© Atayan B., Baghdasaryan T. Data backup security in cloud storage system // Ukrainian Scientific Journal of Information Security, 2016, vol. 22, issue 2, p. 119-122.

БЕЗОПАСНОСТЬ РЕЗЕРВНОГО КОПИРОВАНИЯ ДАННЫХ В ОБЛАЧНОЙ СИСТЕМЕ ХРАНЕНИЯ

Борис Атаян, Татевик Багдасарян

Национальный политехнический университет Армении, Армения

АТАЯН Борис Геннадьевич

Год и место рождения: 1992 год, Ереван, Армения.
Образование: Национальный политехнический университет Армении, 2015 г.
Должность: аспирант 1 года.
Научный интерес: информационная безопасность, облачные системы, криптография, стеганография.
Публикации: около 10 публикаций по теме облачных систем и информационной безопасности.
E-mail: borisn70@gmail.com.

БАГДАСАРЯН Татевик Араевна

Год и место рождения: 1990 год, Ереван, Армения.
Образование: Национальный Политехнический Университет Армении, 2013 г.
Должность: аспирант 3 года.
Научный интерес: информационная безопасность, системы распределения информации, пороговые схемы, криптография, стеганография, облачные системы.
Публикации: более 10 публикаций по теме пороговых схем, стеганографии и криптографии.
E-mail: tatbagg@gmail.com.

Аннотация. *Разработана облачная система резервного копирования данных, которая предоставляет возможность эффективного создания и защищенного хранения в облаке, а также восстановления резервных копий. Для архивации данных в системе используется новый производительный формат хранения данных SGBP, основанный на алгоритме сжатия DEFLATE. Разработанный формат предоставляет возможность быстрого создания архивов, которые могут содержать значительное количество файлов. В статье приведено сравнение формата SGBP с распространённым форматом ZIP, а также проанализированы современные подходы защиты резервных копий. Резервные архивы SGBP защищены методом симметричного шифрования алгоритмом AES. Поскольку при потере, искажении или утечке ключа доступ к резервному архиву будет утерян, то очевидной становится проблема защиты и управления ключей архивов. Защита пользовательских ключей шифрования архивов выполнена с помощью облачной системы распределенного хранения данных, которая основана на алгоритме распределения Шамира. После распределения отдельные части ключа сохраняются на разных облачных сервисах (напр. Dropbox, Google Drive, Sky Drive и т.д.). В систему также включены модули проверки истинности и обновления частей ключа без процедуры восстановления.*

Ключевые слова: *резервное копирование, облачная система, защита резервных копий, распределение ключей, схема Шамира, безопасное обновление, проверка истинности.*

Введение

В настоящее время сложилась ситуация, когда параллельно научно-техническому прогрессу и развитию информационных технологий экспоненциально растет объем данных в компьютерных системах. Согласно статистическим данным о больших данных от UNECE [1], тенденция роста данных не вызывает сомнений и, по прогнозам, к 2019 году объём данных, которые хранятся и обрабатываются в сети Интернет составит примерно сорок петабайт. Необходимость обеспечения безопасности персональных данных в наше время объективная реальность. Информация о человеке всегда имела большую ценность, но сегодня она превратилась в самый дорогой товар. Информация в руках мошенника превращается в орудие

преступления, в руках уволенного сотрудника – в средство мщения, в руках инсайдера – товар для продажи конкуренту и т.д. Именно поэтому персональные данные нуждаются в самой серьезной защите.

Необходимость принятия мер по защите персональных данных вызвана также возросшими техническими возможностями по копированию и распространению информации. Уровень информационных технологий достиг того предела, когда самозащита информационных прав уже не является эффективным средством против посягательств на частную жизнь. Современный человек уже физически не способен скрыться от всего многообразия явно или неявно применяемых в отношении него технических устройств сбора и технологий обработки данных о людях. Появились и эффективно используются злоумышленниками средства интеграции и быстрой обработки персональных данных, создающие угрозу правам и законным интересам человека. Что касается доступности данных, становится весьма актуальной проблема создания и надежного сохранения резервных копий этих данных.

Как и во всех остальных информационных системах, здесь также имеются проблемы. Одной из основных задач является создание и хранение резервных копий данных особо нуждающихся в защите от несанкционированного доступа. При этом важнейшей задачей является обеспечение защиты резервных копий этих конфиденциальных данных. В той же степени важно быстродействие при создании резервных копий, их относительно небольшой объем и эффективность системы при многократном копировании. С учетом вышеуказанного, **целью** работы является разработка защищенной системы резервного копирования основанной на облачных технологиях, которая решит описанные проблемы. Облачная система резервного копирования данных предоставит возможность эффективного создания и защищенного хранения резервных копий в облаке, а также их восстановления.

Облачная система резервного копирования

В наши дни хранение данных невозможно представить без их сжатия. Поскольку объем резервируемых данных может быть достаточно большим, то очевидно, что при ограниченных ресурсах резервные копии необходимо хранить в максимально компактном виде. Сжатие можно организовать с использованием определенного файлового формата, который обеспечит группирование данных, и дальнейшее их сжатие.

После исследования файлового ZIP формата [1], обозначились серьезные проблемы, которые значительным образом препятствуют использованию данного формата для хранения больших объемов резервных копий: ограничение на количество файлов хранящихся в резервном архиве; значительное увеличение заголовка архивного файла, что приводит к замедлению процесса восстановления данных.

Вышеуказанные проблемы чрезвычайно актуальны при резервном копировании особо больших объемов данных. Поскольку ZIP формат не подходит для использования в облачной системе резервного копирования, реализован новый формат SGBP, который больше приспособлен для использования в процессе создания и хранения резервных копий больших объемов данных, а так же их восстановления, в том числе и поэтапного резервирования и восстановления. Формат основан на алгоритме сжатия DEFLATE и имеет следующую структуру: Files **–** данные в сжатом виде; Footer **–** предназначен для хранения информации о CDR (Central Directory Record); CDR **–** в котором содержится детальная информация о резервируемых данных.

CDR запись содержит следующие поля: CRC - Cyclic Redundancy Check, значение для проверки целостности файлов в архиве; Filename **–** имена файлов в архиве; Filename length **–** длина имен файлов в архиве; Offset to the start of file **–** указатель на начало архивированного файла; Compressed file length **–** длина сжатого файла; Uncompressed file length **–** длина оригинального файла.

Как видно из структуры CDR записи, количество файлов хранящихся в резерве не ограничено, и, в отличии от формата ZIP (ZIP32 и ZIP64 [2]), не хранит избыточные или дублирующие данные в заголовке архива.

Защита резервных копий

Как выше было указано, основной задачей в облачной системе резервного копирования является защищенное создание и хранение резервных копий конфиденциальных данных. В настоящее время существуют и широко практикуются следующие подходы защиты резервных копий: аутентификация пользователей в системе резервного копирования; списки контроля доступа на основе ролей для всех операций резервного копирования и восстановления данных; шифрование данных при передаче и хранении резервных копий; защита сервера системы резервного копирования; физическая защита хранилищ резервных копий. В облачной системе резервного копирования все резервные копии (архивы SGBP) шифруются симметричным блочным шифром AES с ключом 256 бит (рис. 1).

Рис. 1. Защищенное хранение архива SGBP

При использовании метода шифрования, возникает необходимость управления ключами. Следовательно, сам ключ архива нуждается в особой защите, при потере, искажении или утечке которого конфиденциальные данные в резервном архиве раскроются; или пользователь вовсе потеряет резервный архив, поскольку при потере ключа расшифровать данные станет невозможно. Проблема управления ключей резервных архивов решается

методом облачного распределенного хранения данных.

Защита ключей архивов SGBP организована при помощи схем распределения секрета. Протоколы распределения секрета [3], которые еще называются схемы распределения, по своей сути являются многосторонними протоколами, функция которых заключается в установлении секретных ключей или паролей. Под установлением ключей подразумевается процесс, в результате исполнения которого общий секрет (в нашем случае ключ) становится доступным определенному количеству субъектов, что позволяет обеспечить необходимый уровень криптографической защиты. В облачной системе резервного копирования используется схема Шамира [4] для распределения ключа архива. Выбор, в качестве алгоритма разделения секрета схемы Шамира, обусловлен следующими причинами: совершенная безопасность (Perfect Security) [4] – информационная теоретическая безопасность. При наличии t частей секрета возможно единственным образом восстановить S. Имея доступ к $t-1$ или менее частям секрета и зная диапазон $0 \le S \le m-1$, восстановление секрета остается сложной задачей; идеальность схемы – битовый размер секрета совпадает с размером каждой из составляющих частей секрета; независимость – в отличие от многих криптографических схем, безопасность схемы разделения информации напрямую не зависит от сложности информации; свойство гомоморфизма – для схемы Шамира имеет место (+,+) гомоморфизм. Например, предположим, что есть два секрета S и R. Они распределены по схеме разделения Шамира: $(f(1),\ldots,f(n))$, определенные из полинома $f(x)$, а $(g(1),\ldots,g(n))$ – из полинома $Å(x)$ для S и R соответственно. Допустим, для каждой i части протокола просуммируется. Каждая из полученных сумм в свою очередь является частью секрета $S+R$, определяемого из полинома $h(x)=f(x)+g(x)$ и $h(0)=S+R$.

Важно, что части распределенного ключа хранятся не локально, а в облаке, одновременно используя множество облачных сервисов хранения (Google, Dropbox и т.д.) (рис. 2), что повышает защищенность ключа, поскольку вероятность потери, искажения или утечки отдельных частей ключа снижается.

На рис. 2. показан пример распределения ключа архива по схеме (3, 5), а именно, ключ шифрования архива SGBP распределяется на 5 частей, любые 3 из которых смогут однозначно восстановить ключ, при этом, меньшее количество частей не предоставляет никакой информации об исходном ключе. После успешного распределения, части ключа сохраняются в разных облачных контейнерах - используя разные облачные сервисы хранения. Преимущество этого метода состоит в том, что даже при потере частей ключа (до 2-х экземпляров), является возможным восстановить исходный ключ целыми частями. И, поскольку отдельные части ключа не предоставляют никакой информации об исходном, то раскрытие отдельных частей не достаточно, чтобы получить исходный ключ. Еще одним преимуществом является то, что при использовании нескольких независимых отдельных облачных сервисов вероятность потери, искажения или утечки отдельных частей ключа снижается.

В результате разработанная система дает возможность распределять и хранить криптографические ключи резервных данных в глобальных сетях в частности в сети Интернет. Так как, Интернет является открытой сетью, нельзя с уверенностью утверждать, что данные хранящиеся в сети полностью защищены, даже при использовании схемы распределения.

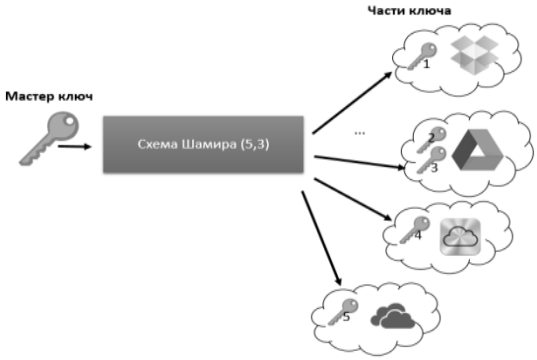

Рис. 2. Пример распределения ключа по схеме (3, 5)

Для повышения безопасности разработанной системы резервного копирования данных, в системе распределения предусмотрены 2 необходимых модуля: модуль проверки истинности частей разделяемого секрета; модуль полного обновления частей секрета, без восстановления самого секрета.

Проверка истинности частей разделенного секрета

Как и в других криптографических системах, так и в системах разделения секрета всегда есть вероятность, что секретная информация может быть скомпрометирована противником. Для обеспечения стойкости в разработанной системы используется алгоритм проверки подлинности частей секрета, с которым возможна проверка каждой части секрета на подлинность и целостность (например, перед процессом восстановления). Алгоритм основывается на следующих действиях.

Как известно в схеме Шамира выбирается любой многочлен $Q(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_t x^t$ степени t, где секрет $S = a_0$. Выбираются любые простые числа p и q, таким образом, что $p = 2*q+1$. Эти числа не секретны. Выбирается число g так, что $g^q mod\ p = 1.$ Вычисляется значение $r_i = g_i^q \,(mod\ p), i = 0,1,2,\ldots,t$ для каждой части $(r_0, r_1, \ldots, r_t)$. Для любых чисел $j = 1,2,\ldots,n$ вычисляются значения частей секрета $S_j = Q(j)$. Для каждой части секрета проверяется уравнение

$$g^{S_j} = r_0 \cdot (r_1)^j \cdot (r_t)^{j^t} \cdot (mod\ p)\ \ g^{S_j} = r_0 \cdot (r_1)^j \cdot (r_t)^{j^t}\ (mod\ p).$$

Для того, чтобы убедится, что проверяемая часть действительно является частью секрета. И действительно, имеет место равенство

$r_0 \cdot (r_1)^j \cdot (r_t)^{j^t} = g^{a_0} \cdot \ldots \cdot g^{a_t \cdot j^t} = g^{a_0 \cdot j + \ldots + a_t \cdot j^t} = g^{Q(j)} (mod\ p)$.

Модуль обновления частей секрета без восстановления самого секрета. Цель данного модуля состоит в том, чтобы пользователям предоставлялась возможность обновить части секрета без процесса восстановления истинного секрета. Благодаря алгоритму обновления получаются новые части секрета, не изменяя секрета, т.е. оставив свободный член многочлена неизменным. Одно из условий схемы восстановления частей является необходимость доступности каждой имеющейся части разделенного секрета. Данная схема основывается на следующих действиях: для каждой $i, i \in [1, n]$ части секрета из конечного поля Fp произвольно выбираются коэффициенты в количестве $t-1$ и получается многочлен $Pi(x)$, чей свободный член равен 0; на основе полученного многочлена $Pi(x)$ для каждой части, а также и для всех остальных частей, вычисляются новые координаты; для каждой части $i, i \in [1, n]$ получаются координаты $P_1(i), \ldots, P_n(i)$; для каждой части суммируются координаты, полученные от других частей и старой части. В итоге получается новая часть секрета. Алгоритм основывается на свойстве гомоморфизма схемы Шамира; каждая $i, i \in [1, n]$ старая часть секрета удаляется.

Выводы

Итак, система облачного резервного копирования данных предоставляет возможность эффективного создания и защищенного хранения резервных копий в облаке, где данные хранятся в архивах формата SGBP. Для обеспечения необходимого уровня защиты резервных данных, архивы SGBP хранятся в облаке в зашифрованном виде. Проблема защиты криптографических ключей зашифрованных архивов решена с помощью облачной системы распределенного хранения данных, обеспечивая должную защиту ключей от утечек, потерь и искажения. Для надежного хранения частей ключей, система распределенного хранения предоставляет возможность проверки истинности частей разделенного ключа, а также периодическое безопасное обновление частей без восстановления исходного ключа.

Литература

[1] Big Data in Official Statistics Vulnerabilities - [Electronic resource] / Mountain View, 2016 - Access mode: World Wide Web. – URL: http://www1.unece.org/stat/platform/display/bigdata/Big+Data+in+Official+Statistics.

[2] ZIP File Format Specification, ver. 6.3.4. - Access mode: World Wide Web. – URL: https://pkware.cachefly.net/webdocs/casestudies/APPNOTE.TXT.

[3] Schneier B. Applied Cryptography – 1994. – P. 34.

[4] Shamir A. How to Share a Secret, Communications of the ACM. – Vol. 22, Issue 11, 1979. – P. 612-613.

УДК 004.056.3 (045)

Атаян Б.Г., Багдасарян Т.А. Безпека резервного копіювання даних в хмарній системі збереження

Анотація. *Розроблена хмарна система резервного копіювання даних, яка надає можливість ефективного створення і захищеного зберігання в хмарі, а також відновлення резервних копій. Для архівації даних в системі використовується новий продуктивний формат зберігання даних SGBP, заснований на алгоритмі стиснення DEFLATE. Розроблений формат надає можливість швидкого створення архівів, які можуть містити значну кількість файлів. У статті наведено порівняння формату SGBP з поширеним форматом ZIP, а також проаналізовані сучасні підходи захисту резервних копій. Резервні архіви SGBP захищені методом симетричного шифрування алгоритмом AES. Оскільки при втраті, спотворенні або витоку ключа доступ до резервного архіву буде втрачено, то очевидною стає проблема захисту і управління ключів архівів. Захист користувальницьких ключів шифрування архівів виконана за допомогою хмарної системи розподіленого зберігання даних, яка заснована на алгоритмі розподілу Шаміра. Після розподілу окремі частини ключа зберігаються на різних хмарних сервісах (напр. Dropbox, Google Drive, Sky Drive тощо). У систему також включені модулі перевірки істинності і поновлення частин ключа без процедури відновлення.*

Ключові слова*: резервне копіювання, хмарна система, захист резервних копій, розподіл ключів, схема Шаміра, безпечне оновлення, перевірка істинності.*

Atayan B., Baghdasaryan T. Data backup security in cloud storage system

Abstract. *Cloud backup system is proposed, which provides means for effective creation, secure storage and restore of backups in Cloud. For data archiving new efficient SGBP file format is being used in the system, which is based on DEFLATE compression algorithm. Proposed format provides means for fast archive creation, which can contain significant amounts of data. Modern approaches of backup archive protection are described in the paper. Also the SGBP format is compared to heavily used ZIP format (both ZIP32 and ZIP64). SGBP backup archives are protected using symmetric encryption AES algorithm. Because in case of loss, distortion or leak of the encryption key backup archive will become inaccessible, there is no doubt that the SGBP encryption keys need special protection and management. The cryptographic keys of SGBP backups are protected using distributed cloud storage system, which is based on Shamir's threshold secret sharing scheme. After secret sharing all individual secret shares are distributed and saved to different cloud storage services (e.g. Dropbox, GoogleDrive, SkyDrive etc.). Modules for key share validity check and share update without key reconstruction are included in the system.*

Key words*: backup, cloud system, backup security, key sharing, Shamir scheme, secure update, validity check.*

Отримано 2 *травня 2016 року*, затверджено редколегією *16 травня 2016 року*