

Delovni okvir za pretočne podatke s standardom common information model (CIM)

Maja Savinek^{1,2}, Matjaz Kukar¹

¹ Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana

² Elektro Ljubljana, Slovenska cesta 56, 1000 Ljubljana

maja.savinek@elektro-ljubljana.si, matjaz.kukar@fri.uni-lj.si

Izveček

Napredno merjenje električne energije je tehnična rešitev za avtomatiziran obračun električne energije. Pametno omrežje je zgolj paradigma, kjer je električno omrežje z njegovimi uporabniki povezano z rešitvami informacijskih in komunikacijskih tehnologij za izboljšanje vseh funkcionalnosti omrežja. Za doseganje tega, je potrebno uporabiti razpoložljive podatke merilnih naprav, senzorjev, polnilnic in drugih naprav v omrežju. Vključevanje teh podatkov predstavlja številne izzive, saj niso vse naprave povezane s tradicionalnimi sistemi zajema podatkov.

Predlagani delovni okvir omogoča zajem podatkov v skoraj realnem času z ustreznim komunikacijskim standardom. Slednje se doseže s pretvorbo podatkov v skladu s standardi za merjenje električne energije na podlagi XSD sheme in transformacije podatkov v formatu CIM. Dognali smo, da s predlaganim delovnim okvirjem čas intervala od zajema, transformiranja do končne razpoložljivosti, je štirikrat učinkovitejši od obstoječih.

Ključne besede: Common information model, napredne merilne naprave, podatkovni tok, upravljanje podatkov

Abstract

Smart metering is a technical solution for emerging markets and billing system. The smart grid is a paradigm where the power grid and its users are connected to information and communication technology solutions, which envisage the improvement of all network functionalities. To achieve this, it is necessary to use the increasingly available data from smart meters, sensors, chargers and other devices. The incorporation of this data is associated with many challenges as not all devices are connected to a traditional data capture system.

The proposed framework enables near real-time data capture with appropriate communication standards. This is achieved by converting values in accordance with the standards for electricity measurement based on the XSD scheme and values transformed into a CIM model. We have found that the usual length of the time interval from capturing the current data on the measuring device to recording the raw current is four times more efficient than the existing frameworks.

Keywords: Common information model, data management, smart meter, stream data

1 UVOD

Elektroenergetska podjetja se soočajo s pospešeno digitalizacijo in posledično z iskanjem novih rešitev za obvladovanje obnovljivih virov energije, dinamičnega razvoja energetskega trga, energetskega storitev ter iskanjem naprednih pristopov upravljanja omrežja. Nekateri izzivi, ki so povezani z obvladljivostjo energije, so stohastično vedenje proizvodnje obnovljivih virov

energije, spremembe smeri pretoka energije v omrežjih ali proizvodnja energije v različnih obdobjih. Prisotnost pametnih števec v nizkonapetostnem omrežju omogoča povezavo končnih deležnikov z omrežjem in sočasno generiranje podatkov. Kljub naprednim tehnologijam in pospešenemu uvajanju naprednih merilnih naprav, predstavlja enovito povezovanje med naprednimi merilnimi sistemi in pametnim omrežjem velik izziv.

V večini primerov napredne merilne naprave komunicirajo preko PLC (Power Line Communication) ali GPRS/3G. Običajni protokol za odčitavanje merilnih podatkov je DLMS/COSEM (standard za izmenjavo merilnih podatkov o električni energiji). Podatki kot so napetost, tok, delovna moč, jalova energija se lahko generirajo skoraj v realnem času. Ti podatki predstavljajo velik potencial za izvedbo napredne analitike in poglobljen pogled v distribucijsko omrežje ter predstavljajo enega od pomembnih vhodov za sistema ADMS (Advanced Distribution Management Systems) in SCADA (Supervisory Control And Data Acquisition), ki sta pogloblitna sistema za upravljanje in nadziranje distribucijskega omrežja.

Iz različnih razlogov kot so zagotavljanje podatkov v standardni obliki za izmenjavo podatkov, nezanesljivosti komunikacije v realnem času z naprednimi merilnimi napravami, zagotavljanje kakovosti masovnih podatkov in izvajanje imputacije manjkajočih vrednosti, smo raziskavo usmerili v razvoj delovnega okvirja, ki omogoča zajem, transformacijo in imputacijo s komunikacijskim standardom za izmenjavo paketnih sporočil v skoraj realnem času.

2 SORODNO DELO

Nekateri avtorji so raziskali področje zajemanja podatkov pametnih števec in svoje delo objavili v prispevkih. Pri pregledu del smo izpostavili ključne točke, ki jih bomo upoštevali pri razvoju delovnega okvirja za zajem in obdelavo merilnih pretočnih podatkov. Avtorji člankov se problematike niso lotili kot celovitega problema, ampak zgolj posameznih področij.

2.1 Delovni okvir za zajem pretočnih podatkov

Koncept velepodatkov je prisoten že več kot desetletje. Kljub temu, da je dobro opredeljen in da na splošno razumemo njegove prednosti, podjetja ne morejo učinkovito izkoristiti podatkov in vseh prednosti, ki jih ta koncept ponuja. Tradicionalne metode upravljanja podatkov, kot so relacijske podatkovne baze, niso sposobne zajeti večletnih, sekundnih in milisekundnih dogodkov, kar vodi do problematike zagotavljanja informacij o dejanskem stanju pametnih omrežij v realnem času [1].

V prispevku [2] avtorji opisujejo problem zajemanja podatkov, ki se generirajo na robu omrežja. Vključevanje novih podatkov predstavlja veliko izzivov, saj niso vse naprave povezane s komunikacijskim

omrežjem. Glavna ideja, ki jo obravnavajo avtorji, je paralelno združiti različne pretočne podatke z zgodovinskimi podatki. Z idejo o virtualnem sistemu SCADA želijo združiti različne formate podatkov na standardiziran vir.

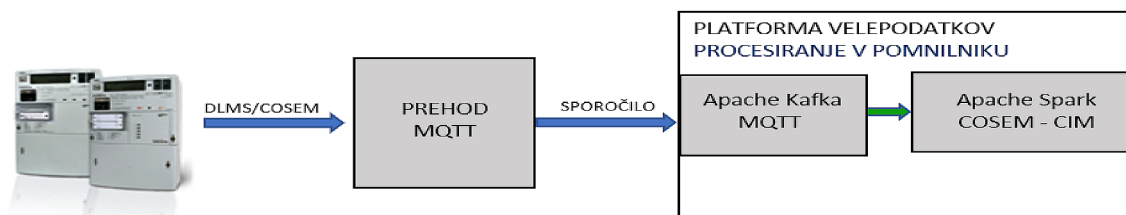
2.2 Izmenjava pretočnih podatkov v formatu CIM

V prispevku avtorji izpostavljajo težave, s katerimi se morajo spoprijeti pri prehodu iz distribucijskega in prenosnega omrežja na pametna omrežja. Pametna omrežja se bodo v prihodnosti soočila s standardizacijo standardnih protokolov, če želijo v celoti izkoristiti ponujene prednosti in funkcionalnosti. Nacionalni inštitut za standarde in tehnologijo (NIST) [4] je opredelil 16 ključnih standardnih protokolov [5] za pametna omrežja, nato dopolnil na 77. Interoperabilnost standardov je mogoče doseči z ontološkim konceptom semantične piramide. To dosežemo z informacijskim modelom, ki združuje definicije, koncepte in povezave med njimi. Prav tako je definiran UML podatkovni model električnega omrežja za CIM (Common Information Model), ki je razširjen na ustrezne protokole in kombiniran z ontologijami drugih protokolov za doseganje interoperabilnosti. V prispevku je poudarek na standardu IEC 60870 [6], kjer v večini primerov izvajajo preslikavo podatkov le za doseganje interoperabilnosti. Predstavili so pristop uporabe SPARQL za merilne naprave, ki ustrezajo formatu CIM.

3 DELOVNI OKVIR ZA PRETOČNE PODATKE

Osnovna dilema predlaganega delovnega okvira je, ali je mogoče vzpostaviti delovni okvir, ki bo v skoraj realnem času zajemala neobdelane podatke, jih prenese v platformo za obvladovanje velepodatkov, kjer se neobdelani podatki v pomnilniku sprocesirajo, imputirajo in transformirajo v format CIM. Delovni okvir vsebuje več komponent, ki so ponazorjene s shematskim prikazom na sliki 1, torej napredne merilne naprave, sistem zajema podatkov (prehod) in procesiranje podatkov v pomnilniku.

Napredne merilne naprave so opremljene s komunikacijskim modulom, ki v skladu s standardom DLMS / COSEM komunicira s sporočilnim sistemom MQTT. MQTT je protokol za prenos podatkov IoT (internet stvari). Na desni strani slike 1 je prikazana obdelava podatkov v pomnilniku na velepodatkovni platformi, ki je sestavljena iz zajemanja pretočnih podatkov s komponento Apache Kafka MQTT in raz-



Slika 1: Semantična predstavitev komponent predlaganega okvira za zajem in obdelavo podatkov v pomnilniku s COSEM-CIM razčlenjevalnikom v formatu CIM

členjevalnikom Apache Spark COSEM-CIM. Apache Kafka je prek protokola MQTT povezan s preходом, kjer pridobi neobdelane pretočne podatke in skrbi za prenos teh podatkov v Apache Spark za procesiranje v pomnilniku. Ko se podatki nahajajo v Apache Spark, se v pomnilniku izvede obdelava z razčlenjevalnikom COSEM-CIM.

Razčlenjevalnik COSEM-CIM vsebuje metodo za pretvorbo neobdelanih podatkov v strukturirane podatke, matematični algoritem za preoblikovanje ter združevanje vrednosti, algoritem za imputacijo manjkajočih vrednosti in komponento za pretvorbo v format CIM. Tako obdelani podatki se v zadnjem koraku zapišejo v platformo za obvladovanje velepodatkov. Razčlenjevalnik COSEM-CIM uporablja standardizirano obliko podatkov CIM XML v standardu IEC61968 [3].

3.1 Opredelitev testnega okolja

Za potrebe vrednotenja delovanja predlaganega delovnega okvirja pretočnih podatkov (COSEM-CIM parser) se je vzpostavilo testno okolje, ki je vsebovalo 50 naprednih merilnih naprav MT880, opremljenih s komunikacijskim modulom UMTS, MQTT sporočilno infrastrukturo ter platformo za upravljanje velepodatkov in procesiranje v pomnilniku. Celotno okolje je delovalo na operacijskem sistemu Linux CentOS z 64 GB RAM, 16 vCPU in 1TB diskovnega prostora.

Testno okolje AMI je vsebovalo tri ključne komponente in sicer 50 naprednih merilnih naprav MT880, sistem za upravljanje merilnih podatkov MDMS (Meter Data Management System) in relacijsko podatkovno bazo. Okolje je delovalo na operacijskem sistemu Windows server 2016 in SQL strežniku z enako konfiguracijo strojne opreme kot delovni okvir COSEM-CIM parser.

RDF SPARQL okolje je delovalo na enaki infrastrukturi kot COSEM-CIM parser. Odstopanje je bilo

v konfiguraciji hranjenja in obdelovanja podatkov v Apache Spark.

Delovanje testnih okolij na različnih operacijskih sistemih bi lahko privedlo do odstopanj v hitrosti delovanja posameznih komponent. Faktor operacijskih sistemov bi bil zelo pomemben pri izvajanju meritev v milisekundah. Izvedba naših meritev je temeljila na najnižji časovni enoti sekundi, ki je uvedena v sistemskih zapisih. Na podlagi teh dejstev smo faktor vpliva operacijskih sistemov zanemarili.

4 REZULTATI IN RAZPRAVA

V tem razdelku predstavljamo primerjavo med okvirji, in sicer AMI (Advance Metering Infrastructure), RDF SPARQL in razčlenjevalnikom COSEM-CIM. Izhodiščna vprašanja raziskovanja so bila:

- primerjava zagotavljanja podatkov v realnem času in
- vpliv parametrov na čas obdelave in pretvorbo podatkov v realnem času v formatu CIM.

Vsaka napredna merilna naprava vsebuje frekvenco prenosa podatkov in registre merilnih vrednosti toka (I) in napetosti (U). Frekvenca intervalov zajema podatkov je dosegla največ 1 minuto. Glavno vprašanje je, ali je mogoče vzpostaviti infrastrukturo, ki bo sposobna v realnem času zajemati neobdelane podatke in jih prenašati v velepodatkovno platformo, kjer se bodo podatki v pomnilniku obdelali in transformirali v format CIM. Test je bil zasnovan tako, da je upošteval natančen čas zajema za vseh 50 števecov točno takrat, ko so se v posredniku pojavili neobdelani podatki, čas obdelave in pretvorbe ter čas ponovne poizvedbe. Testiranje se je ponovilo v 100 iteracijah. Izračunani povprečni čas iz sistemskih zapisov zajema podatkov od posrednika do ponovne poizvedbe, omogoča oceno dejanskega doprinosa obdelave in transformacije podatkov. S standardnim odklonom smo izmerili razpršenost izračunanih povprečnih časov okoli aritmetične sredine vzorcev.

Tabela 1: V tabelarični obliki so prikazani rezultati testiranja časovne primerjave posameznih delovnih okvirjev od zajema do transformacije podatkov. Spremenljivka Stream definira tip pretočnosti podatka. Vrednost t_1 prikazuje čas zajema podatkov, t_2 prikazuje čas obdelave in transformacije v format CIM, t_3 čas, ko so podatki ponovno na voljo v formatu CIM. Skupni čas obdelave je zaveden v stolpcu t_s .

Delovni okvir	Pretočnost podatkov	t_1	σ_1	t_2	σ_2	t_3	σ_3	t_s	σ_s
AMI	ne	6,27 min	0,35	–	–	–	–	–	–
RDF SPARQL	da	0,518 s	0,18	5,527 s	0,23	5,98 s	0,20	12,61 s	0,21
COSEM - CIM	da	0,492 s	0,13	2,648 s	0,19	0 s	–	3,14 s	0,17

Tabela 1 vsebuje primerjavo rezultatov testiranja zajem in transformiranja podatki v skoraj realnem času s tremi različnimi pristopi.

Prvi delovni okvir AMI, zasnovan z inženirskega vidika, ni omogočal zajema podatkov v skoraj realnem času, kar je privedlo do časa zajema podatkov 6,27 min s standardnim odklonom 0,35. Drugi delovni okvir RDF SPARQL, zasnovan z znanstvenega vidika, omogoča zajem podatkov v realnem času. Povprečen čas zajema podatkov je 0,518 s s standardnim odklonom 0,18. V tretjem predlaganem okvirju COSEM-CIM parser je povprečni čas zajema podatkov 0,492 s s standardnim odklonom 0,13.

V naslednjem koraku testiranja smo preverili koliko časa izbrana delovna okvirja RDF SPARQL in COSEM-CIM potrebuje za obdelavo ter transformacijo podatkov v format CIM in omogočanje razpoložljivosti podatkov za nadaljnjo poizvedovanje. RDF SPARQL je potreboval enak povprečni čas za zajem neobdelanih podatkov kot COSEM-CIM parser. V naslednjem koraku, kjer se podatki obdelajo in transformirajo, je COSEM-CIM parser porabil manj časa kot RDF SPARQL, saj slednji neobdelane podatke obdela, transformira, shrani in jih na zahtevo pretvori v CIM format. COSEM CIM parser podatke paralelno obdela in hkrati transformira. Predlagani okvir je štirikrat učinkovitejši od RDF SPARQL, kar je razvidno v tabeli 1, torej COSEM CIM parser porabi 3,14 s s standardnim odklonom 0,17 in RDF SPARQL 12,61 s s standardnim odklonom 0,21. Prednost našega predlaganega okvirja je, da se vsi matematični algoritmi in transformacije izvedejo paralelno v pomnilniku.

4.1 Razprava

Na podlagi izvedenih meritev smo prišli do zaključka, da je s predlaganim delovnim okvirjem izvedljivo zajeti in transformirati pretočne podatke v format CIM v realnem času.

Pri raziskavi in definiranju delovnega okvirja smo se uprli na ugotovitve v sorodnih raziskavah, predvsem na članek [4], kjer so avtorji uporabili hranjenje neobdelanih podatkov in dostop v RDF SPARQL-u. Slabost RDF SPARQL pristopa je povpraševanje po podatkih, saj je časovno zelo potratno. S pristopom COSEM-CIM se podatki zajamejo, vzporedno obdelajo, transformirajo v pomnilniku in shranijo v formatu CIM, kar omogoča štirikrat učinkovitejše delovanje predlaganega delovnega okvirja.

5 ZAKLJUČEK

Raziskovalno delo je usmerjeno v razvoj delovnega okvirja za zajem, obdelavo in procesiranje v pomnilniku pretočnih podatkov v skoraj realnem času. Naš cilj je ponuditi obdelane in standardizirane podatke za shranjevanje ali integracijo z različnimi aplikacijami v realnem okolju.

Ugotovitve temeljijo na rezultatih testiranj in primerjave z drugimi okvirji, ki so pokazale, da predlagan okvir zmanjša čas celotne obdelave za štirikrat. Pri pregledu znanstvenih del s področja predlaganega okvirja ni bilo zaznati, da bi raziskana literatura obravnavala področje kot celoto, ki je predstavljeno v našem predlaganem okvirju.

Na podlagi pregleda znanstvenih člankov smo nekatere rešitve uporabili, jih nadgradili ter jih ponudili kot izboljšane celote. Za testiranje okvirja smo uporabili tehnologijo, komponente in strojno opremo, kar ne pomeni, da je bila optimalno izbrana.

REFERENCE

- [1] Tabue, B., et al.: Proactive transmission and distribution asset management (2016).
- [2] Avo Sevlian, R., et al.: Visualization and Analytics for Distributed Energy Resources, Cornell University (2017).
- [3] Simmons, J.: Common Information Model Primer, EPRI (2015).
- [4] Crapo, A., et al: The semantically enabled smart grid, in Proceedings of the Grid-Interop Forum, pp. 177-185 (2009).

- [5] NIST Framework and Roadmap for Smart Grid Interoperability Standards, 1st ed., National Institute of Standards and Technology (2009).
- [6] Peña, K., et al: Semantic integration of IEC 60870 into CIM, IEEE (2011).
- [7] Rohr, M., et al: Using CIM for Smart Grid ICT integration, BTC, pp. 45-61 (2011).

■

Maja Savinek je doktorska študentka na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Zaposlena je v podjetju Elektro Ljubljana, d. d., kjer se ukvarja z razvojem napredne analitike in algoritmov na področju elektroenergetike, s poudarkom na tehnologijah velepodatkov.

■

Matjaž Kukar je docent na Fakulteti za računalništvo in informatiko, član Laboratorija za kognitivno modeliranje. Leta 2001 je doktoriral iz računalništva in informatike, leta 1996 magistriral in leta 1993 diplomiral na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Njegove raziskave so na področju umetne inteligence, predvsem strojnega učenja, podatkovnega rudarjenja in analize podatkov.