УДК 004.415

МЕТОД КОНТРОЛЯ ПРЯМОГО ДОСТУПА К СЕМАНТИЧЕСКИМ БАЗАМ ДАННЫХ

Хоанг Ван Кует, А.Ф. Тузовский

Томский политехнический университет
E-mail: student8050@sibmail.com

Введены основные понятия, методы определения и создания меток безопасности триплетов семантических данных. Разработаны алгоритмы для формирования покрытия безопасности семантических данных и обеспечения безопасности прямого доступа пользователей к базам данных.

Ключевые слова:
Метка безопасности, покрытие безопасности, несанкционированный доступ.

Key words:
Security label, security cover, unauthorized access.

В связи с активным развитием такого нового направления информатики, как семантические технологии, создаются большие объёмы семантических данных, которые отличаются от реляционных данных тем, что на их основе могут выполняться логические выводы, позволяющие формировать новую информацию [1]. Использование таких данных в работе информационных систем требует решения следующих задач обеспечения их безопасности:

1) контроль прямого доступа пользователей к различным частям данных, определённых политикой доступа к данным для каждого пользователя (управление прямым доступом к данным);
2) контроль логических выводов на доступных данных, на основе которых могут быть логически выведены данные, доступ к которым пользователю не разрешён (управление возможными логическими выводами).

В данной статье рассмотрены метод определения меток безопасности RDF-триплетов данных (RDF – *Resource Description Framework*) и метод определения уровней безопасности RDF-экземпляров данных семантических баз данных для решения задачи контроля прямого доступа.

Постановка задачи контроля прямого доступа к семантическим базам данных

Угроза безопасности информации – это потенциально возможное воздействие на информацию, которое прямо или косвенно может нанести урон пользователям или владельцам информации. При классификации угроз выделяют три основных свойства безопасности информации: конфиденциальность, целостность, доступность [2]. В соответствии с этими свойствами безопасности информации различают три классические угрозы безопасности информации:

- Угроза конфиденциальности информации состоит в нарушении установленных ограничений на доступ к информации.
- Угроза целостности информации – это несанкционированное изменение информации, случайное или преднамеренное.
- Угроза доступности информации осуществляется, когда несанкционированно блокируется доступ к информации (блокирование может быть постоянным или на некоторое время, которое достаточно, чтобы информация стала бесполезной).

В целях обеспечения безопасности работы информации определены основные виды политик: дискреционная политика безопасности, политика безопасности информационных потоков, политика ролевого разграничения доступа, мандатная политика безопасности и политика изолированной программной среды [3].

Основная цель мандатной политики безопасности – предотвращение утечки информации от объектов с высоким уровнем доступа к объектам с низким уровнем доступа, т. е. противодействие возникновению в информационной системе неблагоприятных информационных потоков сверху вниз. Чаще всего мандатную политику безопасности описывают в терминах, понятиях и определениях свойств модели Белла Ла Падула. По сравнению с информационными системами, построенными на основе дискреционной политики безопасности, для систем, реализующих мандатную политику, характерна более высокая степень надёжности. Правила мандатной политики безопасности более ясны и просты для понимания разработчиками и пользователями, что также является фактором, положительно влияющим на уровень безопасности системы.

В семантических данных для контроля прямого доступа пользователей к различным частям данных принимается модель системы мандатного разграничения доступа с помощью использования меток безопасности RDF-триплетов и уровней безопасности RDF-экземпляров данных семантических баз данных. Для достижения поставленной цели необходимо описать основные понятия семантических данных.

Определение 1 (язык описания ресурсов). RDF-язык – это разработанная организацией W3C модель для описания метаданных о ресурсах. В основе данной модели лежит идея об использовании

специального вида утверждений, высказываемых о ресурсе. Каждое утверждение имеет вид «субъект–предикат–объект» и в терминологии RDF называется триплетом.

Определение 2 (модель RDF-триплета). Триплет является основным элементом семантических баз данных и состоит из набора троек: субъект (r), предикат (p), объект (v), где каждый компонент является: константой, такой как $r \in R$ (набор ресурсов), $p \in PR$ (набор свойств) и $v \in R \cup L$ (набор ресурсов и литералов), или переменной, представленной в виде символов $ или ?.

Определение 3 (RDF-схема, RDFS). Язык RDFS является семантическим расширением языка RDF и предоставляет средства для описания групп связанных ресурсов и отношений между этими ресурсами. Все определения языка RDFS выражены с помощью средств языка RDF.

Определение 4 (язык описания онтологий – Web Ontology Language (OWL)). OWL – это язык представления онтологий в Web-сети. Фактически это словарь, расширяющий набор терминов, определённых на языке RDFS.

Определение 5 (семантическая база данных). Семантическая база данных состоит из множеств данных, описанных на языках RDF, RDFS, OWL, и представляет собой множество триплетов. Она включает в себя набор ресурсов R, свойств PR, URI ссылок U, пустых узлов B и литералов L. В семантических базах данных выражения $R=U \cup B$, $PR \subset U$ и P, B и L не пересекаются.

Модель контроля прямого доступа пользователей к семантическим данным

Предлагаемая модель построена на основе мандатной политики безопасности, основанной на мандатном разграничении доступа, определяемом следующими четырьмя условиями:

- Все пользователя и триплеты базы данных однозначно идентифицированы.
- Задана решётка уровней безопасности информации.
- Каждому триплету базы данных присвоен уровень безопасности, определяющий ценность содержащейся в нем информации.
- Каждому пользователю присвоен уровень доступа, определяющий уровень доверия к нему в семантической базе данных.

В семантической базе данных каждый триплет имеет свой уровень безопасности AC, указатель которого может иметь значения из множества меток безопасности L={неклассифицированный (L_U), конфиденциальный (L_C), секретный (L_S) и сверхсекретный (L_{TS})}, где $L_U < L_C < L_S < L_{TS}$. Каждый пользователь имеет уровень доступа AC_s, включающий права выполнения таких операций с данными, как $Rule$={*read*, *write*, *append*, *execute*}, где read – право на чтение триплета, write – права на запись триплета, append – права на запись в конец объекта, execute – права на выполнение добавления или удаления триплета.

В соответствии с мандатной моделью, безопасность базы данных должна обладать следующими свойствами:

- *Свойство простой безопасности («простое» свойство)*: Пользователь может иметь право доступа на чтение триплета только в случае, когда уровень доступа пользователя не ниже уровня безопасности триплета.
- *Свойство «звезда» безопасности (свойство «звезда»)*: Пользователь может иметь доступ к триплету в случае, когда уровень безопасности триплета не ниже его уровня права доступа. Пользователь может иметь право доступа на запись триплета только в случае, когда его уровень доступа равен уровню безопасности триплета. Он может иметь право доступа на чтение триплета только в случае, когда его уровень доступа не ниже уровня безопасности триплета.
- *Свойство дискреционной безопасности (ds-свойство)*: При каждом обращении предоставляется одно право доступа пользователя на триплет.
- *Невозможность полного общения с бездействующим триплетом*: Пользователь не может читать бездействующий триплет.

В связи с этим для семантических баз данных предлагаются следующие основные свойства поддержки безопасности работы с ними:

- *Свойство чтения.* Пользователь может читать триплет только в том случае, когда его уровень доступа не ниже уровня безопасности триплета.
- *Свойство записи.* Только владелец имеет право на запись (модификацию) его данных.
- *Уровень безопасности триплета.* Уровень безопасности триплета должен быть не ниже уровня доступа его владельца.
- *Уровень безопасности элементов триплета.* Уровень безопасности триплета должен быть не ниже уровня безопасности его субъекта, объекта и предиката.
- *Уровень доступа на модификацию триплетов.* Только администратор безопасности базы данных и владелец триплетов имеют право на модификацию уровня безопасности триплетов и уровней доступа пользователей.

Алгоритмы решения задачи

Для решения поставленной задачи разработаны два алгоритма, позволяющие определить уровень безопасности RDF-экземпляров данных и контролировать доступ пользователей к базе данных.

Метка безопасности RDF-триплета данных

В целях безопасности семантических баз данных для каждого компонента триплета могут быть заданы разные права доступа, т. е. они находятся на разных уровнях безопасности согласно своим меткам ($sl_1, sl_2, \ldots, sl_n$), созданным пользователями [4]. Например, sl_1 соответствует открытым данным, sl_2 – конфиденциальным данным, ... , sl_n – сверхсекретным данным.

В триплете (субъект, предикат, объект) каждый из компонентов может иметь одинаковую метку *sl*

или разные метки безопасности (sl_r, sl_p, sl_v) (где sl_r – метка для субъекта, sl_p – метка для предиката, sl_v – метка для объекта). В том случае, когда три элемента имеют одинаковую метку *sl*, то *sl* является меткой безопасности для триплета ($sl=sl_{tr}$, где sl_{tr} – метка для триплета). Если три метки являются разными ($sl_r \neq sl_p \neq sl_v$), то надо образовать наименьшую верхнюю границу значения $sl_{\max}$ этих меток, тогда $sl_{\max}$ является меткой безопасности триплета sl_{tr} ($sl_{tr}=sl_{\max}$).

Один и тот же ресурс может являться субъектом или объектом (иногда является предикатом) в разных ситуациях, следовательно, он может иметь разные метки безопасности ($sl_1,sl_2,...,sl_n$) в зависимости от своего положения. Триплет, обладающий ресурсом в конкретной роли (субъект или объект), должен иметь конкретную метку sl_i, охватывающую соответствующую метку обеспечения ресурса в данной позиции, и обозначаться $sl_i \geq \max(sl_1,sl_2,...,sl_n)$.

Алгоритм 1: создание покрытия безопасности триплетов семантических данных

Определение 6. (модель отображения). Пусть $t=[r,p,v]$ и $t'=[r',p',v']$ являются моделями RDF-триплетов. Модель отображения f: $t \rightarrow t'$, где *r*, *p*, *v* и *r'*, *p'*, *v'* являются константами или переменными, определяется следующим образом:

- *f* отображает одну переменную в другую переменную или в константу;
- *f* отображает константу *e*, находящуюся в D_T (экземпляр данных), в виде такой же константы;
- *f* отображает константу *e*, находящуюся в схеме S_T (схема данных):
  1) в такую же схему констант;
  2) другие схемы констант e_i, подчиняющиеся следующим формам:
     - $[e_i, rdf : type, e]$;
     - $[e_i, rdfs : subPropertyOf, e]$;
     - $[e_i, rdfs : subClassOf, e]$;
     - $[f(r),f[x],f(v)]=[r',p',v']$.

Определение 7 (модель категоризации триплетов). Пусть $t=[r,p,v]$ и $t'=[r',p',v']$ являются моделями RDF-триплетов. Выражение «*t* задаёт категорию (категорирует) *t'*» обозначается как $t' \subseteq t$, если существует функция f: $t \rightarrow t'$. Модель категоризации триплетов является рефлексивной, транзитивной или антисимметричной.

В семантических базах данных каждая модель триплета связана с одной меткой безопасности.

Определение 8 (безопасность RDF-триплета). Пусть S_T является RDF-схемой данных, D_T является RDF-экземплярами и элементы S_T и D_T представляются в виде набора триплетов. Пусть множество всех триплетов в базе данных обозначается как K_T, где $K_T=S_T \cup D_T$. Значение безопасности RDF-триплета определяется как пара $s=(t,sl)$, где $t \in K_T$, $sl \in S_L$, а S_L является множеством меток безопасности семантических баз данных $\{sl_1,sl_2,...,sl_n\}$.

Определение 9 (присвоенная метка безопасности). Пусть *pt* и *t* являются триплетами. Если существует модель отображения f: $pt \rightarrow t$ и *pt* имеет метку безопасность *sl*, то безопасность RDF-триплета будет задаваться как (*t*,*sl*).

Определение 10 (покрытие безопасности для RDF данных). Предполагается, что *S* является множеством значений безопасностей триплетов семантических данных, где $S=\{s_1,s_2,...,s_n\}$, где $s_i=(t_i,sl_i)$, для любой безопасности $s=(t,sl)$ выполняются условия:

1. Не существуют два значения безопасности триплетов (*t*,*sl*) и (*t*,*sl'*), удовлетворяющие выражениям: $t=t'$ и $sl>sl'$.
2. Не существует значения безопасности $s=(t,sl)$, в котором sl является пустой.

Определение 11 (политика безопасности для RDF-данных). Предлагается, что *SP* является множеством безопасностей RDF-триплетов вида (*pt*,*sl*), где *pt* является моделью RDF-триплета, и *sl* – меткой безопасности. Политикой безопасности для RDF-данных является $SP \cup pt_{def}$, где $pt_{def}=([?x_1,?x_2,?x_3], sl_{def})$, x_1, x_2, x_3 являются элементами триплета и sl_{def} является меткой безопасности по умолчанию. Для любых двух пар (*t*,*sl*) и (*t'*,*sl'*), где $t' \subseteq t$, выражение $sl \leq sl'$ выполняется только в случае, когда операция ≤ является доминирующим отношением.

Политика безопасности определяет требования безопасности для примитивных RDF/RDFS утверждений в базе данных и наследованных утверждений из RDF/RDFS-базы данных. Кроме того, политика безопасности по умолчанию также гарантирует, что не существует утверждения RDF/RDFS, которое является неклассифицированным [5].

Таблица. *Модель контроля доступа пользователей*

Модель безопасности	Интерпретация
[*r*,*p*,*v*]	Все элементы триплета являются константами
[*r*,?*x*,*v*]	Субъект и объект являются константами, а предикат – переменной
[*r*,*p*,?*x*]	Субъект и предикат являются константами, а объект – переменной
[?*x*,*p*,*v*]	Объект и предикат являются константами, а субъект – переменной
[*r*,?*x*,?*y*]	Объект и предикат являются переменными, а субъект – константой
[?*x*,*p*,?*y*]	Субъект и объект являются переменными, а предикат – константой
[?*x*,?*y*,*v*]	Субъект и предикат являются переменными, а объект – константой
[?*x*,?*y*,?*z*]	Все элементы являются переменными

Покрытием безопасности *S* является множество всех значений безопасностей триплетов RDF-данных $S=\{s_1,s_2,...,s_n\}$, создающихся путём применения модели отображения и присвоения меток безопасности. На рис. 1 показан алгоритм для создания покрытия безопасности триплетов семантических данных, обладающего следующими свойствами:

- если существует модель отображения от модели к триплету, то метка безопасности триплета доминирует над меткой безопасности модели;
- если триплет имеет метку, не созданную с помощью модели по умолчанию, то должна суще-

ствовать такая модель отображения из модели к триплету, у которой метка безопасности модели доминирует над меткой безопасности триплета.

Данный алгоритм применяется для моделей триплетов, определённых в таблице, где любой элемент из триплета $[r,p,v]$ может быть переменной или константой. В соответствии с данным алгоритмом, входными данными алгоритма являются политика безопасности $SP=\{sp_1,...,sp_n\}\cup([x_1,x_2,x_3],sl_{def})$, где $sp_i=(pt_i,sl_i)$, и RDF база данных $K_T=D_T\cup S_T=\{t_i,...,t_k\}$. Выходными данными является покрытие безопасности $S=\{s_1,s_2,...,s_k\}$, где $s_i=(t_i,sl_i)$, $i=1,...,k$.

Данный алгоритм гарантирует, что все триплеты семантических баз данных получат свои метки безопасности.

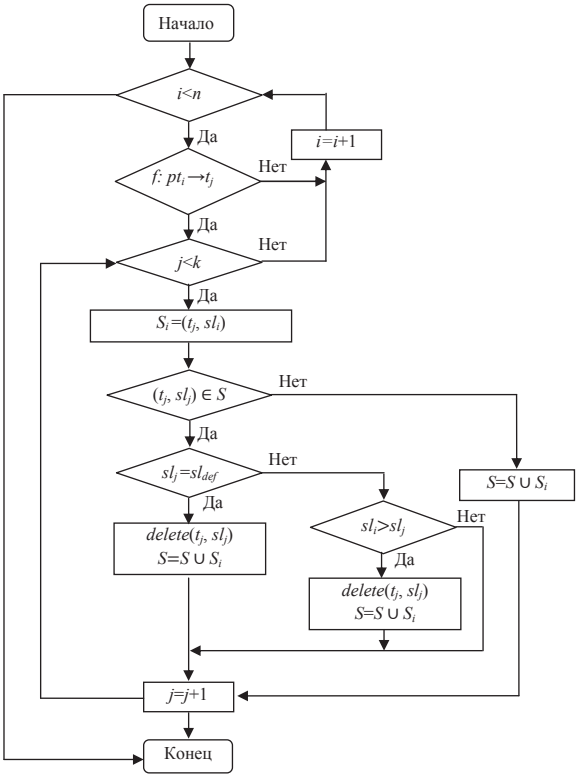

Рис. 1. *Алгоритм определения покрытия безопасности триплетов семантических данных*

Алгоритм 2: контроль прямого доступа пользователей к базе данных

Под контролем прямого доступа пользователей понимается возможность получать данные в соответствии с заданными для них правами на использование данных.

В семантических базах данных хранятся разные категории информации, такие как экономическая, финансовая и т. п. [6]. Если уровень безопасности информации AC содержит только один указатель с меткой безопасности из множества L, то пользователь, имеющий уровень доступа типа $ACs=L_C$ в категории M, не может иметь доступ в категории N к данным, имеющим уровень безопасности $AC=L_C$. В данном случае количество возможных вариантов создания уровней безопасности информации в базе данных будет сильно ограниченным. Для решения данной проблемы в уровни безопасности данных могут быть добавлены другие указатели с метками безопасности из множества L.

Предполагается, что метка безопасности RDF-триплета $AC=(S,P,PS,C)$ (уровень безопасности триплета) включает следующие 4 показателя:

- *чувствительность S* определяет уровень значимости или важности связи (предиката), а также её уязвимости перед несанкционированным лицом;
- *приватность P* определяет права владельца на возможность передачи данной информации другим пользователям;
- *персональная безопасность PS* определяет уровень защиты персональной информации человека или организации;
- *конфиденциальность C* задаёт возможность совместного использования данной информации с другими ресурсами.

Каждый показатель может принимать значения из множества L (в данной работе предлагается, что эти показатели могут принимать значение 0 (unclassified), 1 (confidential)).

Право доступа пользователей к данным $AC_s=(S_s,P_s,PS_s,C_s)$ также включает 4 показателя: чувствительность S, приватность P, персональная безопасность PS и конфиденциальность C. Для контроля доступа пользователя к RDF-триплету необходимо сравнивать уровень права доступа пользователей $AC_s=(S_s,P_s,PS_s,C_s)$ с уровнем безопасности триплета $AC=(S,P,PS,C)$.

Считается, что $AC_s=(S_s,P_s,PS_s,C_s)$ не ниже уровня $AC=(S,P,PS,C)$ только в случае, когда $S_s\geq S$, $P_s\geq P$, $PS_s\geq PS$, $C_s\geq C$. Если $AC_s\geq AC$, то пользователи могут получить доступ к этим данным. Если $AC_s<AC$, то пользователи не имеют права доступа к этим данным и данные являются невидимыми для пользователей. На рис. 2 показан алгоритм контроля прямого доступа к RDF-триплету данных.

В соответствии с этим алгоритмом проверка прав доступа пользователя осуществляется на каждом триплете. Входными данными алгоритма являются: уровень права доступа пользователя и множество индексов чувствительности RDF-триплетов данных. Выходными данными является множество всех триплетов, к которым пользователь имеет право доступа. На каждом шаге происходит процесс сравнения каждого критерия индекса чувствительности с уровнем права доступа пользователя.

Данный алгоритм разработан на основе традиционных алгоритмов управления доступом пользователей к данным. Его достоинствами являются простота и небольшое количество операций, а также возможность обеспечения безопасности каждого триплета данных (как столбца в реляционных базах данных), что позволяет контролировать доступ пользователей к различным частям данных.

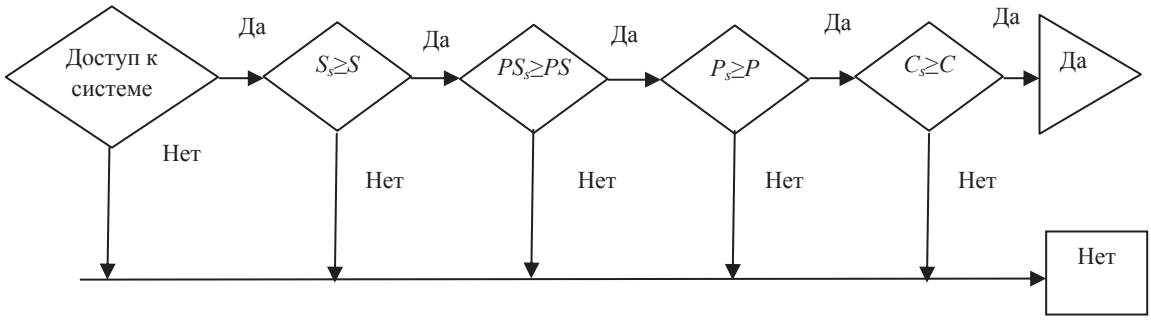

Рис. 2. *Алгоритм контроля прямого доступа пользователей к базе данных*

Заключение

Задача обеспечения безопасности семантических баз данных является более сложной, чем обеспечение безопасности реляционных баз данных, для которых уже разработано большое количество методов. Сложность обеспечения безопасности семантических баз данных заключается в требовании контроля доступа пользователей к разным частям данных (к каждому триплету).

Данная задача решается путём применения вышеописанной модели мандатного разграничения доступа с использованием алгоритмов определения меток безопасности RDF-триплетов данных и уровней безопасности семантических данных, а также алгоритма контроля доступа пользователя к базе данных.

СПИСОК ЛИТЕРАТУРЫ

1. Reddivari P. Policy based Access Control for a RDF Store // Proceedings of the Policy Management for the Web Workshop. – 2005. – V. 120. – № 5. – P. 78–83.
2. Грушо А.А., Тимонина Е.Е. Теоретические основы защиты информации. – М.: Изд-во Агентства «Яхтсмен», 1996. – 250 с.
3. Носов В.А. Основы теории алгоритмов и анализа их сложности. – М.: МГУ, 1992. – 268 с.
4. Stachour P. Design of LDV: A multilevel secure relational database management system IEEE Trans // Knowledge and Data. – 1990. – V. 2. – № 2. – P. 190–209.
5. Тузовский А.Ф., Ямпольский С.В. Системы управления знаниями (методы и технологии) / под общ. ред. В.З. Ямпольского. – Томск: Изд-во НТЛ, 2005. – 260 с.
6. Девянин П.Н. Анализ безопасности управления доступом и информационными потоками в компьютерных системах. – М.: Радио и связь, 2006. – 176 с.

Поступила 08.10.2012 г.