Модель оценки своевременности и полноты обмена информационными ресурсами в корпоративных системах с распределенным реестром

И.Б. Саенко[1], И.Н. Фабияновский[1], А.М. Старков[1], Е.Г. Баленко[2]

[1]Военная академия связи имени Маршала Советского Союза С.М. Буденного, Санкт-Петербург

[2]Донской государственный аграрный университет, Персиановский

Аннотация: В статье рассматривается модель оценки показателей эффективности обмена информационными ресурсами в корпоративных системах, предназначенная для анализа возможности систем с распределенным реестром обеспечить своевременность и полноту информационного обмена. В качестве основного показателя предлагается учитывать вероятность отказа сегмента корпоративной системы для каждой эпохи. Для нахождения этого показателя используются вероятностные оценки сумм, ограниченных сверху гипергеометрическим и биномиальным распределениями с вероятностными границами Чебышева, Хеффдинга и Шватала. Проведен численный и сравнительный анализ предложенных оценок.

Ключевые слова: технология распределенного реестра, информационная система, сегментирование, цепочка блоков.

Технология распределенного реестра (ТРР) – это перспективная информационная технология обмена информационными ресурсами в корпоративных сетях (системах), которая может оказывать положительный эффект практически на все сегменты цифровой экономики, например, банковское дело, здравоохранение, государственный сектор и т.д.

Распределенный реестр (РР) можно определить, как распределенный цифровой журнал, который отслеживает все операции (транзакции) в корпоративной сети. РР не основывается на каком-либо доверенном центральном узле корпоративной сети для проверки операций (транзакций) и расширения цепочки блоков. Сеть, используя протокол согласия, определяет, на каком узле можно создавать действительный блок и добавлять его в систему распределенного реестра (СРР) [1].

Транзакция транслируется на все узлы в корпоративной сети. При получении транзакции узел, который его получает, проверяет, является ли

транзакция действительной. Если ответ положительный, то он отправляет транзакцию своим соседям. В противном случае транзакция отбрасывается. Узел, который получает блок, проверяет его. Если блок подлинный, то копия блока добавляется в CPP и передается соседним узлам. В противном случае копия отбрасывается. Таким образом, в конечном итоге все узлы имеют одну и ту же копию CPP.

Одним из ключевых ограничений CPP, основанных на доказательствах, является масштабируемость. Масштабируемость тесно связана с пропускной способностью CPP. Последний показатель измеряется количествам транзакций, которые могут обрабатываться в единицу времени (например, в секунду). Цель масштабируемости CPP заключается в том, чтобы обрабатывать большое количество транзакций в секунду (т.е. иметь высокую пропускную способность) без ущерба для децентрализации и безопасности. Безопасность, в свою очередь, определяется полнотой и своевременностью информационного обмена. Можно значительно увеличить пропускную способность, однако проиграть при этом в децентрализации и безопасности [2-4].

Учитывая вышеизложенное, для CPP требуется найти баланс между масштабируемостью, безопасностью и децентрализацией. Это является достаточно сложной проблемой. Для ее решения разработана модель, излагаемая в настоящей статье, позволяющая определить условия, которые должны быть выполнены в CPP на основе протокола сегментирования, чтобы вероятность отказа находилась ниже заданного порога.

В протоколах CPP на основе сегментирования предлагается использовать гипергеометрическое распределение вместо биномиального, так как процесс объединения узлов в сегменты можно определить, как выборку без замены (сегменты не пересекаются). В этом случае

гипергеометрическое распределение дает лучшее приближение по сравнению с биномиальным [5].

Обычно биномиальное распределение используется, когда выборка обращается с заменой. Есть несколько вероятностных границ, которые можно применить в СРР. Границы Маркова и Чебышева являются наиболее распространенными неравенствами, используемыми в теории вероятностей. Оценка Маркова применяется только к неотрицательным случайным величинам, тогда как Чебышевская граница может быть применена к любой случайной величине. Кроме того, для любой независимой случайной величины оценка Шватала и оценка Хеффдинга имеют аналогичные границы как для биномиальных, так и для гипергеометрических распределений. Иными словами, границы Маркова, Чебышева, Шваталя и Хеффдинга могут применяться как для биномиального, так и для гипергеометрического распределения. Однако оценка Маркова является наиболее грубой, потому что она постоянна и не меняется по мере того, как число узлов в сегменте увеличивается.

Совокупность гипергеометрического распределения $H(K, N, n, k)$ представляет собой сумму функции распределения вероятностей $h(K, N, n, i)$ для всех $i \geq k$. Обозначим: N – общее количество узлов, n – размер сегмента, K – общее количество вредоносных узлов, r – отказоустойчивость сегмента, R – общая отказоустойчивость, n_c – количество сегментов, p_c – вероятность отказа сегмента, p_e – вероятность отказа эпохи, $p_{bootstrap}$ – вероятность начальной загрузки, $h(K, N, n, k)$ – гипергеометрическое распределение с параметрами K, N и n, $H(K, N, n, k)$ – кумулятивное гипергеометрическое распределение с параметрами K, N и n, $B(n, p, k)$ – кумулятивное биноминальное распределение с параметрами n и p, A – среднее время до отказа (в годах), E_s – ожидаемое количество раундов сегментирования до отказа, Ns – количество раундов сегментирования [6, 7].

Примем следующие допущения:

- вероятность отказа определяется как вероятность того, что количество вредоносных узлов превышают установленный предел, равный 25% от всех узлов) в сети;

- граница вероятности отказа есть функция верхней границы, которая оценивает вероятность отказа;

- безопасность сегмента есть максимальное количество вредоносных узлов, при которых сегмент еще находится в безопасном состоянии, то есть обеспечиваются полнота и своевременность обмена информационными ресурсами.

Пусть X – случайная величина, соответствующая количеству вредоносных узлов в выборке сегмента, и $P(X = k)$ – вероятность того, что сегмент содержит k вредоносных узлов из n сети размером N .

В случае гипергеометрического распределения X следует гипергеометрическому распределению с параметрами K, N и n. Тогда среднее (ожидаемое значение) будет равно:

$$E(X) = n\frac{K}{N} = np,$$

а дисперсия будет иметь следующий вид:

$$\mathrm{var}(X) = \frac{np(1-p)(N-n)}{N-1}.$$

Вероятность $P\ (X = k)$ будет равна:

$$h(K,N,n,k) = \frac{\left(\frac{K}{k}\right)\left(\frac{N-K}{n-k}\right)}{\left(\frac{N}{n}\right)}.$$

Тогда вероятность отказа для сегмента с безопасностью r будет равна:

$$H(K,N,n,nr)=\sum_{k=[nr]}^{n}\frac{\left(\frac{K}{k}\right)\left(\frac{N-K}{n-k}\right)}{\left(\frac{N}{n}\right)}$$

В случае биномиального распределения X следует биномиальному распределению с параметрами n и p, где $p = K / N$ есть вероятность того, что узел является вредоносным. Тогда среднее значение равно:

$$E(X)=np,$$

а дисперсия будет равна:

$$\mathrm{var}(X)=np(1-p). \quad (1)$$

Таким образом, можно провести сравнение гипергеометрического распределения с биномиальным. Если n мало относительно размера сети N, то X может быть аппроксимировано биномиальным распределением. Практически гипергеометрическое распределение приближается к биномиальному, когда размер выборки составляет менее 10% [8-10].

Однако, когда размер выборки увеличивается по сравнению с размером сети, рекомендуется использовать гипергеометрическое распределение, так как оно дает лучшее приближение. Если размер выборки немного меньше размера сети, то тогда для расчета вероятности отказа используется кумулятивное геометрическое распределение или кумулятивное биномиальное распределение вместо кумулятивного гипергеометрического распределения.

Если предположить, что $X \sim B\ (n,\ p)$, т.е. X следует биномиальному распределению с параметрами n и p, то $p = K / N$ есть вероятность того, что узел аномальный. Тогда вероятность отказа одного сегмента с безопасностью r с использованием кумулятивного биномиального распределения может вычисляться следующим образом:

$$P(X \ge nr) = \sum_{k=[nr]}^{n} \left(\frac{n}{k}\right) p^k (1-p)^{n-k}.$$

При оценке вероятности отказа важной задачей является оценка вероятности ограничения функции ошибки для одного сегмента и, следовательно, для одной эпохи, используя три граничные функции, т.е. определение неравенств, которые дают точные результаты.

Пусть X следует гипергеометрическому распределению с параметрами K, N и n. Оценим вероятность отказа для одного сегмента, а также на каждую эпоху по Чебышевской границе.

Введем определение оценки Чебышева: если X есть случайная величина, то для любого a ≥ 0 имеем:

$$P(|X - E(X)| \ge a) \le \frac{\mathrm{var}(X)}{a^2}. \quad (2)$$

Используя (2), для соответствующей гипергеометрической случайной переменной оценим $P(X \ge nr)$:

$$P(X \ge nr) = P(X - E(X) \ge nr - E(X)) = P(X - np \ge nr - np) \le P(|X - np|) \ge nr - p)$$
$$\le \frac{\mathrm{var}(X)}{(nr-np)^2} = \frac{np(1-p)(N-n)}{(N-1)(nr-np)^2} \quad (3)$$

Ограничим вероятность отказа одного сегмента с защищенностью r следующим образом:

$$P(X \ge nr) \le \frac{np(1-p)(N-n)}{(N-1)(nr-np)^2}$$

При оценке Чебышева, соответствующей двучлену случайной переменной, можно ограничить вероятность отказа одного сегмента с защищенностью r. В этом случае, учитывая (1) и (3), оценка $P(X \ge nr)$ имеет следующий вид:

$$P(X \ge nr) \le \frac{np(1-p)}{(nr-np)^2}. \quad (4)$$

Таким образом, при известной оценке вероятности отказа для каждой эпохи можно вычислить объединение n_c с сегментами, где каждый сегмент

может потерпеть неудачу с вероятностью p_c. Если образец меньше 10% от размера сети СРР, то p_c рассчитывается с использованием кумулятивного биномиального распределения. В противном случае используется кумулятивное гипергеометрическое распределение.

В первую эпоху для каждого протокола сегментирования процедура выборов завершается неудачно с вероятностью $p_{bootstrap}$ [11].

Таким образом, вероятность неудачи для одной эпохи, используя оценку Чебышева, соответствующую гипергеометрической случайной величине, ограничена следующим образом:

$$p_{bootstrap} + n_c p_c \le p_{bootstrap} + n_c \frac{np(1-p)(N-n)}{(N-1)(nr-np)^2}. \quad (5)$$

Приведем теперь определение оценки Хеффдинга. Для оценки вероятности отказа оно предполагает использование следующего выражения:

$$H(K,N,n,k) \le G(x),$$

Где:

$$G(x) = \left(\left(\frac{p}{p+x} \right)^{p+x} \left(\frac{1-p}{1-p-x} \right)^{1-p-x} \right)^n,$$

Следовательно, можно оценить вероятность отказа одного сегмента с защищенностью r следующим образом:

$$H(K,N,n,nr) \le G(x). \quad (6)$$

Биномиальное распределение имеет аналогичную хвостовую часть:

$$B(n,p,nr) \le G(x), \quad (7)$$

где

$$B(n,p,nr) = \sum_{k=\lceil nr \rceil}^{n} \left(\frac{n}{k} \right) p^k (1-p)^{n-k}$$

Вероятность отказа для одной эпохи p_e ограничена следующим образом:

$$p_{bootstrap} + n_c p_c \le V(x),$$

где:

$$V(x) = p_{bootstrap} + n_c G(x), \quad n_c = \frac{N}{n}.$$

Оценка Шваталя предполагает еще экспоненциальную функцию, которая дает менее слабую оценку по сравнению с Хеффдингом. Эта оценка имеет следующий вид:

$$H(K, N, n, k) \le F(x),$$

где:

$$F(x) = \exp^{-2x^2 n}.$$

Таким образом, вероятность отказа для одной эпохи ограничена выражением:

$$p_{bootstrap} + n_c p_c \le U(x),$$

где:

$$U(x) = p_{bootstrap} + n_c F(x), \quad n_c = \frac{N}{n}.$$

Оценим теперь безопасность сети, т.е. вычислим среднее количество времени до отказа, используя вероятность отказа. Среднее количество времени до отказа определяется следующим образом:

$$A = \frac{E_s}{N_s},$$

Где:

$$E_s = \frac{1}{p_e}.$$

При этой оценке рассмотрим класс A и класс B протоколов сегментирования. Для всех этих протоколов выбираем гипергеометрическое распределение для расчета вероятности отказа для одного сегмента, а затем и для каждой эпохи. Вероятность отказа для одного сегмента для протоколов

класса *A* (безопасность одного сегмента равна 33%) с использованием кумулятивного гипергеометрического распределения вычисляется следующим образом:

$$H(K,N,n,\frac{n}{3}) = \sum_{k=[\frac{n}{3}]}^{n} \frac{\left(\frac{K}{k}\right)\left(\frac{N-K}{n-k}\right)}{\left(\frac{N}{n}\right)}.$$

Аналогичным образом можно выразить вероятность отказа для класса *B*, используя гипергеометрическое распределение:

$$H(K,N,n,\frac{n}{2}) = \sum_{k=[\frac{n}{3}]}^{n} \frac{\left(\frac{K}{k}\right)\left(\frac{N-K}{n-k}\right)}{\left(\frac{N}{n}\right)}.$$

Вероятность отказа одного сегмента для класса *A* с использованием кумулятивного биномиального распределения вычисляется следующим образом:

$$P(X \ge \frac{n}{3}) = \sum_{k=[\frac{n}{3}]}^{n} \left(\frac{n}{k}\right) p^k (1-p)^{n-k}.$$

Аналогичным образом, кумулятивное биномиальное распределение для класса *B* равно:

$$P(X \ge \frac{n}{2}) = \sum_{k=[\frac{n}{2}]}^{n} \left(\frac{n}{k}\right) p^k (1-p)^{n-k}.$$

Рассчитаем теперь безопасность сети путем вычисления границы вероятности отказа с использованием оценок Чебышева, Хеффдинга и Шваталя.

При использовании оценки Чебышева вероятность отказа одного сегмента для протоколов класса *A* с использованием (3) будет ограничена, как:

$$H(K,N,n,\frac{n}{3}) \le \frac{np(1-p)(N-n)}{(N-1)(\frac{n}{3}-np)^2}.$$

Вероятность отказа одного сегмента для класса B будет ограничена, как:

$$H(K,N,n,\frac{n}{2}) \le \frac{np(1-p)(N-n)}{(N-1)(\frac{n}{2}-np)^2}.$$

Учитывая, что у всех протоколов в классе A имеются одинаковые вероятности начальной загрузки, то, используя (5), можно оценить вероятность отказа для каждой эпохи следующим образом:

- для класса A:

$$p_{bootstrap} + n_c H(K,N,n,\frac{n}{3}) \le 2^{-26.36} + n_c \frac{np(1-p)(N-n)}{(N-1)(\frac{n}{3}-np)^2},$$

- для класса B:

$$p_{bootstrap} + n_c H(K,N,n,\frac{n}{2}) \le 2^{-26.36} + n_c \frac{np(1-p)(N-n)}{(N-1)(\frac{n}{2}-np)^2}.$$

Аналогичным образом производится оценка Чебышева, соответствующая биномиальной случайной величине. Вероятность отказа для каждой эпохи с использованием (4) имеет вид:

- для класса A:

$$p_{bootstrap} + n_c H(K,N,n,\frac{n}{3}) \le 2^{-26.36} + n_c \frac{np(1-p)}{(\frac{n}{3}-np)^2},$$

- для класса B:

$$p_{bootstrap} + n_c H(K,N,n,\frac{n}{2}) \le 2^{-26.36} + n_c \frac{np(1-p)}{(\frac{n}{2}-np)^2}.$$

При использовании оценки Хеффдинга вероятность отказа для одного сегмента по протоколам класса *A* с учетом (6) ограничивается следующим образом:

$$H(K,N,n,\frac{n}{3}) \le G(x),$$

где:

$$x = \frac{1}{3} - p \quad (p \le \frac{1}{4}).$$

Используя (7), в случае биномиального распределения получаем

$$B(n,p,\frac{n}{3}) \le G(x).$$

В таком случае оценка вероятности отказа для одной эпохи протоколов класса *A* можно вычислить следующим образом:

$$p_0 + n_c p_c \le v(x).$$

При использовании оценки Шваталя вероятность отказа границы для одного сегмента для класса *B* может быть вычислена, как:

$$H(K,N,n,\frac{n}{2}) \le F(x),$$

где:

$$F(x) = \exp^{-2x^2 n}.$$

Следовательно, оценку вероятности отказа для сегмента класса *B* можно выразить следующим образом:

$$p_0 + n_c p_c \le U(x),$$

где:

$$U(x) = 2^{-26.36} + n_c F(x), \quad n_c = \frac{N}{n}.$$

Аналогичным образом, можем оценить вероятность отказа для каждого сегмента / эпохи для класса *A* с использованием границы Шватала.

Таким образом, в настоящей статье представлена новая модель оценки безопасности корпоративной сети, в которой реализована СРР на основе сегментирования, предусматривающая оценку своевременности и полноты обмена информационными ресурсами. В частности, предложены три оценки вероятности отказа для одного сегмента и после каждой эпохи, когда используются гипергеометрическое и биномиальное распределения. Кроме того, предложен подход, определяющий условия, которые должны быть удовлетворены протоколами СРР на основе сегментирования, чтобы вероятность отказа была меньше заданного порога. Учитывая порог вероятности отказа, модель позволяет определить среднее количество лет, в течение которых сеть не будет выходить из строя.

В качестве направления дальнейших исследований планируется интеграция разработанной модели с математическим и программным обеспечением автоматизированной системы управления корпоративной информационной системой.

Литература

1. Саенко И.Б., Фабияновский И.Н., Николаев В.В., Ясинский С.А. Построение модели функционирования распределенной информационной системы на основе блокчейн-технологии // Информация и космос. 2020. №4. С. 73-78.

2. Козленко А.В., Авраменко В.С., Саенко И.Б., Кий А.В. Метод оценки уровня защиты информации от НСД в компьютерных сетях на основе графа защищенности // Труды СПИИРАН. 2012. №2(21). С. 41-55.

3. Rabin M. Probabilistic algorithms //Algorithms and Complexity. N. Y.-London: Acad. Press. 1976. pp.21-39.

4. Rabin M. Rondamized byzantine generals // Proc. 24th Symp. on Foundations of Computer Sci. USA. 1983. pp. 393-402.

5. Kotenko, I.V., Saenko, I.B., Kotsynyak, M.A., Lauta, O.S. Assessment Of Cyber-Resilience Of Computer Networks Based On Simulation Of Cyber Attacks By The Stochastic Networks Conversion Method // SPIIRAS Proceedings. 2017. 6(55). pp. 160–184.

6. Robinson P. Hyland-Wood D., Saltini R., Johnson S., Brainard J. Atomic Crosschain Transactions for Ethereum Private Sidechains. URL:semanticscholar.org/paper/Atomic-Crosschain-Transactions-for-Ethereum-Private-Robinson-Hyland-Wood/9a0889a3fd116595a697a180c852817de5caaa50.

7. Rosenfeld M. Analysis of bitcoin pooled mining reward systems// arXiv preprint arXiv: 1112.4980. 2011.

8. Sapirshtein A., Sompolinsky Y. and Zohar A. Optimal selfish mining strategies in Bitcoin // arXiv preprint arXiv: 1507.06183. 2015.

9. Schlichting R., Schneider F. Fault-stop processes: an approach to designing faulttolerant computing systems // ACM Trans. Comput. Syst. 1983. V. 1. № 3. pp. 222-238.

10. Sompolinsky Y and Zohar A. Bitcoin’s security model revisite // arXiv preprint. arXiv: 1605.09193. 2016.

11. Sompolinsky Y., Zohar A. Secure High-Rate Transaction Processing in Bitcoin. Financial Cryptography and Data Security // 19th International Conference. 2015. pp. 26-30.

References

1. Saenko I.B., Fabijanovskij I.N., Nikolaev V.V., Jasinskij S.A. Informacija i kosmos. 2020. №4. pp. 73-78.

2. Kozlenko A.V., Avramenko V.S., Saenko I.B., Kij A.V. Trudy SPIIRAN, 2012, №2(21). pp. 41-55.

3. Rabin M. Probabilistic algorithms Algorithms and Complexity. N. Y.-London: Acad. Press. 1976. pp. 21-39.

4. Rabin M. Rondamized byzantine generals Proc. 24th Symp. on Foundations of Computer Sci. USA. 1983. pp. 393-402.

5. Kotenko, I.V., Saenko, I.B., Kotsynyak, M.A., Lauta, O.S. Assessment Of Cyber-Resilience Of Computer Networks Based On Simulation Of Cyber Attacks By The Stochastic Networks Conversion Method SPIIRAS Proceedings. 2017. 6(55). pp. 160–184.

6. Robinson P. Hyland-Wood D., Saltini R., Johnson S., Brainard J. Atomic Crosschain Transactions for Ethereum Private Sidechains. URL:semanticscholar.org/paper/Atomic-Crosschain-Transactions-for-Ethereum-Private-Robinson-Hyland-Wood/9a0889a3fd116595a697a180c852817de5caaa50.

7. Rosenfeld M. Analysis of bitcoin pooled mining reward systems. arXiv preprint. arXiv: 1112.4980. 2011.

8. Sapirshtein A., Sompolinsky Y. and Zohar A. Optimal selfish mining strategies in Bitcoin. arXiv preprint. arXiv: 1507.06183. 2015.

9. Schlichting R., Schneider F. Fault-stop processes: an approach to designing faulttolerant computing systems ACM Trans. Comput. Syst. 1983. V. 1. № 3. pp. 222-238.

10. Sompolinsky Y and Zohar A. Bitcoin’s security model revisite arXiv preprint arXiv: 1605.09193. 2016.

11. Sompolinsky Y., Zohar A. Secure High-Rate Transaction Processing in Bitcoin. Financial Cryptography and Data Security 19th International Conference. 2015. pp. 26-30.