С. В. Запечников (к. т. н., доцент)
Московский инженерно-физический институт (государственный университет)

КОНТРОЛЬ ЦЕЛОСТНОСТИ ИНФОРМАЦИОННЫХ РЕСУРСОВ ПРИ РАСПРЕДЕЛЕННОМ ХРАНЕНИИ ДАННЫХ

Представлены основные результаты исследования, направленного на решение задачи контроля целостности информационных массивов, подвергаемых биективным преобразованиям, при распределенном способе их хранения в компьютерной системе. Разработаны алгоритмы формирования проверочных кодов преобразованного массива и контроля по ним целостности исходного массива. Алгоритмы реализуют два способа контроля: полный и вероятностный.

Введение

Развитие методов и технических средств распределенного хранения и обработки данных открывает широкие возможности для применения распределенного хранения информационных ресурсов в различных сферах информационных технологий. Организация распределенного хранения информационных ресурсов позволяет решать разнообразные системотехнические и прикладные задачи. В качестве примеров назовем те из них, которые напрямую связаны с задачами обеспечения информационной безопасности:

· повышение живучести распределенных компьютерных систем (РКС) в части устойчивости к утрате пользовательской и системной информации в случае разрушения части аппаратно-программных средств РКС;

· обеспечение требуемых показателей достоверности и сохранности информации в базах данных и системах сетевого хранения данных;

· построение криптосистем, стойких к компрометации, разрушению или утрате ключей, как секретных, так и открытых, за счет применения схем разделения секрета, распределенного хранения репозиториев сертификатов, идентификационной информации и других видов ключевого материала.

Под информационным ресурсом в настоящей работе понимается информационный массив произвольной, но конечной длины, который может содержать информацию любого типа: базу данных, файлы, программный код и т. д. В самом общем виде распределенное хранение информационного массива реализует принцип введения избыточности и аналитических зависимостей между блоками данных, которые сохраняются на носителях устройств хранения данных (УХД) и представляют образ информационного массива. Исходный массив является по отношению к ним прообразом.

1. Задача контроля целостности информационных ресурсов при распределенном хранении данных

Сформулируем задачу контроля целостности информационного ресурса. Преобразование информационного ресурса при распределенном способе его хранения можно представить следующим образом. Пусть F — исходный массив. Обозначим через A алгоритм прямого преобразования массива, выполняемый при сохранении его в РКС, через A^{-1} — обратный алгоритм, выполняемый при восстановлении массива. Алгоритмы A и A^{-1} обязательно должны осуществлять биективное преобразование массива. Примером таких преобразований с доказанным свойством биективности являются алгоритмы безопасного размещения данных в РКС, ранее предложенные автором настоящей работы [1]. Если длина массива значительна, он может разбиваться на последовательность фрагментов {F[1], F[2], ...,F[s]}, каждый из которых обрабатывается РКС за одно обращение к алгоритмам A и A^{-1}. Если его длина невелика, то такой массив считается состоящим из единственного фрагмента.

Запишем уравнение прямого преобразования:

$$\vec{W}[m] = A\,(\vec{V}[m]),$$

и уравнение обратного преобразования:

$$\vec{V}[m] = A^{-1}\,(\vec{W}[m]).$$

Здесь $\vec{V}[m] = (V_m, \ldots, V_{m+k})$ — последовательность блоков исходного фрагмента $F[j]$, $j \in \{1, \ldots, s\}$, которая преобразуется в одну серию блоков $\vec{W}[m] = (W_{I_1}[m], \ldots, W_{I_n}[m])$ размещаемых на носителях УХД, за одно обращение к алгоритму, а $m = \overline{1, s}$.

Таким образом, после преобразования фрагмента массива $F[j]$, $m = \overline{1, s}$ получаются последовательности блоков данных $\{(W_{I_1}[1], W_{I_1}[2], \ldots, W_{I_1}[s]\}$, $\{(W_{I_2}[1], W_{I_2}[2], \ldots, W_{I_2}[s]\}$, ..., $\{(W_{I_n}[1], W_{I_n}[2], \ldots, W_{I_n}[s]\}$, размещенные на узлах РКС с идентификаторами $I_1, I_2, \ldots, I_n$ соответственно. Требуется проверить тот факт, что из этих последовательностей блоков по алгоритму A^{-1} можно в точности восстановить исходный массив.

2. Формирование проверочных кодов массива

Пусть при выполнении алгоритма A для каждого полученного фрагмента с целью контроля его целостности вычисляется по некоторому алгоритму контрольный код $C(W_X[r])$, где $X = I_1, I_2, \ldots, I_n$, $r = \overline{1, s}$. В качестве контрольных кодов могут применяться различные виды кодов, обнаруживающих и исправляющих ошибки, а также хэш-коды, вычисленные с помощью криптографических хэш-функций. В результате преобразования фрагмента массива получается совокупность блоков данных и их контрольных кодов, показанная на рис. 1. Для формирования проверочных кодов блоков данных, распределенных по узлам РКС, и их последующей проверки предлагается использовать следующие алгоритмы.

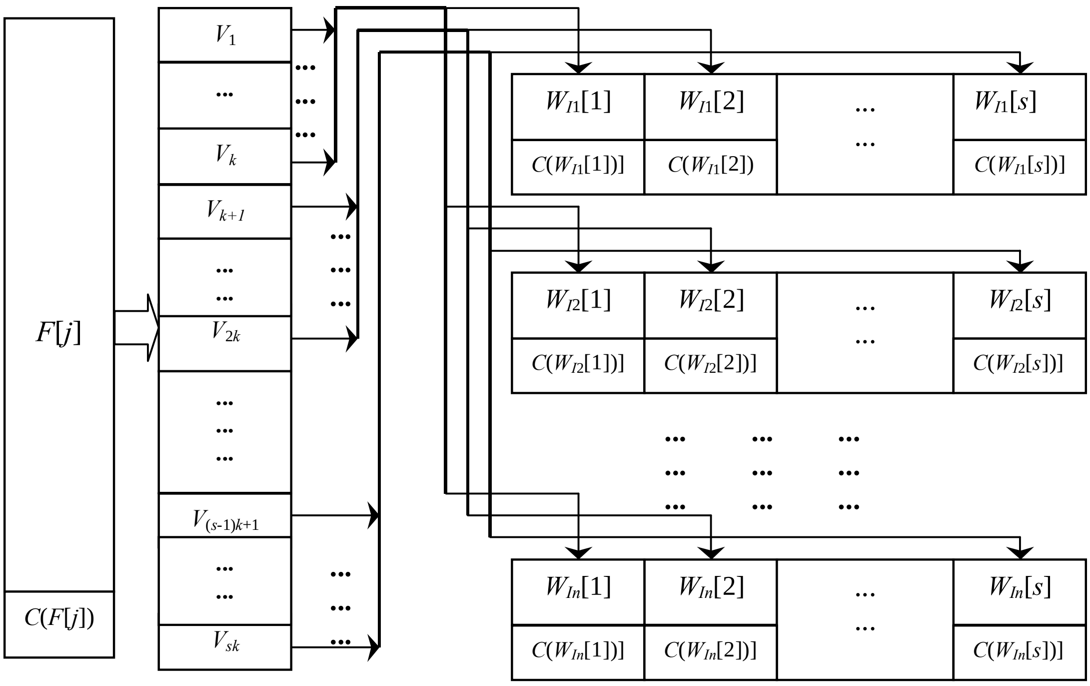


Рис. 1. Схема размещения блоков данных преобразованного массива по узлам РКС

Алгоритм 1 (*формирование проверочных кодов совокупности размещенных на узлах РКС блоков данных*)

1. Интерпретируем каждое значение контрольных кодов блоков данных $C(W_X[r])$, где $X = I_1, I_2, \ldots, I_n$, $r = \overline{1, s}$ как элемент поля $GF(2^m)$, где m — длина $C(W_X[r])$ в битах.

2. Для кодирования каждой совокупности контрольных кодов $\{C(W_{I_1}[r]), C(W_{I_2}[r]), \ldots, C(W_{I_n}[r])\}$, $r = \overline{1, s}$, применим $(n,2t)$ — код Рида — Соломона [2. С. 273–275], где t — задаваемое наперед предельно допустимое количество утрат узлов РКС, при котором должна сохраняться аутентичность фрагмента массива $F[j]$. Обозначим полученное кодовое слово через

$$J[r] = RS\,(C(W_{I_1}[r]), C(W_{I_2}[r]), \ldots, C(W_{I_n}[r])).$$

3. Сохраним кодовое слово $J[r]$ на каждом из узлов РКС с идентификатором $X \in \{I_1, I_2, ..., I_n\}$ вместе с блоком $W_X[r]$ и контрольным кодом $C(W_X[r])$.

Конец алгоритма.

Если выбран систематический код Рида — Соломона, то на каждом из узлов РКС достаточно сохранить только проверочные разряды кодового слова $J[r]$. Тогда объем данных, сохраняемых на каждом из узлов, равен $V_1 = s \cdot \left(|W_X[r]| + |C(W_X[r])| + \left\lceil \frac{n}{n-2t} \right\rceil \cdot |C(W_X[r])| \right)$ бит. Если код несистематический, то кодовые слова необходимо сохранять целиком, и тогда объем сохраняемых на каждом узле данных составит $V_1 = s \cdot \left(|W_X[r]| + \left(n + \left\lceil \frac{n}{n-2t} \right\rceil \right) \cdot |C(W_X[r])| \right)$ бит.

Если дополнительно к обеспечению целостности требуется подтверждение подлинности сохраняемого в РКС информационного ресурса, то всем узлам РКС необходимо выдать доли секретного ключа схемы электронной цифровой подписи (ЭЦП), распределенного с помощью любой (t, n)-пороговой схемы разделения секрета: $sk \xleftrightarrow{(t,n)} (sk_1, ..., sk_n)$, и добавить к алгоритму 1 следующий дополнительный шаг: все узлы с идентификаторами $I_1, I_2, ..., I_n$ совместно выполняют протокол генерации цифровой подписи для кодового слова $J[r]$, используя любую пороговую схему ЭЦП, например, схемы [3, 4].

Формируемые таким образом проверочные коды для преобразованного и распределенного по узлам РКС массива позволяют осуществлять контроль целостности массива, в том числе дистанционный. Такой контроль может совмещаться с процедурой восстановления информационного массива либо выполняться периодически во время хранения его на узлах РКС.

3. Полный контроль целостности массива

Полный контроль целостности может быть осуществлен путем тотальной проверки всех кодовых слов $J[r]$, $r = \overline{1, s}$.

Алгоритм 2 *(полный контроль целостности преобразованного массива, распределенного по узлам РКС)*

1. Положить $i = 1$, $r_i \equiv i$.

2. Для считать $\forall\, m = \overline{1, n}$ с УХД узла РКС блок данных $[\tilde{W}_{I_m}[r_i], \tilde{C}(\tilde{W}_{I_m}[r_i]), ..., \tilde{J}_m[r_i]]$ и вычислить $\tilde{C}(\tilde{W}_{I_m}[r_i])$ по алгоритму формирования контрольных кодов, принятому в системе.

3. Для $\forall\, m = \overline{1, n}$ сравнить: $\tilde{C}(\tilde{W}_{I_m}[r_i]) \overset{?}{=} \tilde{C}(\tilde{W}_{I_m}[r_i])$. Если $\exists m$, для которого равенство не выполнено, то выдать сообщение о нарушении целостности блока данных с указанием значений m, r_i.

4. Если количество блоков, для которых равенство не выполнено, превышает t, то выдать сообщение о нарушении целостности массива с указанием значения r_i.

5. Используя алгоритм формирования $(n, 2t)$-кодов Рида — Соломона, получить $J[r_i] = RS(C(\tilde{W}_{I_1}[r]), C(\tilde{W}_{I_2}[r]), ..., C(\tilde{W}_{I_n}[r]))$.

6. Сравнить величины $\tilde{J}_m[r_i]$, считанные со всех узлов I_m, $m = \overline{1, n}$. Если среди них есть несовпадающие, то принять за $\tilde{J}[r_i]$ такое значение, которое встречается наибольшее число раз. Если таких значений оказалось несколько, то выдать сообщение о нарушении целостности массива с указанием значения r_i.

7. Сравнить найденную на шаге (5) величину с величиной, полученной на шаге (6): $J[r_i] \overset{?}{=} \tilde{J}[r_i]$. Если равенство не имеет места, то выполнить для $\tilde{J}[r_i]$ алгоритм Берлекэмпа — Месси [5. С. 211–221] для отыскания локаторов ошибок $\sigma_1, ..., \sigma_u$. Если $u > 1$, то выдать сообщение о нарушении целостности массива с указанием значения r_i. Если $0 < u \le 1$, то выдать сообщение о нарушении целостности блоков с указанием $\sigma_1, ..., \sigma_u$, r_i.

8. Проверить: $i < s$? Если да, то положить $i \leftarrow i + 1$ и перейти к шагу (1). В противном случае выдать подтверждение целостности массива.

Конец алгоритма.

Вероятность появления ошибки в декодированном символе кодов Рида — Соломона оценивается следующей величиной [6. С. 461]:

$$P_E \approx \frac{1}{2^m - 1} \sum_{v=t+1}^{2^m - 1} v C_{2^m - 1}^{v} p^v (1-p)^{2^m - 1 - v},$$

где ρ — вероятность появления ошибки в символе $(n, 2t)$ кода Рида — Соломона, т. е. в данном случае в контрольном коде блока, m — длина каждого из $C(W_X[r])$ в битах.

Приняв предположение о том, что все нарушения аутентичности вызваны деятельностью противника и происходят с интенсивностью μ, получаем формулу для оценки вероятности необнаружения нарушений целостности фрагмента $F[j]$:

$$\beta(F[j], \lambda) = 1 - \left(1 - \frac{1}{2^m - 1} \sum_{v=t+1}^{2^m - 1} v C_{2^m - 1}^{v} \left(1 - e^{-\mu|\lambda|}\right)^v e^{-\mu|\lambda| \cdot (2^m - 1 - v)}\right)^s \quad (1)$$

где $|\lambda|$ — длина временного интервала, для которого выполняется оценка, выраженная в условных единицах времени.

4. Вероятностный контроль целостности массива

Процедура полного контроля целостности довольно трудоемка: алгоритм 2 требует sn вычислений контрольных кодов блоков данных и s вычислений кодовых слов $(n, 2t)$ кода Рида — Соломона. И поэтому для периодического контроля целостности целесообразно применять вероятностный метод контроля: выборочно опробовать некоторые блоки данных и с определенной вероятностью делать вывод о целостности массива либо обнаруживать нарушение его целостности. Для разработки процедуры вероятностного контроля доказывается вспомогательное утверждение.

Утверждение 1. Пусть s — полное количество блоков данных, размещенных на каждом из узлов $I_1, I_2, \ldots, I_n$. Пусть c — количество контролируемых блоков данных, $c \le s$. Пусть d — количество блоков данных, разрушенных противником, $d \le s$. Тогда вероятность $P_{обн}$ того, что среди контролируемых блоков окажется хотя бы один, разрушенный противником, оценивается следующим выражением:

$$1 - \left(1 - \frac{d}{s}\right)^c \le P_{обн} \le 1 - \left(1 - \frac{d}{s - c + 1}\right)^c \quad (2)$$

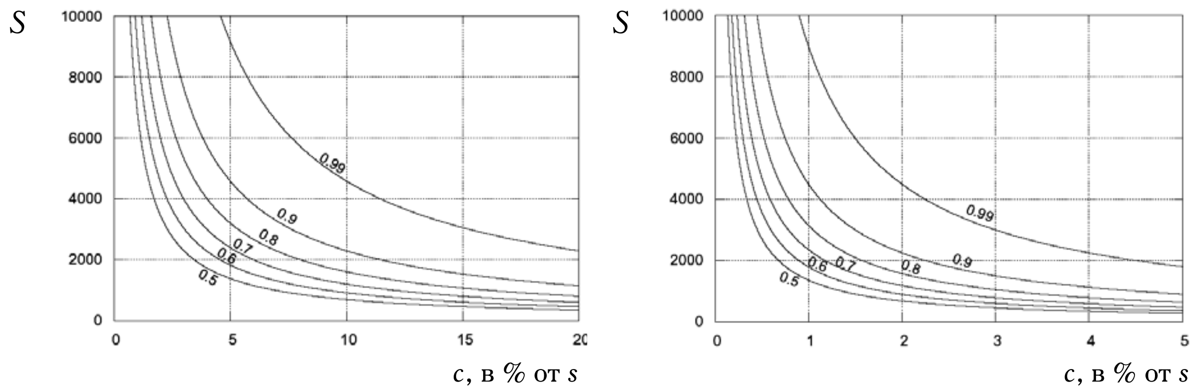
Доказательство: Обозначим через X случайную величину, которая равна количеству выбранных для контроля блоков, совпадающих с одним из блоков, разрушенных противником. Тогда:

$$P_{обн} = P(X \ge 1) = 1 - P(X = 0) = 1 - \frac{s-d}{s} \cdot \frac{s-1-d}{s-1} \cdot \frac{s-2-d}{s-2} \cdot \ldots \cdot \frac{s-(c-1)-d}{s-(c-1)} = 1 - \prod_{u=0}^{c-1} \left(1 - \frac{d}{s-u}\right)$$

Поскольку всегда $\frac{s-u-d}{s-u} \ge \frac{s-u-1-d}{s-u-1}$, то отсюда непосредственно следуют верхняя и нижняя оценки: $P_{обн} \ge 1 - \left(1 - \frac{d}{s}\right)^c$ и $P_{обн} \le 1 - \left(1 - \frac{d}{s-(c-1)}\right)^c$.

Таким образом, утверждение доказано.

На рис. 2 приведены примеры графиков, показывающих величину $P_{обн}$ при различных s, c, d. Исходя из анализа зависимостей, можно сделать следующее наблюдение: когда c и d выражены в процентах от общего количества блоков s, то нарушение целостности обнаруживается с определенной вероятностью при контроле фиксированного количества блоков, которое не зависит от общего количества блоков. Так, если $d = 1\%$ от s, то для того, чтобы достичь вероятности $P_{обн} = 99\%$, необходимо опросить $c = 460$ блоков, а чтобы достичь вероятности $P_{обн} = 95\%$, необходимо $c = 300$ блоков.



а) $d = 1\%$ от s б) $d = 5\%$ от s

Рис. 2. Вероятности обнаружения нарушения целостности массива в зависимости от его длины и числа контролируемых блоков

Из выражения (1) следует, что если задана вероятность $P_{обн}$, известен интервал времени λ и интенсивность атак противника μ, то необходимое количество контролируемых блоков определяется по формуле:

$$c = \left\lceil \frac{\ln\left(1 - P_{обн}\right)}{\ln\left(1 - \mu|\lambda|/s\right)} \right\rceil \quad (3)$$

Алгоритм 3 *(вероятностный контроль целостности преобразованного массива, распределенного по узлам РКС)*

1. Задать $P_{обн}$, найти c по формуле (3) и выбрать случайные числа $r_1, r_2, \ldots, r_c \in Z$ такие, что $1 \le r_l \le s$ для $\forall\, l = \overline{1, c}$.

2. Положить $i = 1$.

3. Выполнить шаги (2) — (7) алгоритма 3.

4. Проверить: $i < c$? Если да, то положить $i \leftarrow i + 1$ и перейти к шагу (3). В противном случае выдать подтверждение целостности массива и завершить работу алгоритма.

Конец алгоритма.

Из утверждения 1 и формул (1), (2) следует, что алгоритм 3 обнаруживает нарушения целостности с вероятностью:

$$P = \max\left\{P_{обн}; \beta\left(F[j], \lambda\right)\right\}.$$

Пусть t — порог допустимого количества утрат УХД, зависящий от алгоритмов A и A^{-1}. Если алгоритмы 2 и 3 не обнаруживают нарушений целостности массива, и существует не менее t узлов I_m, таких, что для $\forall r = \overline{1, s}$ не было обнаружено нарушений целостности блоков $W_{I_m}[r]$, то массив гарантированно может быть восстановлен по алгоритму A^{-1}. Если же таких узлов I_m менее t, то для восстановления такого сильно поврежденного массива необходимо составить карту размещения поврежденных блоков и для $\forall r = \overline{1, s}$ отыскивать такие подмножества узлов $X \subseteq \{I_1, \ldots, I_n\}$, $|X| \ge n - t$ которые содержат неповрежденные блоки $W_{I_x}[r]$, где $I_x \in X$, и из них по отдельности восстанавливать последовательность блоков $\vec{V}[r] = (V_r, \ldots, V_{r+k})$. Если для какого-либо r ни одного такого множества X не существует, то массив не может быть восстановлен.

Заключение

В статье сформулирована и решена задача контроля целостности информационных ресурсов при распределенном способе их хранения. Как показано в работе, в общем случае при распределенном хранении осуществляется биективное преобразование между прообразом — исходным информационным массивом

— и его образом — множеством блоков данных, которые сохраняются на носителях сетевых устройств хранения данных. Разработана система алгоритмов, предназначенных для решения поставленной задачи, а именно:

· алгоритм формирования проверочных кодов совокупности размещенных на узлах РКС блоков данных;

· алгоритм полного контроля целостности преобразованного массива, распределенного по узлам РКС;

· алгоритм вероятностного контроля целостности преобразованного массива, распределенного по узлам РКС.

Для каждого из двух способов контроля — полного и вероятностного — найдены вероятности необнаружения несанкционированных изменений массива (нарушений его целостности). Результаты работы нашли применение при разработке алгоритмического и протокольного обеспечения системы криптографической защиты информации, стойкой к частичному разрушению ключевой системы.

СПИСОК ЛИТЕРАТУРЫ

1. *Запечников С. В.* Живучесть систем защиты информации как фактор обеспечения информационной и функциональной безопасности распределенных компьютерных систем // Безопасность информационных технологий. 2005. № 4. С. 8–17.
2. *Вернер М.* Основы кодирования / Пер. с нем. М.: Техносфера, 2006.
3. *Запечников С. В.* Пороговые схемы цифровой подписи на основе ГОСТ Р 34.10–94 // Безопасность информационных технологий. 2001. № 3. С. 45–51.
4. *Архангельская А. В., Запечников С. В.* Схемы цифровой подписи на основе алгоритмов ГОСТ Р 34.10–2001 с применением аппарата парных отображений // Известия ТРТУ. 2006. № 7 (62). С. 194–201.
5. *Блейхут Р.* Теория и практика кодов, контролирующих ошибки / Пер. с англ. М.: Мир, 1986.
6. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение. 2-е изд., испр. / Пер. с англ. М.: Издательский дом «Вильямс», 2003.