data reliability.

Thus task of designing and application consist of two subtasks. First, subtask is synthesis of HDSS structure and storage methods.

Second, subtask of HDSS management.

In the math form, the problems of constructing and using a heterogeneous storage system can be described as follows.

1. Subtask of HDSS structure synthesis and storage methods.

Given:

1. HDSS operational requirements:

$$\langle D_{\max}, M_{\max}, E_{\max}, V_{\min} \rangle,$$

where
D – HDSS overall dimensions;
M – HDSS mass;
E – HDSS energy consumption;
V – HDSS effective capacity.

2. Requirements for HDSS performance indicators:

$$\langle P_{EL\min}, P_{DNL\min}, U_{R\min}, U_{W\min} \rangle,$$

where
$P_{EL\min}$ – requirements for data errorless probability;
$P_{DNL\min}$ – requirements for data availability probability;
$U_{R\min}$ – requirements for HDSS productivity at read mode;
$U_{W\min}$ – requirements for HDSS productivity at write mode;

3. Structure of data:

$$\langle K, F_k \rangle,$$

where
K – number of importance levels at earth remote sensing data;
F_k – specific capacity of k-th data type.

4. Available set of memory chips:

$$MP = \{mp_j\};$$

5. Available set of regenerating codes: $A = \{a_l\}$, possessing properties:

$$a_l = \langle n_l, k_l \rangle,$$

where
k_l – is number of blocks to which the data is divided;
n_l – number of blocks to which the data is encoded;
$Red(a_l) = \dfrac{n_l}{k_l}$ – information redundancy, introduced by regenerating codes.

6. Error-correcting codes: $C = \{c_n\}$, possessing properties:

$$c_n = \langle n_n, k_n, \mu_n \rangle,$$

where
n_n – length of code block in bytes;
k_n – numbers of information digits in error-correcting code;
μ_n – degree of error-correcting.
$RedC_n = \dfrac{n_n}{k_n}$ – information redundancy, introduced by error-correcting code.

Find:

Structure

$$So = \langle n_{11}, n_{22}, \ldots, n_{JI} \rangle,$$

where
n_{JI} – number of MM j-th type in в i-th SU,
storage methods $Sp = \langle a_l, b, c_n \rangle$, such that:

$$\langle So^*, Sp_k^0(t_{AF})^* \rangle : \hat{Y}_{\langle 4 \rangle}(\langle So, Sp_k(t_{AF}) \rangle) \in \{G_{II}\}$$

$$\hat{Y}_{\langle 4 \rangle} = < P_{ELk}(So, Sp_k), P_{DNLk}(So, Sp_k), U_R(So, Sp_k),$$

$U_W(So, Sp_k), Red(So, Sp_k) >$ – HDSS performance indicator;

$$G_{II} = \{P_{ELk}(So, St, Sp_k) \geq P_{EL\min k} \cap$$

$$\cap P_{DNLk}(So, St, Sp_k) \geq P_{DNL\min k} \cap U_R(So, St, Sp_k) \geq$$

$$\geq U_{R\min} \cap U_W(So, St, Sp_k) \geq U_{W\min}\} \text{ – HDSS fitness criterion.}$$

Under the constraints: HDSS should met requirements for onboard hardware: $Y_{\langle 4 \rangle} \in \{Y_{\langle 4 \rangle}^1\}$, where:

$$Y_{\langle 4 \rangle}^1 = < D(So) \leq D_{\max} \cap M(So) \leq M_{\max} \cap$$

$$\cap E(So) \leq E_{\max} \cap V_{ef}(So, Sp_k) \geq V_{\min} > .$$

2. Subtask of HDDS management when solving target tasks.

Given:

1) HDDS structure СХД: S_o.

2) HDDS state: $SS = \{ss_{IS}\}$.

3) New data: $D_N = \langle V_N, K, F_K \rangle$, where V_N – new data volume.

Find:

Storage methods: $\langle Sp_k(t)^* \rangle$:

$\langle P_{Sk}(So, Ss, Sp_k^*(t)) \arg\max P_{Sk}(So, Ss, Sp_k(t)) \rangle$, under the constraints:

$$\begin{cases} Sp_k^*(t) \in Sp_k^D, \\ Red(So, Ss, Sp_k^*) \leq Red(So, Ss, Sp_k^0). \end{cases}$$

Algorithm of data storage based on combining of the methods of locating and coding non-homogeneous data

Synthesis of the heterogeneous structure occurs under conditions of a priori uncertainty of the functional conditions HDDS. On the one hand, the impact of unfavorable space factors on the most frequently used orbits for spacecraft is well studied and can be accepted as normal operating conditions.

On the other hand, at the spacecraft lifetime cycle, the impact of various unforeseen factors is possible: the effect of a physical collision with other objects (space debris), etc.

According with math formulation of problem Algorithm should solve both subtasks of designing and management of HDDS.

Formation of requirements for HDDS is defined as one of the stages of external design. Based on the research of HDDS systems as an element of onboard aperture, analysis of interaction with other structural elements of the system, necessary technical characteristics are determined and a feasibility study is carried out. The results of the external design stage are formulated in the form of technical requirements for the onboard computing system. In most cases, the requirements for the performance HDDS can be satisfied with a variety of options for the composition and

parameters of the system. In this case, the cost of development and operation, the availability of technological back-up and the used software can become decisive when choosing system options.

At the first stage, the initial data is collected. It consists in a comprehensive analysis of the requirements for the operation conditions of the satellite, analysis of technical requirements for HDDS, as well as available memory options [11].

Also before or at the stage of system-wide design, data are collected on reliability indicators of various drive options. The data can be obtained from the manufacturer on the basis of tests carried out by them or the test data can be produced by the designer in accordance with the approved methodology.

Algorithm of data storage based on combining of the methods of locating and coding non-homogeneous data structurally consist of two parts. Each part solving corresponding subtask.

For first subtask of HDSS structure synthesis and initial storage methods it is necessary to execute the following sequence of steps.

Step 1. Based on $P_{EL\min}$ and available for implementation constructions of error-correcting codes $C = \{c_n\}$ for each of memory modules – mp_{cj} a choice is made of such constructions of error-correcting codes which provide the required probability of non-avoidable errors at k-th type data set:

$$P_{ELkcj}(c_{kcj}) \le P_{EL\min k}.$$

In this case, the minimum information redundancy conditions must be satisfied:

$$\langle c_{kcj}\,* \rangle = \arg\min\{red(c_{kcj})\},$$

where

$red\ (c_{kcj})$ – information redundancy introduced by the error-correcting code used for the k-th data type in the j-th type MM in combination with the c-th MC.

Step 2. Based on chosen types and construction of error-correcting codes for each data type and combination of MM and MC calculating effective data capacity: $Vef_{kcj} = v_{cj} \cdot red(c_{kcj})$.

Step 3. For each data type forms many combinations of substructures and corresponding construction of error-correcting codes. It should be noted that when forming substructures, the assumption that within the substructure, the data is evenly distributed, i. e. $b_{kcj} = \left\langle 1, \dfrac{V_{kcj}}{\min\limits_{cj} V_{kcj}} \right\rangle$. This assumption is possible because of during HDDS application data placement is not even and thus probability of data loss is much less.

During forming substructure and storage methods solving following optimization problem:

$$\{\langle So_k^*, A_k^* \rangle\} = P_{DNLk}(So_k, A_k) \ge P_{DNL\min k},$$

where

So_k – substructure of HDDS.

This problem can be solved by a known method of borders and brunches.

Step 4. Forming of full structure of HDDS can be find by solving classic Multiple-choice knapsack problem:

1) For each data type k set of vectors enumerates from 1 to j.

2) Find: $\max \sum_{i=1}^{K} \sum_{j \in N(So_k)} Vef_{ij} \cdot x_{ij}$, with restrictions:

$$\begin{cases} \sum\limits_{i=1}^{K} \sum\limits_{j \in N(So_k)} m_{ij} \cdot x_{ij} \le M_{\max}; \\ \sum\limits_{i=1}^{K} \sum\limits_{j \in N(So_k)} g_{ij} \cdot x_{ij} \le G_{\max}; \\ \sum\limits_{i=1}^{K} \sum\limits_{j \in N(So_k)} e_{ij} \cdot x_{ij} \le E_{\max}; \\ \sum\limits_{j \in N(So_k)} x_{ij} = 1; \\ i = 1 \ldots K; \\ x_{ij} \in \{0,1\}; \\ j \in N(So_k). \end{cases}$$

Problem can be solved by dynamic programming method. As result forms full HDDS structure that meet operational requirements for system and has maximum effective capacity.

Step 5. To determine the suitability HDDS structure it is necessary to carry out simulation modeling.

According to the results of modeling, the decision on the compliance of the functioning is made – G_{II}.

If following conditions are true:

$$\langle So^*, Sp_k^* \rangle : \hat{Y}_4(\langle So, Sp_k \rangle) \in \{G_{II}\}$$

HDSS structure synthesis and initial storage methods is considered complete. In the case of failing to meet the specified requirements, make the necessary adjustments and the implementation of the synthesis algorithm is repeated.

Main of function time there are exist sufficient information reserve of HDDS capacity because of ERS systems doses not work with full payload and has regular communication sessions. Fault tolerance of HDDS can be enhanced by using this reserve.

At the same time, certain restctions on the quality of operation remains. HDDS should meet fitness criterion.

Optimal operating mode (in terms of HDDS fault tolerance) should meet HDSS optimality criterion:

$$O_{II} = \{< P_{ELk}(So, Ss, Sp_K^*) = \arg\max P_{ELk}(So, Ss, Sp_K) \cap$$
$$\cap P_{DNLk}(So, Ss, Sp_K^*) = \arg\max P_{DNLk}(So, Ss, Sp_K) \cap$$
$$\cap T_R(So, Ss, Sp_K^*) \le T_{R\max} \cap T_W(So, Ss, Sp_K^*) \le T_{W\max} \cap$$
$$\cap Red(So, Ss, Sp_K^*) \le Red(So, Ss, Sp_k^0) >$$

In such way tasks of HDDS management may be formulated as following: it is necessary to find such data placement at HDDS structure and construction of regenerating codes that makes meet HDSS optimality criterion.

For the subtask of HDDS management when solving target tasks it is necessary to execute the following sequence of steps.

Step 1. When new data received by HDDS functional control is carried out.

According to the results of functional control, the HDDS state vector is formed: $Ss_S = \langle V_S, Vu_S \rangle$.

Step 2. New arrived data object $D_N = \langle V_N, K, F_K \rangle$ divided in to k regions according to data type.

Step 3. Based on the initial storage methods – $Sp_k{}^0$ computes permissible redundancy – $Red\ (Sp_k{}^0)$.

Step 4. Based on the HDDS condition and permissible redundancy for regenerating codes – $Red\ (a_k{}^0)$ selected regenerating codes construction:

$$\begin{cases} k_k = \dfrac{V_k}{\min\limits_s (V_s - Vu_s) / Red\ (a_k{}^0)}; \\ n_k = Red\ (a_k{}^0) \cdot k_k, \end{cases}$$

where

$Red\ (a_k{}^0)$ – information redundancy.

Step 5. The next step is to find the placement of data blocks by structure. The task is formulated as follows: it is necessary to find the bijection $b_k : n_k \to s$, such that: $P_{DNLk}(b_k) \to \max$.

This task is proposed to solve the Cauchy Method.

Step 5. According with chosen data placement – b_k, and restrictions on the information redundancy – $Red\ (c_k{}^0)$ calculating constructions of error correction codes:

$$\langle c_{ks}{}^* \rangle = \arg\max P_{ELk}(c_{ks}),$$

where

c_{ks} – chosen type and construction of error correction codes for k-th data type and s-th MM.

Step 7. Based on the way of storage for k-th data type $Sp_k = \langle a_k, b_k, c_k \rangle$ corrected system statement Ss_s for further calculations of storage methods.

Step 8. Similarly, can be chosen storage methods for other data types.

Step 9. Simulation of storage system functioning in accordance with the operating model is carried out.

In case of meeting the requirements:

$$\langle So^*, Sp_k{}^* \rangle : \hat{Y}_4 (\langle So, Sp_k \rangle) \in \{O_{II}\}$$

data writes with calculated storage methods. In the case of failing to meet the specified requirements, it is the necessary to make specified adjustments and repeat an algorithm.

Conclusion

Thus, based on the formulation of problems of designing and application HDDS, they can be attributed to the class of combinatorial optimization problems. There are known ways to solve this type of problem. However, it is worth noting a few features.

Thus problem of searching of HDDS optimal structure and storage methods for full payload can be solved by different ways, including by a full search method.

However, this is unacceptable for the HDDS application, since the choice of the method of coding and placement of data blocks for storage must be made promptly and decided on the control device storage, which has a relatively low performance.

In this regard to solve it, a combined method for selecting the type of regenerating encoding and placing data blocks on the storage is proposed.

The choice of error correction codes based on the type of data block and type of MM on which this block assigned.

The proposed approach to providing fault tolerance of remote sensing information processing system, based on design and application HDDS allows flexible storage process management. Besides that it allows to correct HDDS properties.

This takes into account the initial importance of the remote sensing data areas, the structure and statement of HDDS as well as compliance of the operating characteristics of HDDS requirements.

References

1. Petrov A. G., Ulanova A. V., Chumakov A. I., Vasiliev A. L. The study of information loss in the circuits of flash memory in active and passive modes, when exposed to ionizing [Issledovaniya poteri informatsii v mikroskhemakh flesh-pamyati v aktivnom i passivnom rezhimakh pri ioniziruyushchem vozdeystvii]. *Radiation resistance of electronic systems*. Moscow, 2014, Vol. 17, pp. 175-176. (In Russ.)

2. Savinyh V. P., Solomatin V. A. Electro-optical system for remote sensing [Optiko-elektronnye sistemy distantsionnogo zondiro-vaniya]. Moscow, Nedra, 1996. 315 p. (In Russ.)

3. The concept of development of Russian space system of remote sensing for the period up to 2025 [Kontseptsiya razvitiya rossiyskoy kosmicheskoy sistemy distantsionnogo zondirovaniya zemli na period do 2025 goda]. Moscow, Federal space agency, 2006. 72 p. (In Russ.)

4. Zaharov I. V., Kremez G. V., Maksimov V. A. Construction of distributed onboard storage systems for remote Earth-sensing satellites [Postroenie raspredelennykh zapominayushchikh ustroystv bortovykh vychislitel'nykh sistem kosmicheskikh apparatov distantsionnogo zondirovaniya zemli]. *Proc. Mozhaisky Military Aerospace Acad. [Trudy voenno-kosmicheskoy akademii imeni A. F. Mozhayskogo]*, 2016, Is. 652, pp. 160-166. (In Russ.)

5. Denisov A., Demin A., Letunovskiy A. Optical Digital Systems and Complexes for Space Applications. *Intellectual Technologies on Transport*, 2016, no. 1, pp. 16-22.

6. Dimakis A. G., Ramchandran K., Wu Y., Suh C. A Survey on Network Codes for Distributed Storage, *CSIT*, 2010, pp. 12-18.

7. Wu Y., Dimakis A. G., Ramchandran K. Deterministic Regenerating Codes for Distributed Storage. *Allerton Conf. Control Computing and Communication*, 2007, pp. 236-242.

8. Weatherspoon H., Kubiatowicz J. Erasure coding vs. replication: A quantitative comparison. *Proc. First Int. Workshop on Peer-to-Peer Systems (IPTPS 2015)*, 2015, pp. 65-78.

9. Molette P., Cougnet C., Saint-Aubert Ph., Young R. W., Helas D. Technical and Economical Comparison Between a Modular Geostationary Space Platform and a Cluster of Satellites. *Acta Astronautica (Pergamon Press Ltd.)*, 1984, no. 12 (11), pp. 771-784.

10. Herz E. EO and SAR Constellation Imagery Collection Planning. *Proc. 14th Int. Conf. Space Operations, SpaceOps AIAA*, 2014. 214 p.

11. Haeusser G. B., Osuna A., Bosman Ch., Jahn D., Tarella G. J. ILM Library: Information Lifecycle Management Best Practices. IBM, 2016. 256 p.

Алгоритм хранения информации на основе комбинирования способов размещения и кодирования неоднородных по важности данных

Максимов В. А.
Военно-космическая академия им. А. Ф. Можайского
Санкт-Петербург, Россия
falcon225@yandex.ru

Аннотация. **Предложен подход к построению и применению систем хранения данных перспективных космических аппаратов на основе гетерогенных запоминающих устройств с целью повышения отказоустойчивости функционирования систем обработки информации. Применение гетерогенных накопителей допускает более гибкое управление способами хранения, что позволяет снизить степень информационной избыточности всех видов в системе, а также гибко регулировать эксплуатационные параметры системы. Новизна работы заключается в учете неоднородности важности хранимых данных.**

Ключевые слова: **система хранения данных, гетерогенность, распределенность, отказоустойчивость, безошибочность, дистанционное зондирование Земли, восстанавливающее кодирование.**

Литература

1. Петров А. Г., Уланова А. В., Чумаков А. И., Васильев А. Л. Исследования потери информации в микросхемах флэш-памяти в активном и пассивном режимах при ионизирующем воздействии // Радиационная стойкость электронных систем – «Стойкость-2014». – М., 2014. Вып. 17. С. 175-176.

2. Савиных В. П., Соломатин В. А. Оптико-электронные системы дистанционного зондирования. – М.: Недра, 1996. 315 с.

3. Концепция развития российской космической системы дистанционного зондирования земли на период до 2025 года. – М.: Федеральное космическое агентство, 2006. 72 с.

4. Захаров И. В., Кремез Г. В., Максимов В. А. Построение распределенных запоминающих устройств бортовых вычислительных систем космических аппаратов дистанционного зондирования земли // Тр. ВКА им. А. Ф. Можайского. 2016. Вып. 652. С. 160-166.

5. Denisov A., Demin A., Letunovskiy A. Optical Digital Systems and Complexes for Space Applications // Intellectual Technologies on Transport. 2016. № 1. P. 16-22.

6. Dimakis A. G., Ramchandran K., Wu Y., Suh C. A Survey on Network Codes for Distributed Storage // CSIT. 2010. P. 12-18.

7. Wu Y., Dimakis A. G., Ramchandran K. Deterministic Regenerating Codes for Distributed Storage // *Allerton Conf. Control Computing and Communication.* 2007. P. 236-242.

8. Weatherspoon H., Kubiatowicz J. Erasure coding vs. replication: A quantitative comparison // Proc. First Int. Workshop on Peer-to-Peer Systems (IPTPS 2015). 2015. P. 65-78.

9. Haeusser G. B., Osuna A., Bosman Ch., Jahn D., Tarella G. J. ILM Library: Information Lifecycle Management Best Practices. – IBM, 2016. 256 p.

10. Herz E. EO and SAR Constellation Imagery Collection Planning // Proc. 14-th Int. Conf. Space Operations, SpaceOps AIAA. 2014. 214 p.

11. Molette P., Cougnet C., Saint-Aubert Ph., Young R. W., Helas D. Technical and Economical Comparison Between a Modular Geostationary Space Platform and a Cluster of Satellites // Acta Astronautica (Pergamon Press Ltd.). 1984. № 12 (11). P. 771–784.

Сравнительный анализ популярных систем мониторинга сетевого оборудования, распространяемых по лицензии GPL

Шардаков К. С.
Петербургский государственный университет
путей сообщения Императора Александра I
Санкт-Петербург, Россия
megashok2010@gmail.com

Аннотация. **Рассмотрены наиболее популярные системы мониторинга сетевого оборудования, распространяемые по лицензии GPL – Cacti, Nagios, Zabbix. Выявлены сходства и различия между ними. Описаны архитектура и основные компоненты системы мониторинга. Смоделировано рабочее полнофункциональное состояние каждой рассматриваемой системы. Приведены критические минусы сравниваемых систем мониторинга. Отмечено, что из-за выявленных недостатков и необходимости использования большого количества сторонних плагинов для Cacti и Nagios, а также из-за сложности масштабирования этих систем наилучшим выбором для мониторинга большого количества метрик признана система Zabbix.**

Ключевые слова: **мониторинг, Cacti, Nagios, Zabbix, Docker, система оповещений, метрики, уведомления, GPL, RRDTool.**

Введение

Автоматизированные системы мониторинга играют важную роль в жизни современного общества. Мониторинг – это непрерывный процесс наблюдения и регистрации параметров объекта по сравнению с заданными критериями. В ряде отраслей данные собираются и накапливаются очень интенсивно. Основные сведения о системах мониторинга в различных отраслях, примеры отдельных систем мониторинга и их сравнительные характеристики приводятся в работах [1–5].

Из-за тенденции роста сетей передачи данных повышаются требования к системам мониторинга этих сетей. При расширении сети увеличивается и количество составляющих её объектов, нуждающихся в мониторинге, а следовательно, и нагрузка на систему мониторинга. Такая ситуация влечет за собой замедление реакции на аварийные события в сети, ведет к деградации сетевых сервисов и служб. Для предотвращения этих ситуаций необходим правильный подход к выбору системы мониторинга. В работе рассматриваются нескольких наиболее популярных систем мониторинга с открытым исходным кодом: Cacti, Nagios, Zabbix.

На рис. 1 представлена статистика запросов по ключевым словам на популярном поисковом ресурсе google.com с 2004 г. Как видно из рис. 1, интерес к некогда популярным системам Cacti и Nagios постепенно угасает, в то время как количество запросов с ключевым словом Zabbix растёт и имеет постоянную положительную тенденцию.

Цель работы – выявить наименее ресурсоемкую и наиболее гибко настраиваемую систему мониторинга из имеющихся, распространяемых по лицензии GPL, соответствующую основным требованиям:

- прозрачная масштабируемость;
- мониторинг нескольких тысяч устройств и более 1 000 000 метрик с использованием протокола SNMP [6] (Simple Network Management Protocol);
- обработка нескольких тысяч уведомлений ежедневно.

Основные компоненты системы мотиноринга

Любая система мониторинга, в том числе мониторинга сетевых устройств, – это сложная информационная система, включающая в себя:

- метрики сетевых устройств (CPU, температура, доступность устройств, потери пакетов, ошибки на интерфейсах, доступная полоса пропускания и прочие) – критически важные параметры, за значениями которых необходимо вести наблюдение;
- мониторинг – процесс сбора, агрегирования и анализа метрик для улучшения понимания характеристик и поведения компонентов системы. В этот пункт входит также визуализация собранных данных по метрикам в различные графики, диаграммы, гистограммы [1, 2];
- систему оповещений – не менее важный компонент, поскольку выполняет действия на основе изменений значений наблюдаемых метрик. При достижении критического порога значения метрики может попытаться самостоятельно

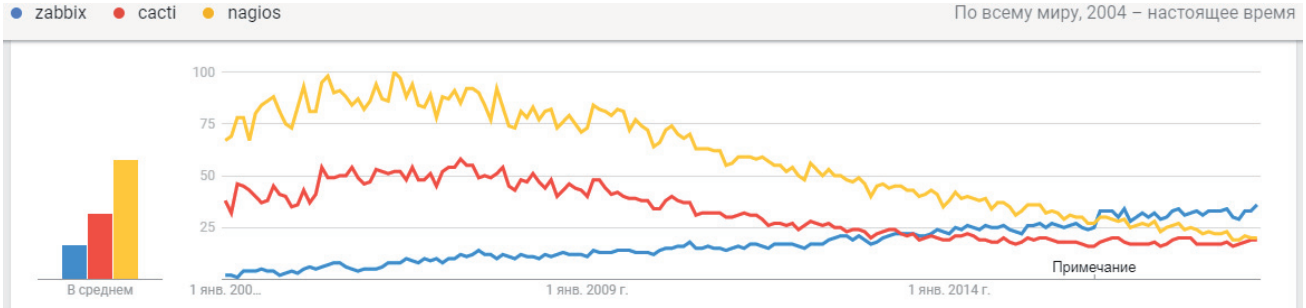

Рис. 1. Статистика запросов в поисковой системе google.com

исправить проблему по заготовленному сценарию или отправить оповещение ответственному лицу средствами SMS, электронной почты и т. д. [1, 2].

Исследование возможностей систем мониторинга

Cacti

Cacti обладает высокой скоростью развертывания при малом количестве объектов для мониторинга. Добавление в мониторинг устройств к стандартным шаблонам происходит быстро и легко [7]. Но процесс добавления устройств с метриками, не предусмотренными стандартным набором функций, трудоёмкий. В качестве хранилища данных Cacti использует RRD (Round-Robin Database), а в качестве визуализатора для построения графиков из собранных данных – утилиту RRDTool [5, 8]. Это является недостатком, поскольку RRD усредняет данные, и невозможно сказать, каково было точное значение параметров несколько месяцев назад. Отсутствует система оповещений, основные функции – сбор метрик с устройств в RRD и их последующая визуализация.

На рис. 2 представлен пример графика утилизации пропускной способности сетевого интерфейса коммутатора, построенный с помощью Cacti.

Nagios

Nagios поддерживает огромное количество плагинов для всевозможных задач, но его слабое место – малая отказоустойчивость. Масштабирование – сложная процедура в Nagios, но возможна при использовании сторонних плагинов [7]. Каждый запущенный плагин является отдельным запущенным процессом в операционной системе. Nagios не имеет встроенных средств визуализации, их тоже необходимо подключать в виде плагинов [9]. Конфигурация изменяется посредством изменения файла конфигурации, но новая конфигурация применяется только при перезапуске службы Nagios. Иными словами, при добавлении нового устройства в систему мониторинга приходится перезапускать службу мониторинга для применения конфигурации. Согласно исследованиям компании Яндекс, при большом количестве объектов мониторинга (несколько тысяч) процедура перезапуска может занять 15–20 минут, в течение которых не снимаются метрики с объектов мониторинга [6]. Это означает, что в указанном промежутке времени система мониторинга не функционирует. В качестве хранилища данных Nagios, как и Cacti, использует RRD [5, 10].

На рис. 3 представлен пример графика утилизации пропускной способности сетевого интерфейса коммутатора, построенный с помощью Nagios.

Zabbix

Zabbix позволяет изменять конфигурацию системы несколькими способами: через веб-интерфейс и посредством API (Application Program Interface) [11]. Конфигурация хранится в базе данных, это позволяет применять её «на лету», в отличие от Nagios [12]. За счет распределенной архитектуры и использования прокси легко масштабируется, поскольку база данных может храниться на одном сервере, сервер приложения – на втором, а фронтенд – на третьем [13]. Также есть возможность распределять нагрузку по нескольким прокси-серверам, которые делят очереди опросов объектов мониторинга между собой, тем самым высвобождая ресурсы основного сервера приложения [4]. Мощная система

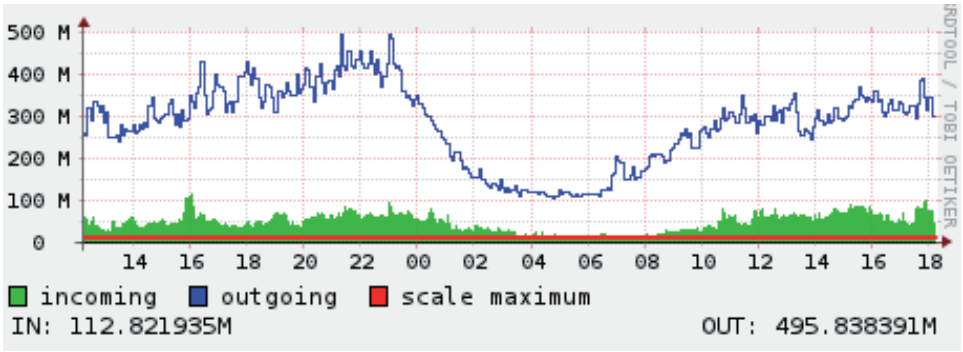

Рис. 2. График утилизации пропускной способности, построенный с помощью Cacti

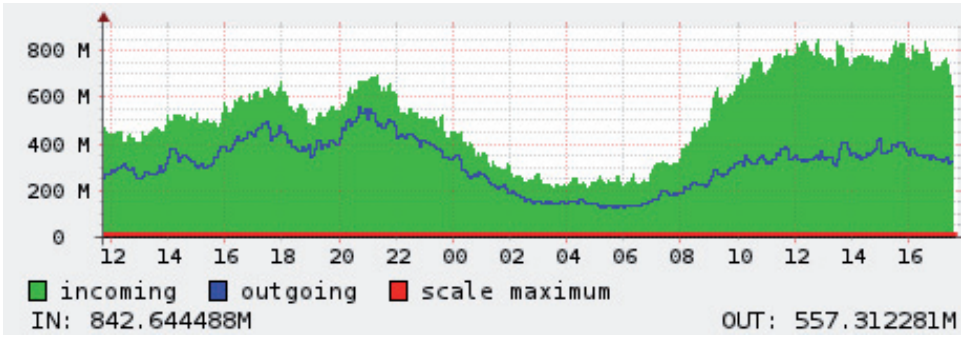

Рис. 3. График утилизации пропускной способности, построенный с помощью Nagios

оповещений позволяет использовать оповещения и автоматические действия на основе заготовленных сценариев при достижении пороговых значений наблюдаемых метрик [14, 15]. В версии 3.4 появилась возможность прогнозировать дальнейшее поведение метрик на основе уже полученных значений [4, 11]. Из недостатков стоит отметить ограниченный функционал для работы с визуализацией данных и большую задержку при отображении большого количества графиков на одном экране.

На рис. 4 представлен пример графика утилизации пропускной способности сетевого интерфейса коммутатора, построенный с помощью Zabbix.

В табл. 1 представлены основные возможности сравниваемых систем мониторинга.

Моделирование системы мониторинга и тестирование

Ещё одним важным критерием является потребление ресурсов каждой системой. В рамках статьи собран тестовый стенд с развернутыми в нем сравниваемыми системами мониторинга. В качестве сервера выступает устройство с процессором семейства Intel Core 2 Duo, HDD 7200 RPM, 8Gb RAM, ОС Debian 9.3 «Stretch». Все тестовые системы мониторинга развернуты на сервере в Docker-контейнерах. Единовременно запущены контейнеры, связанные только с одной системой мониторинга. Наблюдаются 100 хостов по 410 метрик на каждый. Интервал опроса 60 секунд.

В табл. 2 отображены значения, полученные в ходе тестирования исследуемых систем мониторинга.

Заключение

В результате анализа сравниваемых систем мониторинга выявлены следующие критические минусы:

- нет системы оповещений в Cacti;
- при использовании RRD в Cacti и Nagios теряется детализация старых данных;
- конфигурация в Nagios изменяется посредством изменения файла конфигурации, но новая конфигурация применяется только при перезапуске службы Nagios, что при большом количестве собираемых метрик может занимать несколько десятков минут, а следовательно, система не будет функционировать в это время.

Zabbix имеет менее критические недостатки, связанные с визуализацией данных, что слабо влияет на основные функциональные возможности системы.

Как видно из табл. 2, наименее ресурсоемкой системой является Cacti, наиболее ресурсоемкой – Nagios. Но из-за описанных недостатков и необходимости использования большого количества сторонних плагинов для Cacti и Nagios, а также из-за сложности масштабирования этих систем наи-

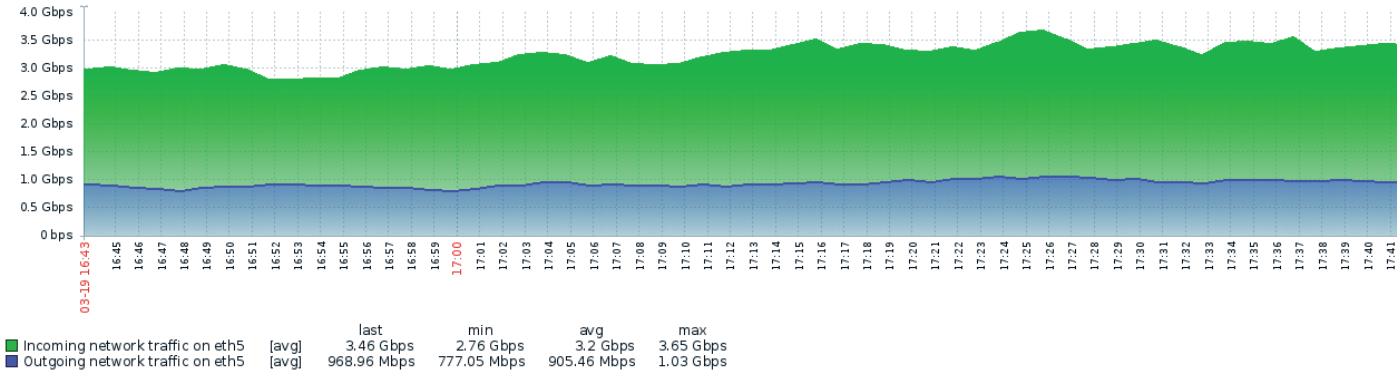

Рис. 4. График утилизации пропускной способности, построенный с помощью Zabbix

ТАБЛИЦА 1. Сводная таблица возможностей сравниваемых систем

Система	Атоматическое обнаружение	Поддержка SNMP	Плагины	Система оповещений	Способ хранения данных	Лицензия
Cacti	Через плагин	Да	Да	Через плагин	RRDTool	GPL
Nagios	Через плагин	Через плагин	Да	Да	RRDTool, MySQL через плагин	GPL
Zabbix	Да	Да	Да	Да	Oracle, MySQL, PostgreSQL, IBM DB2, SQLite	GPL

ТАБЛИЦА 2. Потребление ресурсов системами

Система	CPU, %	RAM, %
Cacti	10	22
Nagios	14	25
Zabbix	12	25

лучшим выбором для мониторинга большого количества метрик признан Zabbix.

Литература

1. Носкова А. И., Токранова М. В. Обзор автоматизированных систем мониторинга // Интеллектуальные технологии на транспорте. 2017. № 1. С. 42-47.

2. Основы мониторинга и сбора метрик. URL: https://www.8host.com/blog/osnovy-monitoringa-i-sbora-metrik.

3. Шмелев В. В. Метод мониторинга технологических процессов на основе структурно-логического подхода // Интеллектуальные технологии на транспорте. 2017. № 2. С. 5-14.

4. Jones D. Creating Unified IT Monitoring and Management in Your Environment. – Realtime Publishers, 2012. – 92 p.

5. Croll A. Complete Web Monitoring. – Sean Power: O'Reilly Media, 2009. 672 p.

6. Vacche A. D., Lee S. K. Zabbix Mastering. – Packt Publishing Ltd., 2013. 358 p.

7. Линикова О. Е. Мониторинг серверного оборудования и приложений. – Екатеринбург, 2014. 123 с.

8. Graphite – как построить миллион графиков. URL: https://events.yandex.ru/lib/talks/1122.

9. Владышев А. Zabbix и миллионы метрик: наилучший опыт масштабного мониторинга. URL: http://www.highload.ru/2015/abstracts/1965.html.

10. Анализ систем централизованного мониторинга с открытым исходным кодом мониторинга. URL: https://sites.google.com/site/teachingandresearchwork/sravnenie-sredstv-monitoringa.

11. Пять ключевых функций систем мониторинга производительности сети. URL: https://networkguru.ru/piat-cliuchevykh-funktcii-sistem-monitoringa-proizvoditelnosti-seti.

12. Использование Zabbix для мониторинга критических систем. URL: https://xakep.ru/2014/08/13/using-zabbix.

13. Сетевая и серверная статистика с использованием Cacti. URL: https://nsrc.org/workshops/2014/caren-nsrc-dante/raw-attachment/wiki/Agenda/cacti-from-packages-vRU.pdf.

14. Мониторинг компьютерной сети или тестирование систем мониторинга. URL: https://united.net.ua/?id=usage&text=3.

15. Nagios vs Zabbix – сравнение систем мониторинга сети. URL: http://amigosteam.ru/blog/item/12-nagios-vs-zabbix.

Comparative Analysis of the Popular Monitoring Systems for Network Equipment Distributed Under the GPL License

Shardakov K. S.
Emperor Alexander I St. Petersburg State Transport University
St. Petersburg, Russia
megashok2010@gmail.com

Abstract. **The most popular network equipment monitoring systems distributed under license GPL – Cacti, Nagios, Zabbix are considered. Similarities and differences between them are revealed. The architecture and main components of the monitoring system are described. The working full-functional state of each considered system is modeled. Critical disadvantages of the compared monitoring systems are given. It is noted that due to the identified shortcomings and the need to use a large number of third-party plugins for Cacti and Nagios, as well as due to the complexity of scaling these systems, Zabbix system is recognized as the best choice for monitoring a large number of metrics.**

Keywords: **monitoring, Cacti, Nagios, Zabbix, Docker, alert manager, metrics, notifications, GPL, RRDTool.**

References

1. Noskova A. I., Tokranova M. V. Overview of Automated Monitoring Systems, *Intellectual Technologies on Transport*, 2017, no. 1, pp. 42-47. (In Russ.)

2. Basics of monitoring and collecting metrics. Available at: https://www.8host .com/blog/osnovy-monitoringa-i-sbora-metrik. (In Russ.)

3. Shmelev V. V. Method for Monitoring the Technological Processes in the Aerospace Industry on the Basis of Structural and Logical Approach, *Intellectual Technologies on Transport*, 2017, no. 2, pp. 5-14. (In Russ.)

4. Jones D. Creating Unified IT Monitoring and Management in Your Environment. Realtime Publishers, 2012. 92 p.

5. Croll A. Complete Web Monitoring. Sean Power, O'Reilly Media, 2009. 672 p.

6. Vacche A. D., Lee S. K. Zabbix Mastering. Packt Publishing Ltd., 2013. 358 p.

7. Linikova O. E., Monitoring of server hardware and applications. Master's thesis. Ural Federal University named after the first President of Russia B. N. Yeltsin. – Ekaterinburg, 2014. – 123 c. (In Russ.)

8. Graphite – how to build a million graphs. Available at: https://events.yandex.ru/lib/talks/1122.

9. Vladyshev A. Zabbix and millions of metrics: the best experience of large-scale monitoring. Available at: http://www.highload.ru/2015/abstracts/1965.html. (In Russ.)

10. Analysis of centralized monitoring systems with open source monitoring. Available at: https://sites.google.com/site/teachingandresearchwork/sravnenie-sredstv-monitoringa. (In Russ.)

11. Five key functions of network performance monitoring systems. Available at: https://networkguru.ru/piat-cliuchevykh-funktcii-sistem-monitoringa-proizvoditelnosti-seti. (In Russ.)

12. Using Zabbix for monitoring of critical systems. Available at: https://xakep.ru/2014/08/13/using-zabbix. (In Russ.)

13. Network and server statistics using Cacti. Available at: https://nsrc.org/workshops/2014/caren-nsrc-dante/raw-attachment/wiki/Agenda/cacti-from-packages-vRU.pdf. (In Russ.)

14. Monitoring a computer network or testing monitoring systems. Available at: https://united.net.ua/?id=usage&text=3.

15. Nagios vs Zabbix – Comparison of network monitoring system. Available at: http://amigosteam.ru/blog/item/12-nagios-vs-zabbix. (In Russ.)