หัวข้อวิทยานิพนธ์	การพัฒนาขั้นตอนวิธีสำหรับการรักษาความเป็นส่วนตัวที่มีประสิทธิภาพ	
ผู้เขียน	นาย บวรศักดิ์ ศรีสังสิทธิสันติ	
ปริญญา	ปรัชญาดุษฎีบัณฑิต (วิศวกรรมคอมพิวเตอร์)	
คณะกรรมการที่ปรึกษา	รศ.ดร.จักรพงศ์ นาทวิชัย	อาจารย์ที่ปรึกษาหลัก
	อ.ดร.พฤษภ์ บุญมา	อาจารย์ที่ปรึกษาร่วม
	รศ.ดร.ตรัสพงศ์ ไทยอุปถัมภ์	อาจารย์ที่ปรึกษาร่วม

บทคัดย่อ

การรักษาความเป็นส่วนตัวของข้อมูลเป็นประเด็นที่สังคมกำลังให้ความสำคัญ เนื่องจากการแลกเปลี่ยนข้อมูลระหว่างองค์กรที่เป็นพันธมิตรมีความสำคัญและเป็นประโยชน์ในการหาองค์ความรู้ใหม่เพื่อใช้ในการปรับตัวทางกลยุทธ์ด้านต่างๆขององค์กร กระบวนการรักษาความเป็นส่วนตัวของข้อมูลจึงจำเป็นเพื่อให้การแลกเปลี่ยนข้อมูลนั้น ไม่เกิดการละเมิดความเป็นส่วนตัวของบุคคลที่เกี่ยวข้องกับข้อมูลที่ถูกแลกเปลี่ยน แบบจำลองที่ใช้รักษาความเป็นส่วนตัวของข้อมูลมีหลายแบบจำลอง ในงานวิจัยนี้ได้เลือกแบบจำลอง (k, e)-Anonymous มาพัฒนาต่อยอด เนื่องจากเป็นแบบจำลองที่มีประสิทธิภาพและประสิทธิผลสูง โดยแบบจำลอง (k, e)-Anonymous ถูกออกแบบมาเพื่อให้ข้อมูลที่ถูกแปลงตามกระบวนการของแบบจำลองแล้ว สามารถนำไปใช้ในการสืบค้นข้อมูลแบบรวมกลุ่ม (Aggregation Query Processing) ได้อย่างมีประสิทธิผล

อย่างไรก็ตามข้อมูลที่ถูกแลกเปลี่ยนนั้นเมื่อผ่านระยะเวลาไปช่วงหนึ่ง อาจมีการเพิ่มเข้ามาของข้อมูลใหม่ ทำให้ข้อมูลที่ถูกแลกเปลี่ยนมีหลายฉบับ โดยแต่ละฉบับอ้างอิงตามเวลาที่ถูกแลกเปลี่ยน หากข้อมูลแต่ละฉบับซึ่งถูกแปลงตามแบบจำลอง (k, e)-Anonymous แล้ว ถูกนำมาเปรียบเทียบกัน จากการศึกษาพบว่าอาจทำให้มีข้อมูลบางส่วนผิดเงื่อนไขการรักษาความเป็นส่วนตัวของข้อมูลตามแบบจำลอง (k, e)-Anonymous ได้ โดยจะอ้างถึงปัญหานี้ว่า “การละเมิดความเป็นส่วนตัวแบบเพิ่มขึ้น”

ในการจัดการกับประเด็นปัญหานี้ ผู้เขียนได้ทำการศึกษาลักษณะของการคงอยู่ของข้อมูลหลายฉบับในแง่ทฤษฎี จากการศึกษา พบว่าการละเมิดความเป็นส่วนตัวแบบเพิ่มขึ้นนั้น เกิดขึ้นเมื่อมีการทับซ้อนกันระหว่างส่วนประกอบในข้อมูลใหม่ และข้อมูลเดิม ในงานวิจัยนี้จึงได้พัฒนาขั้นตอนวิธีที่มีความซับซ้อนทางการคำนวณอยู่ในเวลาของโพลิโนเมียลฟังก์ชัน และสามารถยกเว้นการทำงาน

บางส่วนที่ทับซ้อนกัน และยังตรวจสอบเพียงสำเนาปัจจุบันกับสำเนาข้อมูลก่อนที่ถูกเผยแพร่เพียงสำเนาเดียว ไม่ต้องตรวจสอบกับสำเนาข้อมูลก่อนหน้าทั้งหมด ซึ่งขั้นตอนวิธีมีความซับซ้อนทางการคำนวณลดลงจาก $O(n^m)$ เป็น $O(pn^3)$ โดย n เป็นจำนวนข้อมูลของสำเนาปัจจุบัน m คือจำนวนสำเนาที่เผยแพร่ไปก่อนหน้า และ p คือจำนวนพาร์ทิชันของข้อมูลภายในข้อมูลอินพุต ในขณะเดียวกันสำเนาข้อมูลที่ถูกเผยแพร่นั้นยังสามารถการันตีได้ว่าเป็นสำเนาข้อมูลที่มีค่าผลรวมความผิดพลาดที่น้อยที่สุด นอกจากนี้จากผลการทดลองสามารถยืนยันว่าขั้นตอนวิธีที่นำเสนอนั้นสามารถนำมาใช้กับข้อมูลจริงได้อย่างมีประสิทธิภาพ

Thesis Title	Development of Efficient Privacy-Preservation Algorithms	
Author	Mr. Bowosak Srisungsittisunti	
Degree	Doctor of Philosophy (Computer Engineering)	
Advisory Committee	Assoc. Prof. Dr. Juggapong Natwichai	Advisor
	Dr. Pruet Boonma	Co-advisor
	Assoc. Prof. Dr. Trasapong Thaiupathump	Co-advisor

ABSTRACT

Privacy preservation is one of the important issues that obtain a lot of attention in society. When the collaboration is to be taking place among partners for obtaining the useful knowledge to achieve a good strategic move, the privacy preservation is a necessity for prevent the privacy breach at all cost. Though, there exist several privacy preservation models currently. In this research, the problem of data privacy preservation based on a prominent privacy model, (*k*, *e*)-Anonymous, is addressed. The target data processing which can be applied to the data from the model is aggregated data querying, which is a fundamental data processing of many data analysis and data mining algorithms.

However, when a new dataset is to be released, there may be, at the same time, datasets that were released elsewhere, a problem arises because some attackers might obtain multiple versions of the same dataset and compare them with the newly released dataset. Although the privacy of all of the datasets has been well-preserved individually, such a comparison can lead to a privacy breach, which is a so-called “incremental privacy breach”.

To address this problem effectively, we first study the characteristics of the effects of multiple dataset releases with a theoretical approach. It has been found that a privacy breach that is subjected to an increment occurs when there is overlap between any parts of the new dataset with any parts of an existing dataset. Based on our

proposed studies, a polynomial-time algorithm is proposed. This algorithm needs to consider only one previous version of the dataset, and it can also skip computing the overlapping partitions. Thus, the computational complexity of the proposed algorithm is reduced from $O(n^m)$ to only $O(pn^3)$ where p is the number of partitions, n is the number of tuples, and m is the number of released datasets. At the same time, the privacy of all of the released datasets as well as the optimal solution can be always guaranteed. In addition, experiment results that illustrate the efficiency of our algorithm on real-world datasets are presented.