

暗号化データベースモデルにおける問合せの関連情報を秘匿する範囲検索

川本, 淳平

櫻井, 幸一

<https://hdl.handle.net/2324/1398220>

出版情報 : 暗号と情報セキュリティシンポジウム. 2014, 2014-01

バージョン :

権利関係 : (C) The Institute of Electronics, Information and Communication Engineers

暗号化データベースモデルにおける問合せの関連情報を秘匿する範囲検索

Private Range Query on Encrypted Database Model

川本 淳平 *
Junpei Kawamoto

櫻井 幸一 *
Kouichi Sakurai

あらまし 本稿では、暗号化データベースモデルにおいて問合せの関連情報を秘匿する範囲検索を手法を提案する。暗号化データベースは、完全には信頼できないクラウドデータベースを安全に利用するために、サーバ上には予めクライアントによって暗号化されたデータのみを保存する枠組みである。従来、暗号化データベースに対する問合せを秘匿する方法は提案されてきたが、問合せ間の関連情報を用いた攻撃には必ずしも対応しているわけではない。本稿では、関連情報を用いた攻撃に耐性を持ち範囲検索を実現する手法を提案する。

キーワード クラウドデータベース, セキュリティ, 問合せ保護

1 はじめに

データベース型サービス (DaaS; Database as a Service) における, サーバに預けるデータの安全性や利用者のプライバシーの保護は常に議論されている問題の一つである [2, 3]. その中でも, データベースへの問合せ内容をサーバを含む攻撃者に公開することなく目的のデータを取得したいという要求は古くからある. この要求に関する研究の一つとしては PIR (Private Information Retrieval) [1] が良く知られている. PIR における問題設定では, サーバは 0 から $n-1$ までの数値が割り当てられた計 n 個の 1 ビットデータ $\{x_0, x_1, \dots, x_{n-1}\}$ を保持している. 利用者は x_i の値を問い合わせるのだが, この時問い合わせた番号 i に関する情報が攻撃者に知られないように問合せ処理を実現することが目的である. しかし, 実用的なデータベースを考えた場合, 求むべきデータの番号 i が予め分かっていることは少なく, 何らかの条件を元に求むべきデータを検索することが多い. 我々は, 検索の基本操作の一つである範囲検索を安全に実行する方法について取り組み IPP 法を提案してきた [4].

一方, 長沼らは問合せからの情報漏洩を防ぐ検索可能暗号のための安全性定義である IND-CKQA を提案している [5]. 我々は, この IND-CKQA を満たす範囲問合せを実現するための暗号化範囲問合せフレームワークを定義し, IPP 法の適用を試みる. 特に, IND-CKQA を

満たすためには, サーバ用の鍵を導入し問合せとデータを不正入手した攻撃者が問合せ処理を実行できないようにする仕組みが必要である. 本論文では巡回行列を用いてサーバ, クライアントそれぞれが別々の鍵を持つように IPP 法を拡張する.

以降の節は, 次のような構成になっている. 第 2 節では暗号化範囲問合せのフレームワークを導入する. そして, 第 3 節では IPP 法の基本アイデアについて説明した後, 第 4 節では, 巡回行列を用いた新しい暗号化について説明する. 第 5 節は, まとめと今後の課題である.

2 暗号化範囲問合せ

本節では, 本研究で対象とする暗号化範囲問合せのフレームワークについて述べる. Alice は自分の管理する大量のデータを Bob が管理するサーバに預け信頼できる Carol と共有したいとする. また, Carol も自分の管理するデータを Bob が管理するサーバに預け Alice と共有したいとする. しかし, Alice と Carol は Bob のことを完全には信用しておらず, 二人は Bob に預けるデータを予め暗号化することを考える. 一方で, Bob 預けるデータが大量にあるため, 二人は Bob からデータを取得するために検索を必要とする. 暗号化範囲問合せはこの要求に答えるものである.

Alice と Carol が預ける個々のデータは, 検索用のキー属性値 k とキー属性値に紐付けられたデータ v の組 (k, v) という形をしているとする. また, キー属性値の定義域 \mathbb{D}_k は l bit 自然数とする. データに対する検索は, キー属性の値 k が $[\alpha, \beta] (\alpha \leq \beta \in \mathbb{D}_k)$ に含

* 九州大学大学院システム情報科学研究院, 福岡県福岡市西区元岡 744, Kyushu University, 744 Motoooka, Nishiku, Fukuoka. {kawamoto,sakurai}@inf.kyushu-u.ac.jp

まれるデータを検索する範囲問合せのみを考える．ただし， $\alpha = \beta$ とすれば一致検索も利用できる．本論文では，データに対する操作は考えないため，どのような暗号方式を用いて暗号化してもよい． v は Alice と Carol によって合意された暗号方式を用いて既に暗号化されているものとする．

この暗号化範囲問合せは次の四つのアルゴリズムからなる．

鍵生成

Alice は，Carol と共有する利用者用秘密鍵 sk_c と Bob に渡すサーバ用秘密鍵 sk_s を生成する．鍵生成アルゴリズム KeyGen は，セキュリティパラメータ n を引数とし，秘密鍵のペアを出力する．

$$(sk_c, sk_s) \leftarrow \text{KeyGen}(n)$$

なお，Bob は受け取った秘密鍵を安全に管理するものとする．

属性値暗号化

Alice または Carol は，Bob に預けるデータ (k, v) のキー属性値 $k \in \mathbb{D}_k$ を暗号化キーベクトル \vec{ek} に変換する．キー属性値暗号化アルゴリズム SERQ は，平文キー属性値 k と利用者用の秘密鍵 sk_c を入力とし，暗号化キーベクトル \vec{ek} を出力する．

$$\vec{ek} \leftarrow \text{SERQ}(k, sk_c)$$

Alice または Carol は， (\vec{ek}, v) を Bob へ送信する．

問合せ生成

Alice または Carol は，検索範囲 $[\alpha, \beta] (\alpha \leq \beta \in \mathbb{D}_k)$ を変換した暗号化問合せベクトル \vec{eq} と付加情報 x を生成する．問合せ変換アルゴリズム Query は，問合せ範囲 $[\alpha, \beta]$ と利用者用秘密鍵 sk_c を入力とし，暗号化問合せベクトル \vec{eq} と x の組を出力する．

$$(\vec{eq}, x) \leftarrow \text{Query}([\alpha, \beta], sk_c)$$

Alice または Carol は，得られた (\vec{eq}, x) を Bob へ送信する．

範囲検索

Bob は，Alice または Carol から問合せを受け取ると，保管しているデータ $(\vec{ek}_i, v_i) (i = 1, 2, \dots)$ それぞれが問合せに該当するか否かをアルゴリズム Test を用いて調べる．判定アルゴリズム Test は，暗号化キーベクトル \vec{ek}_i と問合せ (\vec{eq}, x) ，サーバ用の秘密鍵 sk_s を入力とし， $res_i \in \{0, 1\}$ を出力する．

$$res_i \leftarrow \text{Test}(\vec{ek}_i, \vec{eq}, x, sk_s)$$

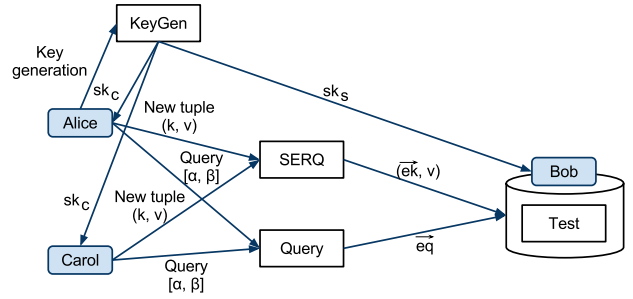


図 1: 暗号化範囲検索のフレームワーク．

判定結果 res が 1 であった i に対応する v_i の集合を作成し，問合せ元である Alice または Carol に送信する．

図 1 は，以上のフレームワークを図示したものである．

3 内積述語を用いた問合せ

我々は，暗号化データベース上で範囲検索を実現する手法として IPP 法を提案している [4]．IPP 法では，キー属性値と問合せの両方にランダムな摂動を加え暗号化することで，同じキー属性値・問合せが複数生成，問合せられた場合で合ってもその頻度分布を秘匿することが出来る．本節では，IPP 法の基本アイデアを導入する．

以降では，Bob が管理するデータベース D は， N 個のタプル $t = (k, v)$ の集合 $D = \{t_1, t_2, \dots, t_N\}$ であるとする．また，関数 key, value を用いて，タプルのキー属性値 k ，データ v を表すことにする．すなわち， $\text{key}(t) = k$ ， $\text{value}(t) = v$ である．これらを用いると，区間 $[\alpha, \beta]$ を問い合わせるデータベース D 上の範囲問合せ $q_D(\alpha, \beta)$ は，

$$q_D(\alpha, \beta) = \{t \in D \mid \alpha \leq \text{key}(t) \leq \beta\}$$

と表すことができる．

IPP 法では，キー属性値の漏えいを防ぐためにキー属性値に摂動を加えている．摂動には $0 < r_k < 1/2$ を満足する乱数 r_k を用い，キー属性値 k に摂動を加えた値を $k + r_k$ とする．キー属性値に摂動を加える操作を

$$\text{per} : \mathbb{D}_k \times R \rightarrow R, \text{per}(x; r_k) = x + r_k$$

と表す．なお，摂動 r_k は毎回異なる値をランダムに選ばれる．つまり同じキー属性値であっても別の値が使用される．

キー属性値に摂動を加えたタプルからなるデータベースを摂動を加えたデータベース PD と呼ぶことにする．この時，元々のデータベース D と摂動を加えたデータベース PD の関係は，

$$PD = \{(\text{per}(\text{key}(t); r_k \xleftarrow{r} [0, 1/2]), \text{value}(t)) \mid \forall t \in D\}$$

である．ただし， $r_k \leftarrow [0, 1/2]$ は，タプル毎に異なるランダムな値を $[0, 1/2]$ から選ぶことを意味する．摂動を加えたデータベースに対する問合せ q_{PD} は，キー属性の定義域が自然数かつ $0 < r_k < 1/2$ という条件から次のように変換できる．すなわち，区間 $[\alpha, \beta]$ を問い合わせる問合せは

$$q_{PD}(\alpha, \beta) = \{t \in PD \mid \alpha - 1/2 \leq \text{key}(t) \leq \beta + 1/2\}$$

となる．

3.1 多項式述語

問合せに対して摂動を加えるために，我々は問合せを内積述語という形で表現する．ここでは，先ず内積述語の元となる多項式述語を導入する．

多項式述語では，範囲問合せを多項式 $p: \mathbb{D}_k \rightarrow \mathbf{R}$ を用いて表現する．述語の値域は実数全体であるが，次の様に定める sign 関数

$$\text{sign}(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (\text{otherwise}) \end{cases}$$

を用い， $\text{sign} \circ p$ とすることで 2 節で定義した Test 関数を実現する．

区間 $[\alpha, \beta]$ の範囲問合せを表す多項式述語 $p_{[\alpha, \beta]}$ は複数存在するが，我々は，最も単純な多項式の一つである $p_{[\alpha, \beta]}(k) = -(k - \alpha)(k - \beta)$ を用いる．この多項式に対して， $\text{sign} \circ p_{[\alpha, \beta]}(k)$ は $k \in [\alpha, \beta]$ の時に限り 1 となることは自明である．

我々が対象としているデータベース PD では，キー属性値には摂動 r_k が加えられている．この摂動付きキー属性値に対し，偽陰性を含まず問合せ処理を行うために多項式述語のパラメータに摂動による増加分を加える必要がある．具体的には，区間 $[\alpha, \beta]$ の範囲問合せを表す先ほどの多項式述語は， $p'_{[\alpha, \beta]}(k) = -(k - \alpha + 1/2)(k - \beta - 1/2)$ となる．

次に，この多項式述語に摂動を加える．我々は，摂動に正の乱数 r_q を用い，多項式述語に $(k + r_q)$ を乗じることで摂動を加える．先ほどの多項式述語 $p'_{[\alpha, \beta]}$ の場合，摂動を加えた述語 $\tilde{p}'_{[\alpha, \beta]; r_q}(k)$ は

$$\begin{aligned} \tilde{p}'_{[\alpha, \beta]; r_q}(k) &= \text{per}(p'_{[\alpha, \beta]}(k); r_q) \\ &= -(k - \alpha + 1/2)(k - \beta - 1/2)(k + r_q) \end{aligned} \quad (1)$$

となる．キー属性値の定義域 \mathbb{D}_k は自然数と仮定しているので， $k \in [\alpha, \beta]$ であれば (1) はゼロ以上となる．すなわち， $\text{sign} \circ \tilde{p}'_{[\alpha, \beta]; r_q}(k)$ は $k \in [\alpha, \beta]$ の時に限り 1 となる．

3.2 内積述語

先ほど定義した多項式述語をベクトルの内積を用いた述語である内積述語として表現する．一般に，任意の一変数多項式は定数ベクトルと変数ベクトルの内積として表すことができる．この性質により，(1) の摂動を加えた多項式述語は，定数ベクトル

$$\vec{q}_{[\alpha, \beta]; r_q} = \begin{pmatrix} -1 \\ \alpha + \beta - r_q \\ -(\alpha - 1/2)(\beta + 1/2) + (\alpha + \beta)r_q \\ -(\alpha - 1/2)(\beta + 1/2)r_q \end{pmatrix} \quad (2)$$

と変数ベクトル $\vec{k} = (k^3, k^2, k, 1)^t$ の内積 $\vec{q}_{[\alpha, \beta]; r_q} \cdot \vec{k}$ と表すことができる¹．

このことを用いて，キー属性値が k であるタプルには， k の代わりにベクトル化した $(k^3, k^2, k, 1)^t$ をキー属性値として与えて置けば，範囲問合せ $[\alpha, \beta]$ に対応する内積述語の一つは

$$\text{IPP}_{[\alpha, \beta]; r_q}(\vec{k}) = \vec{q}_{[\alpha, \beta]; r_q} \cdot \vec{k}$$

となる．実際には，キー属性値 k には摂動が追加され， $\text{per}(k, r_k) = k + r_k$ となっている．したがって，キー属性値が k であるタプルには k の代わりに摂動付きキーベクトル

$$\vec{k}^t = (\text{per}(k; r_k)^3, \text{per}(k; r_k)^2, \text{per}(k; r_k), 1)^t \quad (3)$$

が与えられることになる．前節で議論したように，摂動付きキー属性値に対しても，キー属性値 k が $k \in [\alpha, \beta]$ であれば，多項式述語 $\tilde{p}'_{[\alpha, \beta]; r_q}$ がゼロ以上の値となる．したがって，多項式述語の等価な変形である内積述語 $\text{IPP}_{[\alpha, \beta]; r_q}$ もゼロ以上の値となり， $\text{sign} \circ \text{IPP}_{[\alpha, \beta]; r_q}(\vec{k})$ は， $k \in [\alpha, \beta]$ の時に限り 1 となる．

4 巡回行列を用いた暗号化範囲検索

IPP 法 [4] では，キーベクトル \vec{k}^t と問合せに用いるベクトル $\vec{q}_{[\alpha, \beta]; r_q}$ の値を秘匿するために，正則行列 M を用いて， $M^1 \vec{k}^t$ と $M^t \vec{q}_{[\alpha, \beta]; r_q}$ をそれぞれ暗号化されたキーベクトルと問合せに用いるベクトルとして用いていた．このとき，二つのベクトルの内積は，

$$M^t \vec{q}_{[\alpha, \beta]; r_q} \cdot M^1 \vec{k}^t = \vec{q}_{[\alpha, \beta]; r_q}^t M M^1 \vec{k}^t = \vec{q}_{[\alpha, \beta]; r_q} \cdot \vec{k}^t$$

となり，内積結果は暗号化の前後で変化しない．

一方，この方式では，キーベクトルと問合せを盗み見られる攻撃者はそのキーベクトルが問合せに合致するのかが確認することが出来る．本論文では，Bob にもサーバ用の鍵を渡すことで，サーバになりすます第三者からの攻撃に耐性を持つ方法へ拡張する．

¹ \vec{x}^t でベクトル \vec{x} の転値ベクトルを， $\vec{x} \cdot \vec{y}$ でベクトル \vec{x} と \vec{y} の内積を表す．

4.1 行列を用いた暗号化

本提案手法では、暗号鍵として n 巡回行列を利用する。 n 巡回行列 A とは $A^n = E$ かつ $A^i \neq E$ ($i < n$) を満たす行列 A のことを言う。なお、 E は単位行列である。この巡回行列 A を用いて、Alice や Carol といった利用者のための鍵を (A, n, c, s) とする。ここで、 c と s は n 未満の乱数である。そして、Bob 用の即ちサーバ用の鍵として A^s を用いる。

これらの鍵を用いた場合のキーベクトルの暗号化方法は次の通りである。暗号化前のキーベクトルが \vec{k}^j であるとすると、乱数 r_e を用いて、 $r_e A^c \vec{k}^j$ を暗号化したキーベクトル \vec{ek} とする。また、問合せ用のベクトルについては、乱数 r を用いて $(A^r)^t \vec{q}$ を暗号化した問合せベクトル \vec{eq} とする。さらに、問合せ時には、 x についての方程式

$$sx + c + r = 0 \pmod{n} \quad (4)$$

を解き、得られた x を暗号化した問合せベクトルと併せてサーバに送る。

サーバは自信が持つサーバ用の鍵 A^s を用いて、 $\vec{eq} \cdot (A^s)^x \vec{ek}$ を計算することで暗号化前のベクトルに対する内積を計算できる。すなわち、

$$\begin{aligned} \vec{eq} \cdot (A^s)^x \vec{ek} &= (A^r)^t \vec{q} \cdot (A^s)^x r_e A^c \vec{k}^j \\ &= r_e \vec{q}^t A^r A^{sx} A^c \vec{k}^j \\ &= r_e \vec{q}^t A^{sx+c+r} \vec{k}^j \\ &= r_e \vec{q}^t \vec{k}^j = \vec{q} \cdot \vec{k}^j \end{aligned}$$

である。従って、

$$\text{sign}(\vec{eq} \cdot (A^s)^x \vec{ek}) \quad (5)$$

を判定すればその暗号化キーベクトルを持つタプルが問合せ内容に合致しているか否かを判定できる。また、サーバである Bob 用の暗号鍵 A^s を知らない攻撃者は問合せと暗号化タプルを入手したとしても問合せに合致するタプルか否かの判定はできない。

4.2 暗号化範囲検索のアルゴリズム

Alice は始めに、アルゴリズム KeyGen を用いて利用者用の秘密鍵 sk_c とサーバ用の秘密鍵 sk_s を生成する。この手順をアルゴリズム 1 に示す。KeyGen は、与えられたセキュリティパラメータ n を元に 4 次正方 n 巡回行列と互いに素な整数 c, s を生成する。それらを用いて二つの鍵を計算し出力する。なお、3 次正方 n 巡回行列を生成するひとつの方法を付録に記している。Alice は、作成した利用者用の秘密鍵 sk_c を Carol に、サーバ用の秘密鍵 sk_s を Bob に渡す。

Alice または Carol が Bob に新しいデータ (k, v) を預けたいとする。この時、Alice または Carol は、アル

Algorithm 1 KeyGen

Require: Big prime parameter n

- 1: Generate n -order cyclic matrix A
 - 2: Choose $c, s < n$, where c and s are relatively prime
 - 3: $sk_c \leftarrow (A, n, c, s)$, $sk_s \leftarrow A^s$
 - 4: **return** (sk_c, sk_s)
-

Algorithm 2 SERQ

Require: Attribute value $k \in \mathbb{D}_K$, Encryption key of clients $sk_c = (A, n, c, s)$

- 1: $r_k \leftarrow_r [0, 1/2]$
 - 2: $\vec{k}^j = (\text{per}(k; r_k)^3, \text{per}(k; r_k)^2, \text{per}(k; r_k), 1)^t$
 - 3: $r_e \leftarrow_r R^+$
 - 4: $\vec{ek} \leftarrow r_e A^c \vec{k}^j$
 - 5: **return** \vec{ek}
-

Algorithm 3 Query

Require: Querying range $[\alpha, \beta] (\alpha \leq \beta \in D_K)$, encryption key of clients $sk_c = (A, n, c, s)$

- 1: Compute \vec{q} by (2)
 - 2: $r \leftarrow Z$
 - 3: $\vec{eq} \leftarrow_r (A^r)^t \vec{q}$
 - 4: Solve for x : $sx + c + r = 0 \pmod{n}$
 - 5: **return** (\vec{eq}, x)
-

ゴリズム SERQ を用いてキー属性値 k の代わりに用いる暗号化キーベクトル \vec{ek} を求める。この手順をアルゴリズム 2 に示す。SERQ は、先ず、 $[0, 1/2]$ 内の乱数 r_k を生成し 3 で導入したキーベクトル \vec{k}^j を計算する。その後、新たに生成した正の乱数 r_e と利用者用の秘密鍵 sk_c の中から A と c を用いて暗号化キーベクトルを計算する。Alice または Carol は、SERQ によって得られた暗号化キーベクトルを用いたデータ (\vec{ek}, v) を Bob へ送信する。なお、タプル中の v は一般的な暗号化手法を用いて暗号化されているものとする。

Alice または Carol が Bob からキー属性値 k が区間 $[\alpha, \beta]$ に含まれるデータを取得したいとする。この時、Alice または Carol はアルゴリズム Query を用いて問合せを作成する。この手順をアルゴリズム 3 に示す。Query は、(2) で導入した問合せベクトルを計算する。そして、整数乱数 r を生成し利用者用秘密鍵 sk_c を用いて暗号化問合せベクトル \vec{eq} を計算する。次に、 x についての方程式 (4) を解き x を得る。最後に、それらの組 (\vec{eq}, x) を出力する。Alice または Carol は Query によって得られた問合せを Bob へ送る。

Bob は、Alice または Carol から問合せ (\vec{eq}, x) を受け取ると、保持しているデータすべてに対しアルゴリズム Test を適用し問合せに合致するタプルか否かを判定

Algorithm 4 Test

Require: Encrypted key vector \vec{ek}_i , query (\vec{eq}, x) , and secret key of the server $sk_s = A^s$
1: **return** $\text{sign}(\vec{eq} \cdot (A^s)^x \vec{ek}_i)$

Algorithm 5 Encrypted Range Search with Query

Require: Query (\vec{eq}, x) , database PD , and encryption key of the server $sk_s = A^s$
1: $Res \leftarrow \emptyset$
2: **for** each tuple t in PD **do**
3: $\vec{ek} \leftarrow \text{key}(t)$
4: **if** $\text{Test}(\vec{ek}, \vec{eq}, x, sk_s) = 1$ **then**
5: $Res \leftarrow Res \cup \{\text{value}(t)\}$
6: **end if**
7: **end for**
8: **return** Res

する。この Test は、アルゴリズム 4 に示すとおりで、式 (5) を判定する。Bob は、Test が 1 を返すタプルを集め、Alice または Carol へ送信する。この手順をアルゴリズム 5 に示す。

5 おわりに

本論文では、IND-CKQA を満たす範囲問合せを実現するための暗号化範囲問合せフレームワークを定義し、我々が今まで提案してきた IPP 法を適用した。また、サーバ用鍵を導入することで、暗号化されたデータや問合せを不正入手した攻撃者が問合せ判定を実行することを防ぐ拡張について提案した。今後の課題は、本提案手法が IND-CKQA を満たすことを証明することである。

謝辞

本研究は JSPS 科研費 23300027 並びに中島記念国際交流財団による補助のもとで行われた。ここに記して謝意を表します。

参考文献

- [1] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private Information Retrieval. *Journal of the ACM*, Vol. 45, No. 6, pp. 965–982, 1998.
- [2] George I. Davida, David L. Wells, and John B. Kam. A database encryption system with subkeys. *ACM Transactions on Database Systems*, Vol. 6, No. 2, pp. 312–328, June 1981.
- [3] Hakan Hacigumus, Bala Iyer, Chen Li, and Sharad Mehrotra. Executing SQL over Encrypted Data in

the Database-Service-Provider Model. In *Proc. of the 21st ACM SIGMOD International Conference on Management of Data*, pp. 216–227, Madison, WI, USA, 2002. ACM Press.

- [4] Junpei Kawamoto and Masatoshi Yoshikawa. Private Range Query by Perturbation and Matrix Based Encryption. In *Proc. of the Sixth IEEE International Conference on Digital Information Management*, pp. 211–216, Melbourne, Australia, 2011. IEEE Computer Society.
- [5] Ken Naganuma, Masayuki Yoshino, and Hisayoshi Satoh. Research of Database Security on Keyword-Searchable Encryption (1). In *Proc. of the 2011 Symposium on Cryptography and Information Security*, pp. 1–5, Kokura, Japan, 2011. The Institute of Electronics, Information and Communication Engineers.