P-ISSN: 2548-1932
e-ISSN: 2549-7758

KEAMANAN BASIS DATA BERDASARKAN TEORI HIMPUNAN

Novianti Indah Putri[1], Yudi Herdiana[2], Zen Munawar[3], Dadad Zainal Musadad[4]

[1,2]Teknik Informatika, Fakultas Teknologi Informasi, Universitas Bale Bandung
[3,4]Manajemen Informatika, Politeknik LP3I Bandung
[1]noviantiindahputri2021@gmail.com, [2]ydherdn@gmail.com, [3]munawarzen@gmail.com, [4]dadadzainal@gmail.com

ABSTRAK

Banyaknya data yang tersimpan dalam basis data maka penting bagi organisasi untuk mengamankan data dalam basis data. Organisasi perlu memastikan keamanan dan kerahasiaan data, oleh sebab itu organisasi mengimplementasikan aplikasi dan sistem yang menyediakan layanan, fungsi, dan alat untuk pengelolaan dan pemeliharaan data. Penelitian ini menyampaikan masalah keamanan yang muncul dalam basis data yang tidak tepat berdasarkan teori himpunan. Dalam penelitian ini bertujuan menyelidiki area keamanan untuk basis data kasar, yang memiliki masalah keamanan yang mirip dengan basis data statistik. Aspek keamanan yang dipertimbangkan mirip dengan database statistik yang kombinasi kuerinya tidak dapat mengungkapkan nilai atribut yang tepat. Langkah-langkah teori informasi digunakan untuk mengkarakterisasi keamanan untuk database yang tidak tepat. Hasil penelitian telah menunjukkan bagaimana sifat basis data relasional kasar menyediakan beberapa keamanan yang melekat melalui penggunaan struktur non-first bentuk normal.

Kata Kunci: *Keamanan Basis Data; Teori Informasi; Entropi; Database Relasional*

I. PENDAHULUAN

Basis data terus tumbuh dalam ukuran dan kompleksitas, dan digunakan dalam berbagai aplikasi yang beragam. Ada hal penting untuk melakukan pekerjaan dengan baik dalam pencegahan keamanan jaringan komputer, untuk meminimalkan kemungkinan terjadinya kejahatan komputer [1]. Untuk penggunaan basis data aplikasi di dunia nyata, maka perlu untuk memasukkan beberapa jenis manajemen ketidakpastian ke dalam model data yang mendasarinya. Salah satu karakteristik dari banyak basis data yang tidak tepat adalah dengan mengizinkan kumpulan nilai dalam tupel. Hal ini disebut sebagai bentuk non-first atau basis data bersarang [2]. Jika nilai suatu atribut adalah non-atomik, yaitu bernilai himpunan, maka ada ketidakpastian tentang salah satu nilai dalam himpunan yang sesuai dengan atribut, atau apakah himpunan tersebut lebih dari satu. Terdapat aspek-aspek tertentu dalam model basis data yang tidak pasti yang berbeda tetapi semua berbagi penggunaan nilai-nilai yang ditetapkan. Ada hal menarik dalam penelitian ini yaitu basis data relasional kasar, model yang didasarkan pada himpunan kasar [3].

Basis data relational kasar merupakan kasus bahwa keamanan menjadi lebih dan lebih menjadi masalah dengan aplikasi basis data, terutama mengingat luasnya masalah yang terkait dengan pencurian identitas dan penipuan, pelacak riwayat kunjungan situs web, privasi dan aplikasi penambangan data, dan kebanyakan spam [4]. Dalam penelitian ini bertujuan menyelidiki area keamanan untuk basis data kasar, yang memiliki masalah keamanan yang mirip dengan basis data statistik.

Penelitian ini tidak berbicara tentang perlindungan umum data dari penggunaan yang tidak sah, tetapi dalam mengendalikan jenis data yang dapat diakses oleh pengguna yang valid. Misalnya, pengguna harus dicegah untuk menyimpulkan nilai atribut non-kunci tertentu yang terkait dengan nilai kunci.

Akan selalu ada kompromi antara manfaat berbagi informasi dan privasi, dan meskipun sering mengharapkan dengan cara memaksimalkan pembagian dan penggunaan data, hal ini tidak dapat membiarkan data yang dilindungi dikompromikan. Tahap perancangan basis data setelah pengumpulan dan analisis kebutuhan pengguna adalah perancangan basis data konseptual [5].

Dalam penelitian ini membahas masalah keamanan dalam basis data relasional kasar dan pengukuran untuk menentukan keamanan relatif dari hubungan kasar.

P-ISSN: 2548-1932
e-ISSN: 2549-7758

II. TINJAUAN PUSTAKA

2.1 Teori Himpunan

Teori himpunan adalah formalisme matematika untuk mewakili ketidakpastian [6]. Wilayah aproksimasi dalam himpunan kasar membagi beberapa alam semesta ke dalam kelas ekuivalensi. Partisi ini dapat disesuaikan untuk menambah atau mengurangi granularitasnya, untuk mengelompokkan item bersama-sama yang dianggap tidak dapat dibedakan untuk tujuan tertentu, atau untuk menyimpan domain yang dipesan ke dalam grup rentang. *U* adalah alam semesta yang tidak bisa kosong, *R* : relasi ekivalen, *A = (U,R)*, pasangan terurut, disebut ruang aproksimasi, $[x]_R$ menyatakan kelas ekivalen dari *R* yang mengandung *x*, untuk setiap elemen *x* dari *U*, himpunan elementer di *A* - kelas ekuivalensi dari *R*, himpunan terdefinisi dalam *A* - setiap penyatuan berhingga dari himpunan elementer di *A*. Setiap penyatuan berhingga dari himpunan elementer ini disebut himpunan yang dapat ditentukan. Namun, himpunan kasar $X \subseteq U$ didefinisikan dalam bentuk himpunan terdefinisi dengan menentukan daerah aproksimasi bawah ($\underline{R}X$) dan atas ($\underline{R}X$):

$$\underline{R}X = \{x \in U | [x]_R \subseteq X\} \text{ dan}$$
$$\underline{R}X = \{x \in U | [x]_R \cap X \neq \emptyset\}$$

$\underline{R}X$ adalah wilayah positif, U- $\underline{R}X$ adalah wilayah negatif, dan $\underline{R}X - \underline{R}X$ adalah batas atau wilayah garis batas dari himpunan kasar X, memungkinkan untuk perbedaan antara inklusi tertentu dan mungkin dalam himpunan kasar.

Misal: *U* = {sedang, kecil, kecil, kecil, besar, besar, besar, besar}, dan relasi ekuivalensi R didefinisikan sebagai berikut: R* = {[sedang], [kecil, kecil, kecil] , [besar, besar], [besar, besar sekali]}. Himpunan *X* = {sedang, kecil, kecil, kecil, besar, besar}, dapat didefinisikan dalam pendekatan bawah dan atas:

$\underline{\text{R}}$X = { sedang, kecil, kecil, kecil }, dan $\underline{R}X$ = { sedang, kecil, kecil, mungil, besar, besar, besar, besar }. Konsep himpunan utama yang menarik adalah penggunaan hubungan *indiscernibility* untuk domain partisi ke dalam kelas ekuivalensi dan konsep wilayah aproksimasi bawah dan atas untuk memungkinkan perbedaan antara inklusi tertentu dan mungkin, atau parsial, dalam suatu set kasar. Hubungan tidak dapat dibedakan memungkinkan pengelompokan item berdasarkan beberapa definisi kesetaraan yang berkaitan dengan domain aplikasi.

Kelas-kelas ekuivalensi yang seluruhnya termasuk dalam X termasuk ke dalam wilayah aproksimasi yang lebih rendah. Wilayah aproksimasi atas mencakup kelas ekuivalensi yang termasuk seluruhnya atau sebagian dalam X. Hasil di wilayah aproksimasi bawah adalah pasti, sesuai dengan kecocokan tepat. Wilayah batas dari aproksimasi atas berisi hasil yang mungkin, tetapi tidak pasti.

2.2 Basis Data Relasional

Model basis data relasional standar Codd [7]. Pada semua fitur penting dari teori himpunan kasar termasuk ketidakterbedaan elemen yang dilambangkan dengan kelas ekuivalensi dan daerah aproksimasi bawah dan atas untuk mendefinisikan himpunan yang tidak dapat ditentukan dalam hal ketidakterbedaan. Setiap domain atribut dipartisi oleh beberapa relasi ekuivalensi yang ditunjuk oleh basis data desainer atau pengguna. Dalam setiap domain, nilai-nilai yang dianggap tidak dapat dibedakan termasuk dalam kelas ekuivalensi. Ini kompatibel dengan model relasional tradisional karena setiap nilai milik kelasnya sendiri. Informasi ini digunakan oleh mekanisme kueri untuk mengambil informasi berdasarkan kesetaraan dengan kelas yang memiliki nilai daripada kesetaraan, menghasilkan kata-kata kueri yang kurang kritis seperti yang ditunjukkan pada [3].

Recall juga ditingkatkan dalam basis data relasional kasar karena relasi kasar menyediakan kemungkinan kecocokan dengan kueri selain kecocokan tertentu yang diperoleh dalam basis data relasional standar. Hal ini dicapai dengan menggunakan set penahanan di samping kesetaraan atribut dalam perhitungan wilayah aproksimasi bawah dan atas dari hasil kueri. Basis data relasional kasar memiliki beberapa fitur yang sama dengan basis data relasional biasa. Kedua model merepresentasikan data sebagai kumpulan relasi yang mengandung tupel. Hubungan ini adalah himpunan. Tupel dari suatu relasi adalah elemennya, dan seperti elemen himpunan pada umumnya, tidak beraturan dan tidak terduplikasi. Sebuah tuple ti mengambil bentuk (d_{i1}, d_{i2}, ..., d_{im}), di mana d_{ij} adalah nilai domain dari himpunan domain tertentu D_j. Dalam basis data relasional biasa, dij D_j. Namun, dalam basis data kasar, seperti dalam ekstensi bentuk normal non-pertama lainnya ke model relasional [8][9], d_{ij} D_j, dan

P-ISSN: 2548-1932
e-ISSN: 2549-7758

meskipun tidak mengharuskan dij menjadi lajang, dij . Misalkan $P(D_i)$ menyatakan himpunan daya (D_i).

Definisi. Relasi kasar R adalah himpunan bagian dari perkalian silang himpunan $P(D_1) \times P(D_2) \times \ldots \times P(D_m)$. Sebuah tupel kasar **t** adalah sembarang anggota dari R, yang menyiratkan bahwa ia juga merupakan anggota dari $P(D_1) \times P(D_2) \times \cdot \cdot \times P(Dm)$. Jika ti adalah beberapa tupel arbitrer, maka ti = $(d_{i1}, d_{i2}, ..., d_{im})$ di mana $d_{ij} \subseteq D_j$. Tuple dalam model ini berbeda dari basis data biasa karena komponen tupel mungkin: set nilai domain daripada nilai tunggal. Kawat gigi set dihilangkan dari singletons untuk kesederhanaan notasi. Definisi. Interpretasi = $(a_1, a_2, ..., a_m)$ dari tupel kasar $t_i = (d_{i1}, d_{i2}, ..., d_{im})$ adalah sembarang penetapan nilai sedemikian sehingga aj dij untuk semua j. Ruang interpretasi adalah hasil kali silang $D_1 \times D_2 \times \times D_m$, tetapi terbatas untuk relasi tertentu R ke himpunan tupel yang valid menurut semantik dasar R. Dalam basis data relasional biasa, karena domain nilainya atom, hanya ada satu kemungkinan interpretasi untuk setiap tupel t_i, tupel itu sendiri. Di basis data relasional kasar, ini tidak selalu terjadi ketika ada satu set nilai. Biarkan $[d_{xy}]$ menunjukkan kelas ekuivalensi yang dimiliki d_{xy}. Ketika d_{xy} adalah himpunan nilai, ekuivalensi kelas dibentuk dengan mengambil serikat kelas kesetaraan anggota himpunan; jika $[d_{xy}] = [c_1] \cup [c_2] \cup ... \cup [c_n]$

Definisi. Tupel $t_i = (d_{i1}, d_{i2}, ..., d_{im})$ dan $t_k = (d_{k1}, d_{k2}, ..., d_{km})$ redundan jika $[d_{ij}] = [d_{kj}]$ untuk semua j = 1,..., m.

Dalam basis data relasional kasar, tupel yang berlebihan dihapus dalam proses penggabungan karena duplikat tidak diperbolehkan dalam set, struktur yang menjadi dasar model relasional.

Ada dua tipe dasar dari operator relasional. Tipe pertama muncul dari fakta bahwa relasi dianggap sebagai kumpulan tupel. Oleh karena itu, operasi yang dapat diterapkan pada himpunan juga berlaku untuk relasi. Yang paling berguna untuk tujuan basis data adalah set difference, union, dan interseksi. Operator yang tidak berasal dari teori himpunan, tetapi berguna untuk pengambilan data relasional adalah pilih, proyek, dan gabung. Dalam basis data relasional kasar, relasi adalah himpunan kasar yang bertentangan dengan himpunan biasa. Oleh karena itu, operator kasar baru (—, ∩, ∪, σ, π, ⋈), sebanding dengan operator relasional standar, dikembangkan untuk relasional kasar. basis data. Properti dari operator relasional kasar dapat ditemukan di [6].

III. KEAMANAN BASIS DATA

Ada banyak keuntungan teknologi basis data seperti kemampuan untuk berbagi data dan informasi dan untuk memungkinkan akses terkontrol ke data untuk tujuan data mining. Namun, dengan kelebihan ini juga datang kekurangan. Secara khusus, ada masalah keamanan. Keamanan biasanya didefinisikan sebagai perlindungan data terhadap akses yang tidak sah [7]. Namun, juga harus melindungi data dari pengguna yang berwenang mengakses data dengan mengontrol apa yang dapat diakses dan bagaimana caranya.

Beberapa peneliti telah mempelajari masalah yang terkait dengan jenis keamanan basis data ini [10][11][12]. Keamanan dalam basis data *fuzzy* ditangani. Setiap tupel dalam basis data kasar berpotensi mewakili sejumlah besar interpretasi karena interpretasi merupakan elemen dari produk silang dari set nilai domain. Keamanan yang melekat ini terjadi karena kemampuan basis data relasional kasar untuk memungkinkan set nilai untuk atribut. Ketika data digabungkan ke dalam set ini, asosiasi spesifik nilai berdasarkan interpretasi menjadi kabur.

Oleh karena itu jika beberapa item data b ∈ D_i dilindungi, maka nilai x ∈ D_j yang berasosiasi dengan b tidak dapat ditentukan. Jadi dengan b dan x seharusnya tidak mungkin untuk menurunkan set tunggal. Salah satu bidang keamanan basis data berkaitan dengan tumpang tindih hasil kueri yang memungkinkan kesimpulan dibuat. Dimungkinkan untuk memanipulasi data dalam suatu relasi untuk menghasilkan asosiasi eksplisit untuk nilai-nilai yang dilindungi. Misalnya, jika gaji Smith akan dilindungi, pelanggaran keamanan terjadi jika tupel (…{Jusuf, David, Harun} …{65000, 85000} …) dan (.., {David, Wawan, Juan } …{54000, 75000, 85000} …) berpotongan, menghasilkan { …David …85000 …). Dalam satu kueri jenis pelanggaran keamanan ini disebabkan oleh persimpangan set yang menghasilkan satu tupel. Pada kenyataannya sulit untuk menangani masalah ini sepenuhnya karena kueri dapat dibuat pada waktu yang berbeda atau oleh pengguna yang berbeda, dan masing-masing kueri itu sendiri mungkin bukan masalah keamanan, tetapi jika dilakukan bersamaan,

P-ISSN: 2548-1932
e-ISSN: 2549-7758

mungkin melanggar beberapa privasi data. Namun, dalam basis data relasional yang kasar, persimpangan tupel dalam satu relasi tidak dapat menghasilkan pelanggaran keamanan.

Karena tupel redundan tidak diperbolehkan dalam relasi kasar, tidak mungkin ada dua tupel yang memiliki interpretasi yang sama. Dalam hal ini terbukti untuk basis data *fuzzy*. Bukti untuk basis data relasional kasar mengikuti dengan cara yang sama: TEOREMA: Perpotongan tupel dalam satu relasi kasar R tidak dapat menyebabkan pelanggaran keamanan. BUKTI: Pertimbangkan persimpangan tupel t_1, t_2, t_3, cin R atas domain D_i dan D_j . Agar pelanggaran keamanan terjadi, harus benar bahwa $|d_{1i} \cap d_{2i} \cap d_{3i} \cap \dots| = 1$ dan $|d_{1j} \cap d_{2j} \cap d_{3j} \cap \dots| = 1$.

Di sini himpunan yang dihasilkan adalah bebas, {b} dan {x}, misalnya. Ini berarti b ∈ d_{ki} dan x ∈ d_{kj} untuk semua tupel di persimpangan. Interpretasi yang menghubungkan b dan x harus merupakan interpretasi dari semua tupel yang berpotongan. Namun, hubungan kasar tidak bisa memiliki lebih dari satu tupel yang memiliki interpretasi yang sama. Oleh karena itu, perpotongan tupel dalam relasi kasar tunggal tidak dapat menghasilkan pelanggaran keamanan. Pelanggaran keamanan dalam basis data relasional kasar dalam hal akses item data yang dilindungi berhubungan langsung dengan ketidakpastian tentang asosiasi item data tertentu. Langkah-langkah informasi-teoretis sering digunakan untuk mengukur ketidakpastian, dan mereka telah digunakan dalam basis data statistik, dan untuk basis data fuzzy [13]. Dalam langkah-langkah informasi-teoritis basis data relasional kasar untuk ketidakpastian adalah didefinisikan untuk skema kasar dan hubungan kasar:

Definisi. Entropi skema kasar untuk skema relasi kasar S adalah

$$E_s(S) = -\sum_j [\sum Q_i \; log(P_i)]$$

untuk i = 1,... n; j = 1,..., m dimana terdapat n kelas ekivalen dari domain j, dan m atribut dalam skema R(A_1, A_2, ...,A_m). Definisi. Entropi relasi kasar dari perpanjangan skema tertentu adalah

$$E_R(R) = -\sum_j Dp_j(R)[\sum Q_i \; log(DP_i)]$$

untuk i = 1,... n; j = 1,..., m di mana $D_{\rho j}(R)$ mewakili jenis kekasaran basis data untuk himpunan kasar nilai domain untuk atribut j dari relasi, m adalah jumlah atribut dalam relasi basis data, dan n adalah jumlah kelas kesetaraan untuk domain yang diberikan untuk basis data.

Entropi skema memberikan ukuran ketidakpastian yang melekat dalam definisi skema hubungan kasar dengan mempertimbangkan partisi domain di mana skema atribut didefinisikan. Entropi dari contoh relasi kasar aktual $E_R(R)$ dari beberapa basis data D adalah penambahan dari skema entropi yang diperoleh dengan mengalikan setiap suku dalam produk dengan kekasaran himpunan nilai kasar untuk domain dari atribut yang diberikan. Diperoleh nilai $D_{\rho j}(R)$ dengan membiarkan nilai domain non singleton mewakili elemen wilayah batas, menghitung kasar asli. akurasi dan menguranginya dari satu untuk mendapatkan kekasaran. DQ_i adalah probabilitas sebuah tuple dalam relasi basis data yang memiliki nilai dari kelas i, dan DP_i adalah probabilitas sebuah nilai untuk kelas i muncul dalam relasi basis data dari semua nilai yang diberikan. Pertimbangkan basis data sampel di bawah ini di mana domain untuk warna dan ukuran tanah telah didefinisikan sebagai:

WARNA = {[hitam, hitam gelap], [coklat, coklat kemerahan, coklat kekuningan],[putih], [abu-abu], [oranye]}, dan UKURAN-PARTIKEL = {[besar, besar lebar], [besar-besar, besar sekali], [sedang], [kecil, kecil sedang, kecil kurus]}.

Tabel 1. Sampel 1

BIN	Warna	Ukuran Partikel
B11	coklat	Sedang
B12	{hitam,coklat kemerahan}	besar
		{ sedang, kecil }
B13	abu-abu	
K01	hitam	
K04	{abu-abu, coklat}	

Tabel 2. Sample 2

BIN	Warna	Ukuran Partikel
L42	{ hitam, coklat kemerahan, putih}	{ besar , besar-besar, sedang }
		{ sedang, kecil }
L45	{coklat, oranye, putih, abu-abu}	

DOI: 10.32897/infotronik.2021.6.2.695

P-ISSN: 2548-1932
e-ISSN: 2549-7758

Entropi relasi kasar dari relasi Sampel-1 dan Sampel-2 yang ditunjukkan pada tabel dihitung sebagai berikut:

E_R(Sample-1) = -(4/7)[(2/5)log(2/7) + (3/5)log(3/7) + 0 + (2/5)log(2/7) + 0] - (2/6)[(2/5log(2/6) + 0 + (2/5))log(2/6) + (2/5)log(2/6)] = 0,56

E_R(Sampel-2) = -(7/7)[(1/2)log(1/7) + (2/2)log(2/7) + (2/2)log(2/7) + (1/2)log(1/7) + (1/2)log(1/7)] – (5/5) [(1/2)log(1/5) + (1/2)log(1/5) + (2/2)log(2/5) + (1/2)log(1,5) = 3,78

Dari contoh ini jelas bahwa konsep keamanan dalam basis data relasional kasar sesuai dengan ketidakpastian dalam pengertian ini, sehingga dapat menggunakan ukuran entropi ini sebagai ukuran kuantitatif untuk keamanan dalam basis data relasional kasar.

IV. KESIMPULAN

Keamanan merupakan masalah penting dalam basis data, dan aplikasi yang melibatkan basis data statistik dan penambangan data sangat penting dalam menjaga keamanan data yang dilindungi. Aspek yang terkait dengan jenis keamanan basis data ini juga berlaku untuk basis data relasional yang kasar. Hasil penelitian telah menunjukkan bagaimana sifat basis data relasional kasar menyediakan beberapa keamanan yang melekat melalui penggunaan struktur non-first bentuk normal. Selain itu, menyediakan langkah-langkah keamanan basis data berdasarkan langkah-langkah teoretis informasi yang memungkinkan evaluasi langkah-langkah numerik untuk entropi. Diperlukan Penelitian selanjutnya untuk menyelidiki perluasan topik ini untuk basis data relasional *fuzzy* kasar dan intuitif.

DAFTAR PUSTAKA

[1] N. I. Munawar, Zen and Putri, “Keamanan Jaringan Komputer Pada Era Big Data,” *J-SIKA| J. Sist. Inf. Karya Anak Bangsa*, vol. 02, no. 01, pp. 14–20, 2020.

[2] A. Makinouchi, “A Consideration on Normal Form of Not-Necessarily-Normalized Relation in the Relational Data Model,” in *VLDB*, 1977, pp. 447–453.

[3] F. E. Petry and B. P. Buckles, “Extension of the Relational Database and its Algebra with Rough Set Techniques,” in *Computational Intelligence*, 1995, vol. 11, no. 2, pp. 233–245.

[4] W. Stalling and L. Brown, *Computer Security: Principles and Practice*. Prentice Hall, 2007.

[5] Z. Munawar, M. I. Fudsyi, and D. Z. Musadad, “Perancangan Basis Data untuk Sistem Informasi Persediaan ATK pada PT. SPP,” *Temat. - J. Teknol. Inf. dan Komun.*, vol. 6, no. 1, pp. 75–94, Jun. 2019.

[6] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Norwell: Kluwer Academic Publishers, MA, 1991.

[7] R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, Seventh. Boston: Pearson, 2016.

[8] M. A. Roth, H. F. Korth, and D. S. Batory, “SQL/NF: A query language for ¬1NF relational databases,” *Inf. Syst.*, vol. 12, no. 1, pp. 99–114, 1987.

[9] S. J. Thomas and P. Fischer, “Nested Relational Structures,” *Adv. Comput. Res.*, vol. 3, pp. 269–307, 1986.

[10] D. E. Denning, “Secure statistical databases with random sample queries,” *ACM Trans. Database Syst.*, vol. 5, no. 3, pp. 291–315, 1980.

[11] F. Y. Chin and G. Ozsoyoglu, “Statistical Database Design,” *ACM Trans. Database Syst.*, vol. 6, no. 1, pp. 113–139, Mar. 1981.

[12] M. McLeish, “Further Results on the Security of Partitioned Dynamic Statistical Databases,” *ACM Trans. Database Syst.*, vol. 14, no. 1, pp. 98–113, 1989.

[13] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell Syst. Tech. J.*, vol. 27, no. 4, pp. 623–656, 1948