Ara bellek tabanlı yöntem ile akan mekânsal verilerin ön işlemden geçirilerek kaydedilmesi

Murat TAŞYÜREK[1]*

[1]*Kayseri Üniversitesi, Mühendislik Mimarlık ve Tasarım Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, TÜRKİYE*

Özet

Nesnelerin interneti (IoT - internet of things) teknolojisinin yaygın olarak kullanılması ile birlikte anlık olarak üretilen veri boyutu ve veri üretim hızı sürekli artmaktadır. IoT tarafından sürekli olarak üretilen bu veriler akan veri (streaming data) olarak adlandırılmaktadır. Akan veri, genel olarak eşzamanlı ve küçük boyutlarda (Kilobayt seviyesinde) gönderilen ve binlerce veri kaynağı tarafından sürekli olarak üretilen verilerdir. Akan veriler, sürekli olarak üretildiğinden dolayı güncel olmaktadır ve günümüzde sıklıkla kullanılan veri madenciliği yöntemleri gibi analiz yöntemleri ile incelenmesi durumunda faydalı bilgiler elde edilmektedir. Akan verilerin analiz edilebilmesi için verilerin uygun formatta veri tabanı sistemlerine veya dosya sistemlerine kaydedilmesi gerekmektedir. Ancak, akan verilerin sistemlere kaydedilmesi verinin hızından ve sürekliliğinden dolayı ve aynı zamanda verinin kaydedilmeden önce veri üzerinde işlem yapılması gereken durumlar olabileceğinden dolayı oldukça zor bir problemdir. Bu çalışmada, Kayseri Büyükşehir Belediyesi (KBB) ulaşım araçlarında bulunan validator cihazları tarafından sürekli olarak gönderilen ve akan veri sınıfında olan verilerin mekânsal veri tabanı sistemlerine ön işlemden geçirilerek kaydedilmesi için ara bellek (buffer) tabanlı bir sistem önerilmiştir. Önerilen yöntemin performansı gerçek veriler üzerinde klasik yöntem ile karşılaştırılmıştır. Deneysel sonuçlar, önerilen ara bellek tabanlı yöntemin klasik yönteme göre daha hızlı cevap verdiğini göstermiştir.

Anahtar Kelimeler: *Akan Veri, Mekânsal Veriler, Veri Tabanı*

Pre-Processing and Saving Streaming Spatial Data Using a Buffer Method

Abstract

With the widespread use of the Internet of Things (IoT), the instantly generated data size and data production speed are constantly increasing. These data continuously produced by IoT are called streaming data. Streaming data is data that is generally sent simultaneously and in small sizes (in kilobyte size) and continuously generated by thousands of data sources. Flowing data is up-to-date because it is constantly produced, and useful information is obtained when analyzed with analysis methods such as data mining methods frequently used today. In order to analyze the flowing data, the data must be recorded in database systems or file systems in an appropriate format. However, saving the flowing data in the systems is a very difficult problem due to the speed and continuity of the data and also because there may be situations where the data must be processed before the data is saved. In this study, a buffer-based system is proposed for pre-processing and recording of data in the data class that is continuously sent and flowing by the validator devices in Kayseri Metropolitan Municipality (KBB) transportation vehicles. The performance of the proposed method was compared with the classical method on real data. Experimental results have shown that the proposed buffer-based method responds faster than the classical method.

Keywords: *Streaming Data, Spatial Data, Database*

1 Giriş

Nesnelerin interneti (IoT- internet of things) teknolojisinin yaygın kullanılması ile birlikte anlık olarak üretilen veri boyutu ve veri üretim hızı sürekli artmaktadır [1, 2]. IoT tarafından sürekli olarak üretilen bu veriler akan veri (streaming data) olarak adlandırılmaktadır [3]. Akan veri, genel olarak eşzamanlı ve küçük boyutlarda (Kilobayt seviyesinde) gönderilen ve binlerce veri kaynağı tarafından sürekli olarak üretilen verilerdir. Akan veriler, sürekli olarak üretildiğinden dolayı güncel olmaktadır ve günümüzde sıklıkla kullanılan veri madenciliği yöntemleri gibi analiz yöntemleri ile incelenmesi durumunda faydalı bilgiler elde edilmektedir [4]. Akan verilerin analiz edilebilmesi için verilerin uygun formatta veri tabanı sistemlerine veya dosya sistemlerine kaydedilmesi gerekmektedir. Bu çalışmada, Kayseri Büyükşehir Belediyesi (KBB) ulaşım araçlarında bulunan validator cihazları tarafından sürekli olarak gönderilen ve akan veri sınıfında olan verilerin mekânsal veri tabanı sistemlerine ön işlemden geçirilerek kaydedilmesi için ara bellek (buffer) tabanlı bir sistem önerilmiştir. Önerilen yöntemin performansı gerçek veriler üzerinde klasik yöntem ile karşılaştırılmıştır.

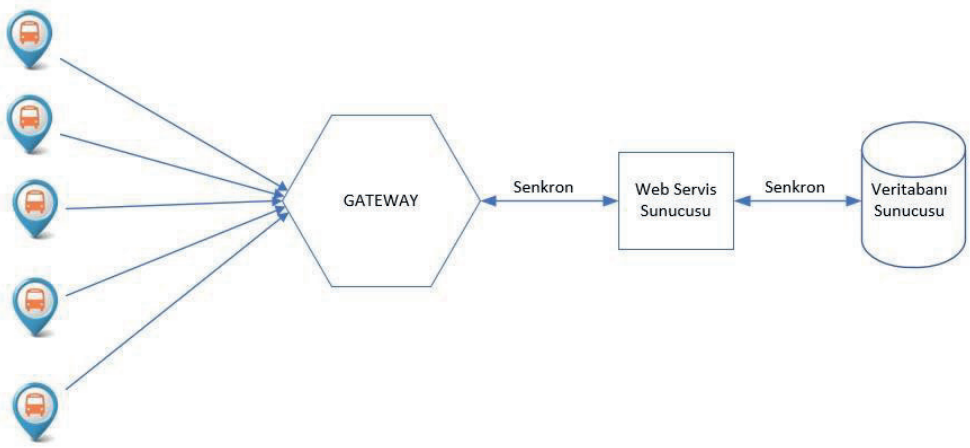
Akan verilerin sistemlere kaydedilmesi verinin hızından ve sürekliliğinden dolayı ve aynı zamanda verinin kaydedilmeden önce veri üzerinde işlem yapılması gereken durumlar olabileceğinden dolayı oldukça zor bir problemdir [5,6]. Akan veriler ilgili sisteme senkron (synchronous) veya asenkron (asynchronous) programlama yöntemleri ile kaydedilebilir [7]. Senkron programlama, yazılımda her bir işlemin sıra ile yapılmasıdır ve bir işlem adımı bitmeden diğer işlem adımına geçilemez [8]. Asenkron programlama; işlemlerin birbirini beklemeden, bağımsız olarak gerçekleşmesidir [9]. Asenkron programlama son yıllarda yaygın olarak kullanılmaktadır. Azadbakht vd. akan verinin özelliklerini yakalamak için asenkron programlama yöntemini kullanmıştır [10]. Hendricks, finans piyasalarında çevrimiçi durum tanımlayıcılarını akış özelliklerini kullanıp gerçek zamanlı kümeleme için asenkron programlama yöntemini kullanmıştır [11].

2 Problem ve Çözümler

Problem, KBB ulaşım sistemi [12] tarafından üretilen akan mekânsal verilerin kaydedilmesi ve ulaşım sistemlerinde online olarak kullanıcılılara sunulması işleminde ortaya çıkmıştır. İlk olarak akan mekânsal veriler, senkron programlama yöntemi ile ön işlemden geçirilerek veritabanına kaydedilmeye çalışılmıştır. Ancak, geliştirme aşamasında bu sistemin çok uzun sürede cevap verdiği ve akan verileri kaydedemediği görülmüştür. Bu sorunun üstesinden gelmek ve akan mekânsal verileri ön işlemden geçirerek veritabanına kaydetmek için asenkron programlama yöntemini kullanan araba bellek tabanlı yöntem önerilmiştir.

2.1 Senkron Yöntem ile Klasik Çözüm

KBB ulaşım araçlarında bulunan validator cihazları tarafından sürekli olarak gönderilen ve akan veri sınıfında olan veriler otobüs yan numarası, GPS konum bilgisi, anlık yolcu sayısı ve otobüs hat numarasından oluşmaktadır. Gelen bu verilerden aracın hangi hattın gidiş veya dönüş güzergahında çalıştığını bulmak ve aracın güzergâh üzerindeki



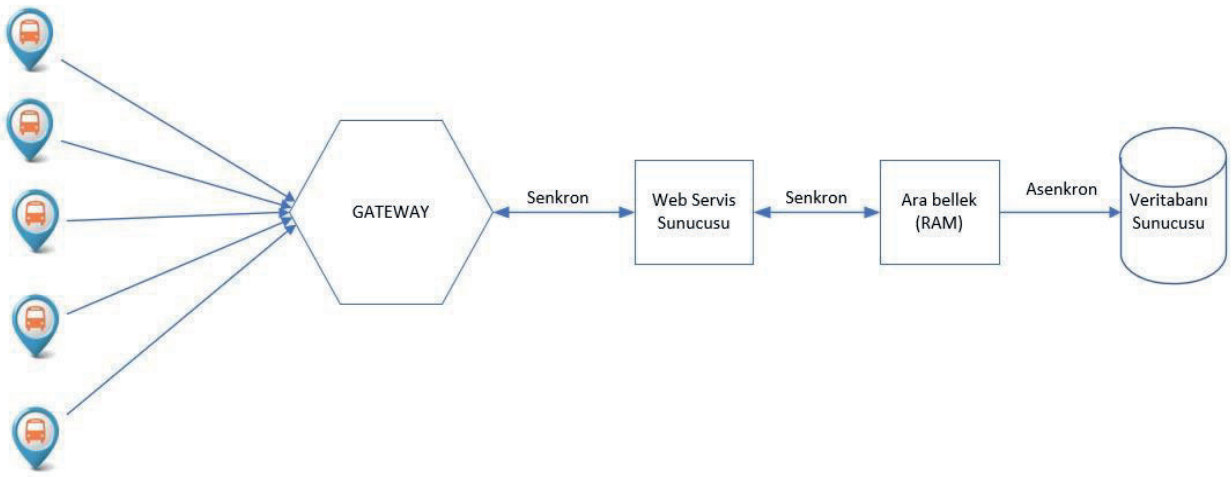
Şekil 1. Senkron Yöntem ile Akan Mekânsal Verilerin Kaydedilmesi

konumunu tespit ederek kullanıcılara harita üzerinde göstermek için verinin ön işlemden geçirilerek kaydedilmesi gerekmektedir. Hat güzergâh bilgisi mekânsal olarak veritabanında tutulduğundan dolayı ön işlemlerin veritabanında gerçekleştirilmesi işlemin süresini hızlandırmaktadır. Şekil 1'de KBB ulaşım validatör cihazları tarafından üretilen akan mekânsal verilerin senkron yöntem ile ön işlemden geçirilerek kaydedilmesi gösterilmiştir. KBB Ulaşım sisteminde 1200 üzerinde araç aktif olarak çalışmakta ve saniye bazında anlık olarak veri göndermektedir. Gateway sunucusu bu verileri senkron yöntem ile web servis sunucusuna iletmekte ve yöntem senkron olduğu için ancak işlem bitince bir sonraki verileri gönderebilmektedir. Web servis sunucusu verileri veritabanına senkron olarak kaydedip cevap verdiğinden dolayı işlemin süresi uzamaktadır. Bu durumda Gateway sunucusu verili geç göndermekte ve akan veriler sisteme zamanında kaydedilememektedir. Bu sorunun üstesinden gelmek ve akan verileri anında sisteme kaydedebilmek için asenkron yöntemi kullanan ara bellek tabanlı bir model önerilmiştir.

2.2 Asenkron Yöntem ile Ara Bellek Tabanlı Çözüm

KBB ulaşım sisteminde saniyede yüzlerce otobüse ait akan mekânsal veriler oluşmakta ve bu verilerin saniyeler içerinde ön işlemden geçirilerek kaydedilmesi gerekmektedir. Gateway sunucusu verileri senkron yöntem ile gönderdiğinden dolayı yönteme cevap gelince ancak sıradaki veriyi gönderilebilmektedir. Bu sorunun üstesinden gelmek için asenkron yöntem kullanan ara bellek tabanlı sistem önerilmiş ve önerilen sistemin mimarisi Şekil 2'de gösterilmiştir. Şekil 2'de Gateway sunucusundan gelen veriler web servis sunucusu tarafından senkron yöntem ile ön belleğe aktarılmaktadır ve Gateway sunucusuna mili saniyeler içerisinde verinin kaydedildiğine dair cevap dönmektedir. Bu işlem sayesinde Gateway sunucusu akan verileri sürekli olarak gönderebilmektedir. Ancak, yüzlerce akan mekânsal verilerin verileri ön işlemden geçirerek veritabanına kaydedilmesi mili saniyelerden daha uzun sürmektedir. Önerilen sistemde akan mekânsal veriler ön bellekte yığın şeklinde tutulmaktadır. Yeni bir araç mekânsal bilgisi geldiği zaman eğer ön bellekte daha önceden araca ait bir bilgi var ise o bilgi güncellenmekte yok ise yığına eklenmektedir. Ara bellekte bulunan veriler asenkron yöntem ile veritabanı sunucusunda ön işlemden geçirilerek kaydedilmektedir. Bu ön işlem esnasında oluşan zaman kaybı işlemler birbirini etkilemediğinden dolayı Gateway sunucusunun veri gönderme sıklığını etkilememektedir.

Ara bellek ve veritabanı sunucusu arasındaki asenkron yöntem belirli aralıklara otomatik olarak çalışmaktadır. Bu işlem sayesinde ara bellekte bulunan bütün veriler veri tabanına ön işlemden geçirilerek diğer sunucularda herhangi bir bekleme veya kitlenmeye yol açmadan veritabanı sunucusuna kaydedilmektedir.



Şekil 2 Asenkron Yöntem ile Akan Mekânsal Verilerin Ara Bellek Kullanarak Kaydedilmesi

3 Araştırma Sonuçları ve Tartışma

Bu çalışmada KBB ulaşım sistemine ait akan mekânsal veriler kullanılmıştır. KBB ulaşım sistemine ait verilerin bilgisi Tablo 1'de sunulmuştur.

Tablo 1. KBB Ulaşım Sistemi Bilgileri

Tablo Bilgisi	Veri Türü	Kayıt Sayısı
Hat güzergâh bilgisi	Çizgi formatında geometrik veri	1280
Otobüs durak sayısı	Nokta formatında mekânsal veri	6600
Otobüs araçları	Akan mekânsal veri	1207

KBB ulaşım sisteminde 1280 adet otobüs güzergahı, 6600 adet otobüs durağı ve 1207 adet bu sistem üzerinde çalışan araç bulunmaktadır. KBB ulaşım sisteminde 1207 araç tarafından anlık olarak akan mekânsal veri üretilmektedir. Bu mekânsal veriler ön işlemden geçirilerek veritabanına kaydedilmekte ve kullanıcılara sunulmaktadır.

Bu çalışmada yapılan deneyler Intel Xenon Gold 5118 CPU 2.30 GHz (8 Çekirdek) 32 GB Ram 300 GB SSD disk özellikleri ve Windows Server 2016 işletim sistemine sahip sunucu üzerinde yapılmıştır.

3.1 Deneysel Sonuçlar

Deneysel değerlendirme için kullanılan 20 adet akan verinin veri geliş sırası ve veri sayısı Tablo 2'de sunulmuştur.

Tablo 2. KBB Akan Mekânsal Veri

Veri Geliş Sırası	Veri Sayısı
1	90
2	123
3	40
4	150
5	60
6	152
7	139
8	144
9	98
10	150
11	146
12	128
13	133
14	118
15	150
16	142
17	124
18	145
19	130
20	125

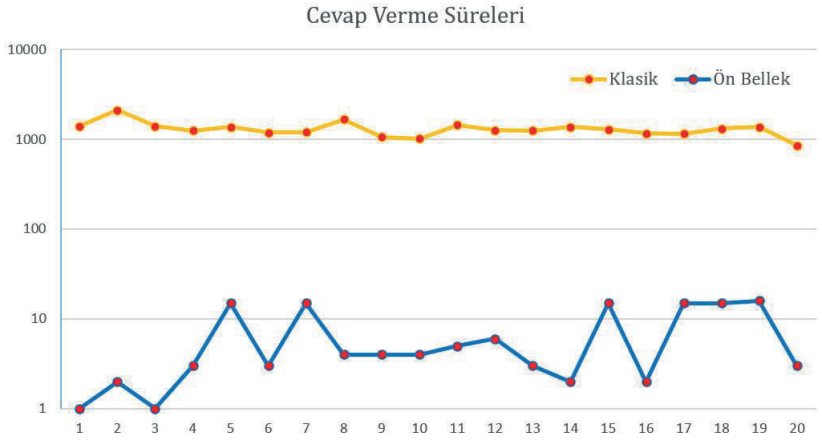
Tablo 2'de bulunan 20 adet akan veri için senkron yöntem ile verileri ön işlemden geçirerek veritabanına kaydeden klasik yöntem ve asenkron yöntem ile verileri ön bellekte tutarak ön işlemden geçiren ön bellek tabanlı yöntemin cevap verme süreleri salise cinsinden incelenmiştir. Klasik yöntem ve ön bellek tabanlı yöntemlerin akan mekânsal verilerin ön işlemden geçirilerek veritabanına kaydetme işlemi esnasında her bir akan veriye cevap verme süresi Tablo 3'te sunulmuştur.

Tablo 3. Yöntemlerin Cevap Verme Süreleri

Veri Geliş Sırası	Klasik	Ön Bellek
1	1406	1
2	2129	2
3	1404	1
4	1250	3
5	1375	15
6	1187	3
7	1203	15
8	1681	4
9	1062	4
10	1020	4
11	1453	5
12	1265	6
13	1249	3
14	1374	2
15	1296	15
16	1171	2
17	1156	15
18	1317	15
19	1374	16
20	859	3

Klasik yöntem ve önerilen ön bellek tabanlı yöntemin cevap verme süreleri Şekil 3'te logaritmik ölçekte gösterilmiştir. Asenkron yöntem kullanarak verileri ön bellekte tutup ön işlemden geçiren önerilen

yöntem, senkron yöntem kullanarak verileri ön işlemden geçiren klasik yönteme göre çok daha hızlı sürede cevap vermiştir. Senkron tabanlı yöntem veriler ön işlemden geçip bütün işlem adımları bittiğinde cevap verdiğinden dolayı cevap verme süresi çok uzundur. Asenkron tabanlı yöntem gelen verilere anında ön belleğe alarak verilerin ön işlem adımlarını beklemeden anında cevap verdiğinden dolayı cevap verme süresi hızlıdır.



Şekil 3 Yöntemlerin Cevap Verme Süreleri

4 Sonuçlar

Akan verilerin analiz edilebilmesi için verilerin uygun formatta veri tabanı sistemlerine veya dosya sistemlerine kaydedilmesi gerekmektedir. Bu çalışmada KBB ulaşım sisteminde validatör cihazları tarafından üretilen gerçek akan mekânsal verilerin ön işlemden geçirilerek veritabanına kaydedilmesi incelenmiştir. Deneysel değerlendirmeler, asenkron yöntem kullanan verileri ön bellekte tutarak ön işlemden geçiren önerilen yöntemin senkron yöntem ile verileri ön işlemden geçirerek veritabanına kaydeden klasik yönteme göre çok daha hızlı sürede cevap verdiğini göstermiştir.

Teşekkür

Bu çalışma kapsamında kullanılan veri temininde sık güncellenen mekânsal verileri paylaştığı için Kayseri Büyükşehir Belediyesi'ne teşekkür ederiz.

Kaynaklar

[1] Tasyurek M, Celik M. RNN-GWR: A geographically weighted regression approach for frequently updated data. *Neurocomputing*, 399, 258-270, 2020.

[2] Xia F, Yang LT, Wang L, Vinel A. Internet of things. *International journal of communication systems*. 25(9):1101, 2012.

[3] Calbimonte JP, Corcho O, Gray AJ. Enabling ontology-based access to streaming data sources. *International Semantic Web Conference, Springer, Berlin, Heidelberg*. 96-111, 2010.

[4] Taşyürek M, Çelik M. FastGTWR: Hızlı coğrafi ve zamansal ağırlıklı regresyon yaklaşımı. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*. 36(2):715-26, 2021.

[5] Siekkinen M, Hoque MA, Nurminen JK, Aalto M. Streaming over 3G and LTE: How to save smartphone energy in radio access network-friendly way. *Proceedings of the 5th Workshop on Mobile Video*.13-18, 2013.

[6] Ostrovsky R, Skeith WE. Private searching on streaming data. *Annual International Cryptology Conference, Springer, Berlin, Heidelberg*. 223-240, 2005.

[7] Cristian F. Synchronous and asynchronous. *Communications of the ACM*. 39(4):88-97, 1996.

[8] Halbwachs N. Synchronous programming of reactive systems. *Springer Science & Business Media*, 2013.

[9] Syme D, Petricek T, Lomov D. The F# asynchronous programming model. *International Symposium on Practical Aspects of Declarative Languages, Springer, Berlin, Heidelberg*. 175-189, 2011.

[10] Azadbakht K, Bezirgiannis N, de Boer FS. On futures for streaming data in abs. *International Conference on Formal Techniques for Distributed Objects, Components, and Systems, Springer, Cham*. 67-73, 2017.

[11] Hendricks D. Using real-time cluster configurations of streaming asynchronous features as online state descriptors in financial markets. *Pattern Recognition Letters.* 97:21-8, 2017.

[12] Taşyürek M, Çelik M. Akıllı Durak Sistemindeki Araç Seyahat Sürelerinin Birleşik Yapay Sinir Ağları Kullanarak Tahmini. *Avrupa Bilim ve Teknoloji Dergisi.*72-79, 2020.