Инженерный вестник Дона, №9 (2020)
ivdon.ru/ru/magazine/archive/n9y2020/6612

Увеличение надежности схем порогового разделения данных

Ю.Н. Кочеров, Э.Е. Тихонов, Д.В. Самойленко

Невинномысский Технологический Институт (филиал) федерального государственного автономного образовательного учреждения высшего образования «Северо-Кавказский федеральный университет»

Аннотация: В работе рассмотрено применение схем порогового разделения данных, как одного из перспективных направлений в области защиты как частной, так и коммерческой информации. В таких схемах части информации распределяются среди пространственно распределенных хранилищ данных, а восстановление информации возможно с использованием не менее k из n частей. Главным недостатком таких схем является то, что они могут быть легко скомпрометированы, например, при утере какой-либо части данных, либо при фальсификации данных злоумышленником невозможно локализировать ошибку, а, следовательно, восстановить исходную информацию. Для решения указанного недостатка предлагается использовать схемы, основанные на системе остаточных классов (СОК). Использование алгоритмов разделения данных, основанных на СОК, приведет к снижению вычислительной сложности, и, следовательно, снижению нагрузки на среды передачи данных. Другим достоинством СОК является то, что они обладают корректирующими свойствами. Пропорциональное увеличение количества информационных и избыточных модулей при использовании классических схем разделения данных приведет к резкому снижению надежности всей схемы обмена данными. Для решения указанного недостатка предлагается разделять информацию на группы, а после этого распределять ее среди участников схемы обмена данных.
Ключевые слова: пороговое разделение данных, система остаточных классов, надежность схем обмена данными.

Введение

Схемы разделения данных применяются в случае, когда необходимо разделить исходную информацию на n частей и распространить ее среди территориально удаленных участников. Восстановление информации возможно только в том случае, когда будут собраны все ее части.

Частным случаем схем разделения данных являются схемы с пороговым разделением данных. В таких схемах информацию можно восстановить только в том случае, когда участник схемы обмена сообщения соберет k из n частей данных. Родоначальниками таких схем были Адди Шамир и Джордж Блэкли.

Область применения таких схем это: распределенный доступ к ресурсам; распределенные вычисления; безопасное хранение секретных

ключей. Также такие схемы позволяют ограничивать доступ к закрытым ресурсам участников в случае утраты к ним доверия.

Схемы разделения данных подвержены как пассивным, так и активным атакам.

Отсутствие внедренных в схемы разделения данных алгоритмов предупреждения о некорректном восстановлении исходной информации может привести к достижению злоумышленником своих целей. Отсюда следует, что для достижения достоверности восстановления данных необходимо применять алгоритмы, в работе которых предусмотрена возможность обнаружения ошибок и их коррекции. Среди таких алгоритмов выделяют алгоритмы, основанные на СОК [1,2].

СОК широко изучена в теории чисел, она активно используется при цифровой обработке сигналов, графической обработке изображений, в кодах с обнаружением и коррекцией ошибок, и в криптографических системах.

Методы разделения данных и их восстановление по остаткам

Чтобы методы разделения данных, основанные на СОК, обладали возможностью локализации и коррекции ошибок, требуется два набора множеств: информационных и контрольных.

В информационное множество включаются числа, составляющие значение разделяемой величины, контрольные избыточные числа, вводимые для локализации и коррекции ошибок при передаче данных.

СОК позволяет использовать единый помехоустойчивый код для локализации и коррекции ошибок, возникающих при передаче данных по каналам связи и ее преобразования в системах обработки информации.

Числа, которые преобразуются в СОК, должны удовлетворять следующему условию: $\prod_{i=1}^{n-1} p_i < A \le \prod_{i=1}^{n} p_i$ где n - общее количество оснований СОК, A - значение, преобразуемое в СОК.

Пример 1. Рассмотрим представление числа в СОК. Дано число $A=66$ и система оснований $p_1=2$, $p_2=5$, $p_3=7$, $p_4=13$. Проверим, удовлетворяет ли A условию $\prod_{i=1}^{n-1} p_i < A \le \prod_{i=1}^{n} p_i$. Так как $A<P$, $P=\prod_{i=1}^{n} p_i = p_1 \cdot p_2 \cdot p_3 \cdot p_4 = 2 \cdot 5 \cdot 7 \cdot 13 = 910$ то условие соблюдены.

Таким образом, число $A=66$ можно представить, как остатки [3] от деления на рассмотренную систему оснований СОК:

$$\alpha_1 = A \bmod p_1 = 66 \bmod 2 = 0;$$

$$\alpha_2 = A \bmod p_2 = 66 \bmod 5 = 1;$$

$$\alpha_3 = A \bmod p_3 = 66 \bmod 7 = 3;$$

$$\alpha_4 = A \bmod p_4 = 66 \bmod 13 = 1.$$

К методам разделения данных, основанных на СОК, относятся такие, как схема Асмута-Блума [4], и схема Миньотта [5].

В основе работы схемы Миньотта лежит применение Китайской теоремы об остатках (КТО) [6,7]. КТО позволяет участникам схемы обмена данными, имеющим некоторое количество частей информации, восстановить исходные данные, причём единственным образом.

В схеме разделения данных Миньотта применяются специальные числовые ряды, названные последовательностями Миньотта.

Пусть n – целое число такое что $n \ge 2$ и $2 \le k \le n$. Тогда (n,k) - это последовательность Миньотта – ряды взаимно простых принадлежащих N таких, что: $p_1 < p_2 \ldots < p_n$ и $\prod_{i=0}^{k-2} p_{n-i} < \prod_{i=1}^{k} p_i$.

Недостатком такой схемы является то, что невозможно разделить весь диапазон чисел $[0,P)$, также такая схема является неидеальной и несовершенной.

Схема Асмута-Блума исключает указанные недостатки. Она описывается следующим образом:

Пусть A – некоторая информация, которую необходимо разделить среди участников группы. Выбирается простое число q, больше A.

Из условий:

$$\forall_i : p_i > q;$$

$$\forall_i : p_i < p_i + 1;$$

$$p_1 \cdot p_2 \cdot \ldots \cdot p_k > q \cdot p_{n-m+2} \cdot p_{n-m+3} \cdot \ldots \cdot p_n.$$

Проводится выбор ряда простых чисел $p_1, p_2, p_3, \ldots, p_n$. Далее генерируется случайное число r и вычисляется $A' = A + q \cdot r$. Затем вычисляются доли информации $\alpha_i = A \bmod p_i$.

При восстановлении чисел из СОК в позиционную систему счисления (ПСС) могут применяться различные методы, такие как: методы основанные на КТО; метод Гарнера (метод, основанный на обобщенной полиадической системе счисления (ОПСС)) или метод совместного использования КТО и ОПСС.

Восстановление исходной информации с применением КТО основывается на вычислении значения по формуле:

$$A = \left| \sum_{i-1}^{n} \alpha_i b_i \right|_P$$

где: $\alpha_i = A(\bmod p_i)$; b_i -ортогональные базисы, рассчитываемые по формуле $b_i = \frac{m_i P}{p_i}$, m_i - положительные, целые числа называемые весами базиса, их, определяют из приближения $P_i m_i = 1(\bmod p_i)$.

Недостаток этого метода заключается в том, что для преобразования из системы СОК в позиционную систему счисления требуются операции умножения и сложения больших чисел и нахождение остатка по модулю P.

Для снижения вычислительной сложности стоит применять метод Гарнера, в котором операция нахождения остатка вычисляется не от полного диапазона, а по множеству p_i.

В методе Гарнера используется полиадическая система счисления, где любое число представляется в системе взаимно простых чисел p_1, … p_n, следующим образом:

$$A = a_1 + a_2 \cdot p_1 + a_3 \cdot p_1 \cdot p_2 + ... + a_{n-1} \cdot p_1 \cdot p_2 \cdot ... \cdot p_{n-2} + a_n \cdot p_1 \cdot p_2 \cdot ... \cdot p_{n-1}.$$

коэффициенты a_i, $i = [0:n]$ вычисляются следующим образом:

$$a_1 \equiv s_1 (\mathrm{mod} p_1);$$

$$a_2 \equiv (s_2 - a_1)\ \tau_{12} (\mathrm{mod} p_2);$$

$$a_3 \equiv ((s_3 - a_1)\ \tau_{13} - a_2) \tau_{23} (\mathrm{mod} p_3);$$

$$a_n \equiv (((...(s_n - a_1)\ \tau_{1n} - a_2) \tau_{2n} - ... a_{n-1}) \tau_{n-1n} (\mathrm{mod} p_n)$$

Константы τ_{kj} рассчитываются из условий

$$\tau_{kj} = \left| \frac{1}{p_k} \right|_{p_j} \text{ где } 1 \le k < j \le n.$$

Подставив константы τ_{kj}, получим:

$$a_1 \equiv \alpha_1 (\mathrm{mod} p_1);$$

$$a_2 = ((p_1^{-1}) \mathrm{mod} p_2 \cdot (s_2 - a_1)) \mathrm{mod} p_2;$$

$$a_3 = ((p_2^{-1}) \mathrm{mod} p_3 \cdot ((p_1^{-1}) \mathrm{mod} p3 \cdot (s_3 - a_1) - a_2)) \mathrm{mod} p_3;$$

$$a_4 = ((p_3^{-1}) \mathrm{mod} p_4 \cdot ((p_2^{-1}) \mathrm{mod} p_4 \cdot ((p_1^{-1}) \mathrm{mod} p_4 \cdot (s_4 - a_1) - a_2) - a_3)) \mathrm{mod} p_3;$$

…

Недостатком метода Гарнера является то, что в каждой итерации применяются операции вычитания и умножения по модулю p_i.

Метод, основанный на совместном применении КТО и ОПСС, исключает операцию вычитания.

Для решения этим методом, ортогональные базисы необходимо представить в ОПСС:

$$b_i = b_{i1} + b_{i2} \cdot p_1 + b_{i3} \cdot p_1 \cdot p_2 + ... + b_{in} \cdot p_1 \cdot p_2 \cdot ... \cdot p_{n-1}$$

где: b_{ij} — это коэффициенты ОПСС; $i, j = 1,2,..,n$.

В связи с тем, что $b_i \mathrm{mod} p_i = 0$, $\forall j > i$, то перед первым значащим значением будет $i-1$ нулей.

Для удобства базисы можно представить в виде матрицы размерностью $[n,n]$.

$$\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ 0 & 0 & \dots & b_{nn} \end{bmatrix}.$$

Тогда

$$A \to \begin{bmatrix} |\alpha_1 b_{11}|_{p_1}^{+} & |\alpha_1 b_{12}|_{p_2}^{+} & \dots & |\alpha_1 b_{1n}|_{p_n}^{+} \\ 0 & |\alpha_2 b_{22}|_{p_2}^{+} & \dots & |\alpha_1 b_{2n}|_{p_n}^{+} \\ 0 & 0 & \dots & |\alpha_1 b_{nn}|_{p_n}^{+} \end{bmatrix}.$$

При этом $a_i = \left| \sum_{\substack{i=1 \\ j=1}}^{n} \alpha_i b_{ij} \right|_{\bmod p_i}$.

Из вышесказанного можно сделать вывод о привлекательности для практической реализации метода совместного использования КТО и ОПСС.

Метод разделения данных на множество подгрупп

Для повышения надежности схем разделения данных внедряется алгоритм разделения на множество подгрупп (рис. 1).

Рис. 1. – Структурная схема разделения данных на множество подгрупп

Ниже представлена работа алгоритма с разделением на множество подгрупп. Она состоит из двух этапов:

1) Информация A разделяется на множество, состоящее из n частей «лидеров групп» $F_1, F_2, ..., F_n$.

2) Каждый «лидер группы» $F_1, F_2, ..., F_n$ разделяется на свое новое множество, состоящее из m частей $(F_{1_1}, F_{1_2}, ..., F_{1_m})(F_{2_1}, F_{2_2}, ..., F_{2_m})..(F_{n_1}, F_{n_2}, ..., F_{n_m})$.

Полученные $n \times m$ частей информации $(F_{1_1}, F_{1_2}, ..., F_{1_m})(F_{2_1}, F_{2_2}, ..., F_{2_m})..(F_{n_1}, F_{n_2}, ..., F_{n_m})$ передаются по линиям связи на удаленные сервера [8].

При хранении и передаче информации возможно использовать СОК. Для этого исходная информация будет разделена на части согласно количеству оснований СОК [9].

Для надежного хранения и передачи рассчитанных данных вводятся избыточные основания, тем самым увеличивается избыточность информации. В результате этого получаем классическую пороговую схему разделения данных, где для восстановления информации достаточно получить от k групп K частей. Избыточность введенной информации позволяет локализовать скомпрометировавшие себя хранилища информации и заблокировать их для дальнейшего использования и восстановления исходной информации.

Оценка надежности групповой схемы разделения данных и линейной

В работе были сравнены надежность групповой и линейной схемы разделения данных. Расчёт надёжности представляет собой процедуру определения значений показателей надежности объекта с использованием методов, основанных на их вычислении по справочным данным о надежности элементов объекта, по данным о надежности объектов-аналогов, данным о свойствах материалов и другой информации, имеющейся к моменту расчета [10].

Вероятность того, что в системе, состоящей из n одинаковых и равно надёжных элементов, безотказно работают не менее k элементов, может быть вычислена по формуле:

$$P(t) = \sum_{i=k}^{n} \binom{n}{i} p(t)^i q(n)^{n-1} \quad (1)$$

где: $p(t)$ – вероятность безотказной работы одного элемента системы; $q(t)=1-p(t)$; $\binom{n}{i}=\frac{n!}{k!(n-k)!}$ – биноминальный коэффициент из n по k.

Рассмотрим пример для вычисления надежности функции при $n=5$, а $k=3$. В качестве времени распределения безотказной работы одного элемента системы применим экспоненциальный закон распределения $p(t)=\lambda e^{-\lambda t}$ при $\lambda=1$. Тогда, подставив значение в формулу (1), получим: $P(t)=\sum_{i=3}^{5}\frac{5!}{3!(5-3)!}(\lambda e^{-\lambda t})^i(1-\lambda e^{-\lambda t})^{n-1}$.

Так же рассмотрим вероятность безотказной работы системы при пропорциональном увеличении информационных и контрольных модулей $n_x=25$, а $k_x=15$ тогда: $P_x(t)=\sum_{i=15}^{25}\frac{25!}{15!(25-15)!}(\lambda e^{-\lambda t})^i(1-\lambda e^{-\lambda t})^{n-1}$.

На рис. 2 показаны графики зависимости времени безотказной работы системы $P(t)$ и $P_x(t)$.

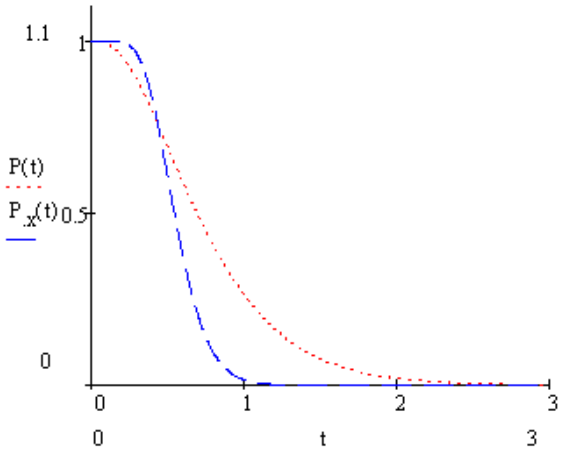

Рис. 2. – графики безотказной работы систем $P(t)$ и $P_x(t)$

Из графика видно, что время работы системы $P_x(t)$ намного меньше, чем $P(t)$.

Проинтегрировав $P(t)$ и $P_x(t)$ получим среднее время безотказной работы системы: $\int_0^\infty P(t)dt=0{,}783$, $\int_0^\infty P_x(t)dt=0.564$. Следовательно, при пропорциональном увеличении информационных и контрольных модулей надежность системы

падает. Для рассмотренного нами примера снижение среднего времени работы системы составляет 72%.

Далее рассмотрим пример для определения среднего времени работы схемы с групповым разделением данных. Для этого разобьем схему из 25 элементов на 5 групп по 5 частей. В каждой группе 2 элемента будут избыточны и 2 группы будут избыточны. Тогда вероятность безотказной работы любой группы может быть оценена как $P_g(t)=\sum_{i=3}^{5}\frac{5!}{3!(5-3)!}(\lambda e^{-\lambda t})^i(1-\lambda e^{-\lambda t})^{n-1}$, а системы в целом $P_{ob}(t)=\sum_{i=3}^{5}\frac{5!}{3!(5-3)!}(P_g(t))^i(1-\lambda e^{-\lambda t})^{n-1}$.

На рис. 3 показаны графики зависимости времени безотказной работы системы $P_{ob}(t)$ и $P_x(t)$.

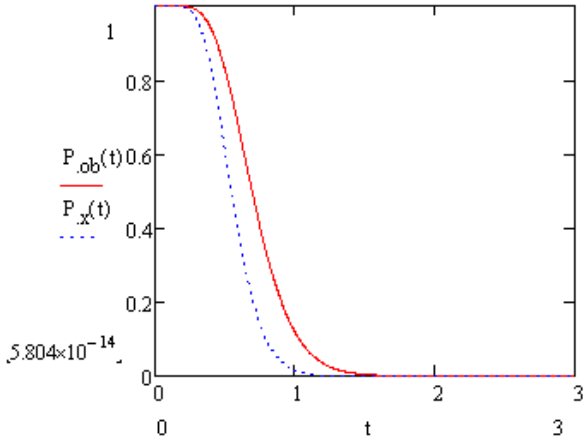

Рис. 3 – графики безотказной работы систем $P_{ob}(t)$ и $P_x(t)$

Из графика видно, что время работы системы групповой схемы разделения сданных $P_{ob}(t)$ выше чем линейной $P(t)$.

Проинтегрировав $P_{ob}(t)$ и $P_x(t)$ получим среднее время безотказной работы системы: $\int_0^\infty P_{ob}(t)dt=0{,}719$, $\int_0^\infty P_x(t)dt=0.564$. Следовательно, при одинаковом числе участников схемы обмена сообщениями целесообразно использование групповой схемы разделение данных. Для рассмотренного нами примера преимущество групповой схемы составляет 78%.

Вывод

Рассмотренный метод разделения информации на группы подмножеств, основанный на СОК, показал что его обнаруживающая способность кода составляет $\left(\frac{P_{n1}-P_{k1}}{P_{n1}}+\frac{P_{n2}-P_{k2}}{P_{n2}}+...+\frac{P_{nm}-P_{km}}{P_{nm}}\right)+\frac{P_n-P_k}{P_n}$ в отличие от классического $\frac{P_n-P_k}{P_n}=\frac{P_n-1}{P_n}$. Также уменьшение разрядности частей информации приведет к снижению нагрузки на сети передачи данных, либо сервера.

Рассмотрены способы восстановления информации из СОК в ПСС. Оценка надежности показала, что среднее время работы групповой схемы разделения информации выше, на рассмотренных нами примерах она увеличивается на 78%.

Литература

1. Акушский И.Я. и Юдицкий Д.И, Машинная арифметика в остаточных классах. Москва: Советское радио, 1968. 440 с.

2. Goldreich O, Ron D., and Sudan M., Chinese remaindering with errors // IEEE Trans. Inf. Theory, 2000, № 46, Т. 4, pp. 1330–1338

3. Соловьев Р.А., Тельпухов Д.В., Балака Е.С., Рухлов В.С., и Михмель А.С. Аппаратная реализация операции нахождения остатка от деления для входных данных большой разрядности на основе блоков редукции и коррекции // Инженерный вестник Дона, 2017, №2. URL: ivdon.ru/ru/magazine/archive/N2y2018/5042.

4. Asmuth C. and Bloom J., A modular approach to key safeguarding // IEEE Trans. Inf. Theory, 1983, №. 29, Т. 2, pp. 208–210.

5. Mignotte M., How to Share a Secret, // Cryptography, Springer, Berlin 1982, pp. 371–375.

6. Кочеров Ю. Н. и Червяков Н. И., Разработка методов и алгоритмов разделения и восстановления данных в модулярных пороговых структурах

для распределенных вычислительных сетей // Ставрополь: Северо-Кавказский федеральный университет, 2016. 236 с.

7. Iftene S., General Secret Sharing Based on the Chinese Remainder Theorem with Applications in E-Voting // Electron Notes Theor Comput Sci, 2007, №186, pp. 67–84.

8. Kocherov Y.N., Samoylenko D.V., and Koldaev A. I., Development of an Antinoise Method of Data Sharing Based on the Application of a Two-Step-Up System of Residual Classes // 2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), 2018, pp. 1–5.

9. Krasnobaev V., Koshman S., Kononchenko A., Kuznetsova K., and Kuznetsova T., The Formulation and Solution of the Task of the Optimum Reservation in the System of Residual Classes // 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), 2019, pp. 1–4,

10. Назаров А.С., Дерябин М.А., Бабенко М.Г., и Тарасенко Е.О., Вероятностный подход к оценке отказоустойчивости различных моделей распределенного хранения данных // Инженерный вестник Дона, 2019, №8. URL: ivdon.ru/ru/magazine/archive/N8y2019/6137.

References

1. Akushskij I.YA. i YUdickij D.I, Mashinnaya arifmetika v ostatochnyh klassah [Machine arithmetic in residual classes]. Moskva: Sovetskoe radio, 1968. 440 p.

2. Goldreich O, Ron D., and Sudan M. IEEE Trans. Inf. Theory, 2000, № 46, T. 4, pp. 1330–1338

3. Solov'ev R.A., Tel'puhov D.V., Balaka E.S., Ruhlov V.S., i Mihmel' A.S., Inzhenernyj vestnik Dona, 2017, №2. URL: ivdon.ru/ru/magazine/archive/N2y2018/5042.

4. Asmuth C. and Bloom J. IEEE Trans. Inf. Theory, 1983, №. 29, T. 2, pp. 208–210.

5. Mignotte M. Cryptography, Springer, Berlin 1982, pp. 371–375.

6. Kocherov YU. N. i CHervyakov N. I., Razrabotka metodov i algoritmov razdeleniya i vosstanovleniya dannyh v modulyarnyh porogovyh strukturah dlya raspredelennyh vychislitel'nyh setej [Development of methods and algorithms for data separation and recovery in modular threshold structures for distributed computing networks], Stavropol': Severo-Kavkazskij federal'nyj universitet, 2016. 236 p.

7. Iftene S. Electron Notes Theor Comput Sci, 2007, №186, pp. 67–84.

8. Kocherov Y.N., Samoylenko D.V., and Koldaev A. I. International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), 2018, pp. 1–5.

9. Krasnobaev V., Koshman S., Kononchenko A., Kuznetsova K., and Kuznetsova T., 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), 2019, pp. 1–4.

10. Nazarov A.S., Deryabin M.A., Babenko M.G., i Tarasenko E.O., Inzhenernyj vestnik Dona, 2019, №8. URL: ivdon.ru/ru/magazine/archive/N8y2019/6137.