Способ контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений

А. О. Лачинов, email: avatar31221@yandex.ru
А. В. Решотка, email: ReshA_Vl@mail.ru
А. В. Лавриненко, email: Lavrinenko312@mail.ru
Н. С. Савельев, email: nssavelyev01@ya.ru
Н. А. Сипатров, email: nikita.Sipatrov@yandex.ru
А. Д. Ваничкин, email: vanamaka@mail.ru
Т. В. Стариков, email: TVS_7@mail.ru

Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М. Штеменко

Аннотация. *В данной работе рассматривается способ контроля целостности данных, использующий комбинации хэш-значений, полученных от данных, подлежащих защите от деструктивных воздействий среды и злоумышленника.*

Ключевые слова: *контроль целостности данных, снижение избыточности, хэш-функция, система хранения данных, Big Data.*

Введение

Способ контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений заключается в том, что применяя хэш-функцию к определенному количеству подблоков памяти M_i с помощью сочетаний полученных хэш-значений H_j можно осуществлять контроль целостности данных для поиска однократной ошибки.

В настоящее время одним из приоритетных направлений исследований и разработок государства в области информационных технологий является сфера обработки больших данных (Big Data). В связи с этим одной из основных задач государства по развитию отрасли информационных технологий в этом направлении является развитие защищенных информационно-аналитических систем (ИАС), требующих высокого уровня достоверности информации [1-3].

На сегодняшний день самым популярным из существующих решений реализации контроля целостности данных в системах хранения данных (СХД) является способ применения хэш-функции, при которой от каждого из блоков данных вычисляется хэш-значение. Недостатком данного способа является высокая избыточность при контроле целостности блоков данных [4-9].

1. Способ контроля целостности данных

Основная идея способа контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений заключается в использовании формулы комбинаторики C_n^m для вычисления сочетаний хэш-значений таким образом, что каждому из подблоков памяти M_i ставится в соответствии комбинация из H_j. При этом количество используемых хэш-значений всегда меньше количества подблоков памяти M_i, которые необходимо защитить. Следовательно, одним из преимуществ такого способ является снижение вводимой избыточности при контроле целостности данных.

Блок данных M, который необходимо защитить, представляется в виде конечного счетного множества подблоков памяти v.

Количество сочетаний вычисляется по формуле

$$C_n^m = \frac{n!}{m!(n-m)!},$$

где n – количество хэш-значений, используемых для защиты всех подблоков блока данных M, а m – количество хэш-значений, используемых для защиты одного подблока данных M_i.

Количества поблоков памяти k меньше или равно количеству сочетаний хэш-значений H_j, необходимых для защиты блока данных. Это значит, что потребуется использовать меньше ресурсов СХД для хранения хэш-значений, следовательно уменьшит избыточность хранимой информации.

Для того чтобы определить оптимальное количество хэш-значений необходимых для защиты блока данных M, а также количество хэш-значений необходимых для защиты одного подблока памяти M_i можно воспользоваться арифметическим треугольником Паскаля, где значение элемента треугольника всегда должно быть больше чем численное значение количества поблоков памяти. Треугольник Паскаля можно представить в виде таблицы, как показано в таблице 1.

Таблица 1

Арифметический треугольник Паскаля

n	m						
	0	1	2	3	4	5	…
0	1						

1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
…	…	…	…	…	…	…	…

Из арифметического треугольника Паскаля видно, что оптимальное количество хэш-значений, необходимых для защиты одного подблока данных M_i, достигается при $m = \frac{n}{2}$.

2. Пример

Для защиты блока памяти M, состоящего из 6 подблоков $M = M_1, M_2 ... M_6$, будем использовать сочетания из 4 хэш-значений $H = H_1, H_2, H_3, H_4$ по 2 хэш-значения для каждого подблока M_i.

Схема, поясняющая метод контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений представлена на рисунке.

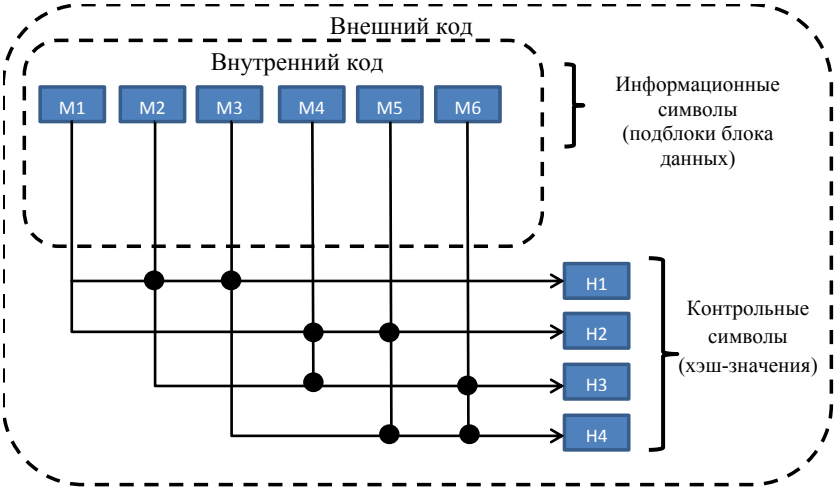

Рисунок. Схема, поясняющая способ контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений, используемый для защиты 6 подблоков памяти.

Для контроля целостности защищенного блока данных вычисляется синдром $\vec{S} = (s_1, s_2 \ldots s_3)$, согласно предикату:

$$P(\vec{S}) = \begin{cases} 1, \text{ если } H_i \neq H_i^*; \\ 0, \text{ если } H_i = H_i^*, \end{cases}$$

где H_i – хэш-значение вычисленное от данных и подлежащее проверке, а H_i^* – эталонное хэш-значение.

В соответствии с таблицей 2 по полученному синдрому легко определить подблок блока данных, целостность которого нарушена.

Таблица 2

Таблица синдромов для локализации однократной ошибки

Синдром S_i					**Результат**
i	H_1	H_2	H_3	H_4	
0	0	0	0	0	Нет ошибок
1	1	1	0	0	M_1

2	1	0	1	0	M_2
3	1	0	0	1	M_3
4	0	1	1	0	M_4
5	0	1	0	1	M_5
6	0	0	1	1	M_6

3. Оценка разработанного способа контроля целостности данных

Оценка разработанного способа будет осуществляться с помощью сравнения способа контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений со способами контроля целостности данных основанных на криптографических преобразованиях кодов исправляющих ошибки, а именно прямоугольных кодов и кодов Хемминга [3-10].

В качестве критерия эффективности будет использоваться коэффициент избыточности информации $K_{изб}$, характеризующий отношение объема избыточных данных, необходимых для осуществления контроля целостности ($V_{изб}$) к объему данных, подлежащих защите ($V_{защ}$). Определение наиболее эффективного метода будет осуществляться по правилу $K_{изб} \rightarrow \min$. Коэффициент избыточности вычисляется в соответствии с формулой:

$$K_{изб} = \frac{V_{изб}}{V_{защ}}$$

Результаты вычислений для различного количества защищенных подблоков данных представлены в таблице 3.

Таблица 3

Таблица значений коэффициента избыточности для различных методов контроля целостности данных

Количество защищенных подблоков данных	**Значение коэффициента избыточности**		
	метод контроля целостности с применением прямоугольных кодов	**метод контроля целостности с применением сочетаний хэш-значений**	**метод контроля целостности с применением кодов Хемминга**
10	0,7	0,5	0,4
100	0,1	0,07	0,07

1 000	0,032	0,013	0,01
10 000	0,01	0,0016	0,0014

Из результатов оценки эффективности видно, что способ контроля целостности данных на основе применения хэш-функции по принципу сочетаний хэш-значений уступает способу контроля целостности данных основанному на криптографических преобразованиях кодов Хемминга [3, 5, 6, 10], но значительно опережает способ контроля целостности данных на основе прямоугольных кодов [4, 7-9].

4. Заключение

Не смотря на то, что современные технологии позволяют хранить невероятные объемы данных, условия обеспечения безопасности информации накладывают ограничения, которые не позволяют, например, использовать технологии облачных хранилищ или приводят к большим экономическим затратам на создание нового оборудования. Но в тоже время данные проблемы вынуждают инженеров, математиков и программистов создавать новые методы хранения информации, приводя в движение механизм, именуемый прогрессом. Каждый шаг, сделанный в сторону решения проблемы хранения защищенных данных, приводит к экономии ресурсов. Одним из таких шагов является полученный в результате данной работы метод, позволяющий снизить избыточность хранимой в СХД памяти за счет сокращения числа эталонных хэш-значений, используемых для обеспечения контроля целостности данных.

Список литературы

1. Стратегией развития отрасли информационных технологий в Российской Федерации на 2014-2020 годы и на перспективу до 2025 года.

2. Диченко, С. А. Концептуальная модель обеспечения целостности информации в современных системах хранения данных / С. А. Диченко // Информатика: проблемы, методология, технологии. Сборник материалов XIX международной научно-методической конференции. Под ред. Д. Н. Борисова. – Воронеж, 2019. – С. 697-701.

3. Диченко, С. А. Контроль и обеспечение целостности информации в системах хранения данных / С. Диченко // Наукоемкие технологии в космических исследованиях Земли. – 2019. – Т. 11. – № 1. – С. 49-57.

4. Левин, В. Ю. О повышении криптостойкости однонаправленных хэш-функций / В. Ю. Левин // Фундамент. и прикл. матем. – 2009. – Т. 15. Выпуск 5. – С. 171-179.

5. Диченко, С. А. Гибридный крипто-кодовый метод контроля и восстановления целостности данных для защищённых информационно-аналитических систем / С. Диченко, О. Финько // Вопросы кибербезопасности. – 2019. – № 6(34). – С. 17-36.

6. Dichenko, S. Two-dimensional control and assurance of data integrity in information systems based on residue number system codes and cryptographic hash functions / S. Dichenko, O. Finko // Integrating Research Agendas and Devising Joint Challenges International Multidisciplinary Symposium ICT Research in Russian Federation and Europe. – 2018. – P. 139-146.

7. Диченко, С. А. Системный анализ проблемы обеспечения целостности данных в информационно-аналитических системах / С. А. Диченко // Информатика: проблемы, методы, технологии. Сборник материалов XX Международной научно-методической конференции. Под ред. А.А. Зацаринного, Д.Н. Борисова. – Воронеж, 2020. – С. 1001-1005.

8. Диченко, С. А. Алгоритм проверки достоверности контрольной информации, используемой при обеспечении целостности данных в условиях деструктивных воздействий злоумышленника и среды / С. А. Диченко // Информатика: проблемы, методы, технологии. Сборник материалов XX Международной научно-методической конференции. Под ред. А.А. Зацаринного, Д.Н. Борисова. – Воронеж, 2020. – С. 996-1000.

9. Диченко, С. А. Разработка алгоритма контроля и обеспечения целостности данных при их хранении в центрах обработки данных / С. А. Диченко [и др.] // Сб. науч. статей VIII Междунар. молод. научнопр. конф. с элементами науч. шк. – Омск: Омский ГТУ, 2018. – С. 40-43.

10. Хэмминг, Р. В.: Теория кодирования и теория информации: Пер. с англ. – М.: Радио и связь, 1983. – 176 с.