Ľubomír Pallaj, Jaroslav Petráš, Ardian Hyseni, František Margita

Analýza integrity veľkých dát v systémoch inteligentného merania pomocou detekčných mechanizmov

Článok sa zaoberá analýzou integrity veľkých dát v systémoch inteligentného merania (AMI) so zameraním na detekciu cielených útokov na namerané hodnoty. Vzhľadom na obmedzenú dostupnosť reálnych datasetov bol navrhnutý generátor syntetických dát reprezentujúci virtuálny inteligentný merač, ktorého výstupy aproximujú charakteristiky reálnych meraní a poskytujú dostatočný rozsah dát pre vývoj a testovanie detekčných mechanizmov. Na syntetických datasetoch bola aplikovaná sada útočných algoritmov simulujúcich rôzne scenáre narušenia integrity dát. Pre jednotlivé typy útokov boli následne navrhnuté detekčné mechanizmy založené na štatistickej analýze časových radov a detekcii anomálií. Dosiahnuté výsledky potvrdzujú schopnosť systému detegovať široké spektrum útokov pri súčasnej minimalizácii falošne pozitívnych detekcií.

Kľúčové slová: inteligentné meranie, AMI, integrita dát, detekcia anomálií, útočné algoritmy, syntetické dáta, smart grid

This paper addresses the analysis of big data integrity in Advanced Metering Infrastructure (AMI) systems, with a focus on the detection of targeted attacks on measured data. Due to the limited availability of real-world datasets, a synthetic data generator representing a virtual smart meter was developed. The generated data approximate the characteristics of real measurements and provide a sufficiently large dataset for the development and testing of detection mechanisms. A set of attack algorithms simulating various data integrity breach scenarios was applied to the synthetic datasets. For each type of attack, corresponding detection mechanisms based on statistical time series analysis and anomaly detection were designed. The obtained results demonstrate the ability of the proposed system to detect a wide range of data integrity attacks while minimizing the rate of false positive detections. **(Big Data Integrity Analysis in Smart Metering Systems Using Detection Mechanisms)**

Keywords**:** smart metering, AMI, data integrity, anomaly detection, attack algorithms, synthetic data, smart grid

I. ÚVOD

Moderné elektroenergetické systémy prechádzajú transformáciou smerom k inteligentným sieťam (smart grid), ktoré integrujú tradičnú energetickú infraštruktúru s pokročilými informačnými a komunikačnými technológiami, čím vzniká komplexný kyber fyzikálny systém [3], [6]. Táto transformácia umožňuje zlepšenie monitorovania, riadenia, efektívnosti a spoľahlivosti prevádzky elektrizačnej sústavy, ako aj integráciu obnoviteľných zdrojov energie a distribuovaných energetických zdrojov. Dôležitou súčasťou inteligentných sietí je pokročilá meracia infraštruktúra (AMI), ktorá využíva inteligentné merače na zber dát v reálnom čase a umožňuje obojsmernú komunikáciu medzi odberateľmi a prevádzkovateľmi siete [1], [7].

Zvyšujúca sa závislosť na digitálnej komunikácii a dátovo riadených algoritmoch však prináša nové bezpečnostné výzvy, najmä v oblasti integrity a spoľahlivosti dát. Medzi najzávažnejšie kybernetické hrozby patria útoky typu false data injection (FDI), ktoré umožňujú útočníkovi manipulovať merané dáta a ovplyvniť proces odhadu stavu bez detekcie tradičnými metódami [3], [9]. Dôsledky takýchto útokov môžu zahŕňať nesprávne fakturovanie, destabilizáciu siete, manipuláciu s trhom s energiou alebo dokonca rozsiahle výpadky elektrickej energie. Tradičné metódy detekcie chybných dát založené na reziduálnej analýze často nedokážu identifikovať sofistikované útoky, čo vytvára potrebu vývoja pokročilejších detekčných mechanizmov. Zároveň je vývoj a testovanie týchto metód obmedzené nedostupnosťou reálnych dát, ktoré sú často chránené z dôvodu bezpečnosti a ochrany kritickej infraštruktúry [4], [8].

Cieľom tohto článku je analyzovať integritu dát v systémoch AMI prostredníctvom generovania syntetických dát, modelovania útokov a návrhu detekčných mechanizmov založených na štatistickej analýze a detekcii anomálií.

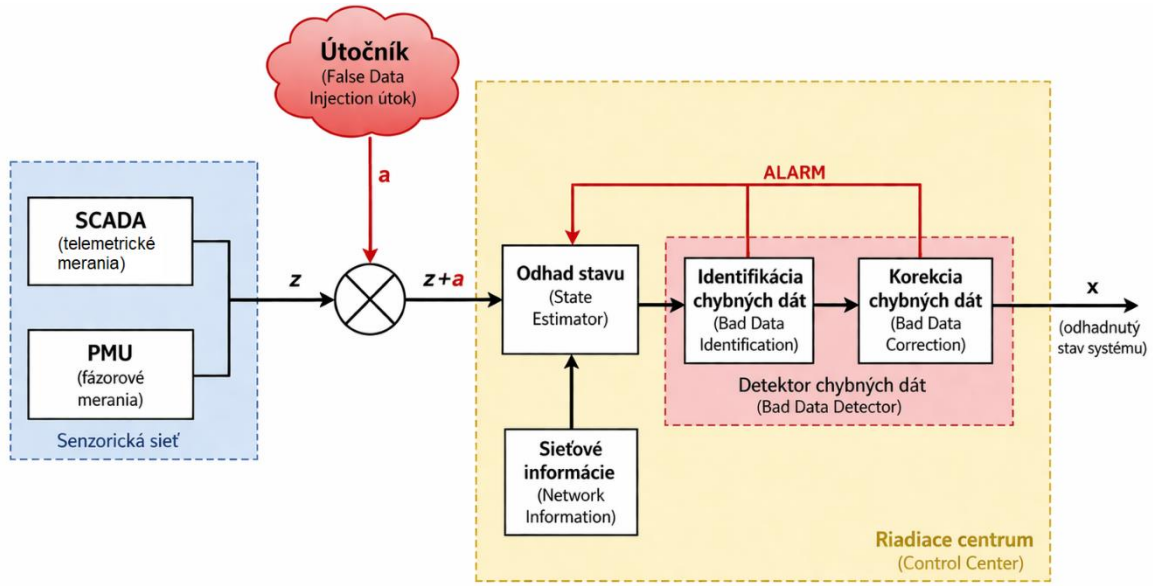
II. FORMULÁCIA PROBLÉMU A POPIS DÁT

S rastúcou digitalizáciou elektrizačných sústav sa zvyšuje význam spoľahlivosti a bezpečnosti spracovania meraných dát, ktoré sú základom pre riadenie a optimalizáciu prevádzky inteligentných sietí. Integrácia pokročilých meracích a komunikačných technológií síce umožňuje detailné monitorovanie systému, no zároveň rozširuje priestor pre potenciálne kybernetické útoky zamerané na manipuláciu dát [3], [6].

Z tohto dôvodu je potrebné detailne analyzovať architektúru AMI systémov, identifikovať zraniteľné miesta a pochopiť charakter dostupných dát, ktoré sú využívané pri návrhu detekčných mechanizmov. Táto kapitola sa preto zameriava na opis štruktúry AMI a problematiky integrity dát, ako aj na charakteristiky a obmedzenia reálnych datasetov používaných pri analýze.

1. Pokročilá meracia infraštruktúra a integrita dát

Inteligentná elektrizačná sieť predstavuje komplexný kyber-fyzikálny systém, ktorý prepája energetickú infraštruktúru s komunikačnými sieťami, senzorickými zariadeniami a riadiacimi systémami [2], [3], [6].



Obr. 1. Vývojový diagram odhadu stavu elektrizačnej sústavy.

V rámci tejto architektúry zohráva AMI kľúčovú úlohu, keďže umožňuje kontinuálne monitorovanie spotreby elektrickej energie a prenos meraných dát do riadiacich centier. Inteligentné merače zhromažďujú dáta o spotrebe a odosielajú ich do systémov, ako je SCADA, kde sa vykonáva odhad stavu a prijímajú sa rozhodnutia o riadení siete [3].

Presnosť týchto meraní je zásadná, pretože mnohé riadiace a optimalizačné procesy v energetike sú priamo závislé od správnosti odhadnutého stavu systému. Zavedenie komunikačných technológií však zároveň zvyšuje zraniteľnosť systému voči kybernetickým útokom, ako sú spoofing alebo útoky typu FDI [9]. FDI útoky sú obzvlášť nebezpečné, pretože umožňujú koordinovanú manipuláciu meraní tak, že uniknú detekcii klasických metód. Tieto útoky cielia na proces odhadu stavu, čím môžu viesť k nesprávnym rozhodnutiam riadiaceho systému a ohrozeniu stability elektrizačnej sústavy [7].

Proces odhadu stavu elektrizačnej sústavy je znázornený na Obr. 1 a pozostáva z dvoch hlavných častí: senzorickej siete a riadiaceho centra. Senzory v sieti prenášajú merania do riadiaceho centra, ktoré spracúva sieťové informácie, pričom algoritmus odhadu stavu využíva metódu najmenších štvorcov na určenie stavu siete, ako sú napríklad napätie a fázový uhol jednotlivých uzlov. Ak merania prejdú testom detektora chybných dát, výsledný odhad stavu sa považuje za finálny. V opačnom prípade je potrebné merania opätovne skontrolovať a vykonať ich korekciu.

2. *Charakteristiky a obmedzenia reálnych dát*

Pre návrh a testovanie detekčných mechanizmov je nevyhnutné pracovať s dátami, ktoré reprezentujú reálne prevádzkové podmienky elektrizačnej sústavy. Takéto dáta sú získavané z inteligentných meračov, SCADA systémov alebo PMU jednotiek, ktoré poskytujú detailný obraz o stave systému v čase [5], V praxi sú však tieto datasety často obmedzené z hľadiska rozsahu aj dostupnosti, keďže ide o citlivé údaje kritickej infraštruktúry, ktoré prevádzkovatelia distribučných sietí z bezpečnostných a regulačných dôvodov spravidla nezverejňujú.

Ďalším zásadným problémom je absencia označených kybernetických útokov v reálnych dátach sa reálne prevádzkové merania prirodzene neobsahujú príklady útokov, čo znemožňuje ich priame využitie na objektívne hodnotenie a testovanie detekčných mechanizmov [5]. Z tohto dôvodu sa v súčasnosti čoraz viac využívajú syntetické dáta, ktoré umožňujú simulovať realistické scenáre spotreby aj rôzne typy útokov a predstavujú štandardný a široko akceptovaný prístup v oblasti výskumu bezpečnosti inteligentných sietí [4], [5], [10].

III. GENEROVANIE SYNTETICKÝCH DÁT A MODELOVANIE ÚTOKOV

Vzhľadom na obmedzenú dostupnosť reálnych meraní obsahujúcich kybernetické útoky je pre účely návrhu a testovania detekčných mechanizmov, ktoré sú definované ako binárne funkcie $f_i(x) \in \{0,1\}$, a sú vhodné využívať syntetický generované dáta. Model syntetického inteligentného merača bol implementovaný ako parametrizovateľný generátor dát, ktorý umožňuje nastavovať charakteristiky simulovaných meraní. Medzi hlavné parametre patria počet simulovaných meračov, časový rozsah generovaných dát, časový krok merania a typ denného profilu spotreby ako rezidenčný alebo industriálny profil. Okrem toho model umožňuje definovať priemernú spotrebu energie a časové obdobie jej vyhodnocovania, čo umožňuje prispôsobiť generované dáta rôznym typom odberateľov. Na simuláciu realistického správania systému je do modelu zahrnutá aj náhodná zložka riadená parametrom náhodnosti a typom šumu, ktorý reprezentuje prirodzené fluktuácie meraní.

TABUĽKA I
Príklad konfigurácie generátora

Parameter	Hodnota	Popis
num_smart_ meters	10	Počet simulovaných inteligentných meračov
start	01. 01. 2025	Dátum začiatku simulačného obdobia
end	31. 12. 2025	Dátum ukončenia simulačného obdobia
interval_min	15	Interval merania v minútach
randomness	0.88	Stupeň náhodnosti aplikovaný na profily
noise_type	gaussian	Typ šumu pridaného k signálu
daily_profile	residential	Šablóna profilu dennej spotreby
average_energy_ kwh	12.5	Priemerná spotreba energie za obdobie (kWh)
average_energy _period	day	Referenčné obdobie pre priemernú energiu

Príklad konfigurácie generátora dát je znázornený na TABUĽKA I, kde sú uvedené základné parametre použité pri generovaní syntetických dát.

1. Scenáre útokov

Na synteticky generované dáta boli aplikované viaceré typy útokov simulujúce rôzne formy narušenia integrity meraní. Jednotlivé útoky sa líšia spôsobom modifikácie dát, pričom zasahujú buď do hodnôt meraní, časových značiek, alebo do úplnosti dátového záznamu.

1.1. **Data tampering** (narušenie hodnôt merania pomocou šumu)

Tento útok predstavuje priamu manipuláciu nameraných hodnôt vo vybraných stĺpcoch datasetu. V definovanom časovom intervale, prípadne v celom datasete, je ku každej hodnote pripočítaná náhodná zložka generovaná z normálneho rozdelenia. Veľkosť tejto poruchy je riadená parametrom *intensity* a je úmerná smerodajnej odchýlke daného stĺpca. Útok nemení základný trend signálu, ale narúša jeho priebeh pridaním šumu, čím vznikajú lokálne odchýlky indikujúce narušenie integrity dát.

1.2. **Replay attack** (opakovanie historických dát)

Replay útok nahrádza aktuálne merania historickými hodnotami. Implementácia spočíva vo výbere časového okna, z ktorého sa skopírujú pôvodné dáta, a ich následnom vložení do iného časového úseku datasetu. Veľkosť okna je určená parametrom *replay_frac* alebo explicitne definovaným časovým intervalom.

Napadnutý úsek obsahuje konzistentné, avšak neaktuálne hodnoty, ktoré nezodpovedajú skutočnému stavu systému. Útok preto nevytvára výrazné odľahlé hodnoty, ale narúša časovú autenticitu dát, čo sťažuje jeho detekciu.

1.3. **Time-shift** (posun časových značiek)

Útok typu *time-shift* nemení hodnoty meraní, ale modifikuje ich časové značky. Vybrané záznamy alebo celý dataset sú posunuté o časový interval definovaný parametrom *shift*. Posun môže byť aplikovaný globálne alebo len na definované časové okno.

Táto modifikácia narúša časovú synchronizáciu meraní a môže viesť k nesprávnej interpretácii dát v procesoch odhadu stavu. Pri aktivovaní parametra *wrap* sa časové značky cyklicky mapujú do pôvodného intervalu, čím vzniká ťažšie identifikovateľná forma narušenia.

1.4. **Selective dropping** (selektívne vynechávanie dát)

Tento útok simuluje situáciu, keď časť meraní nie je dostupná v dôsledku poruchy alebo úmyselného zásahu. Výber odstránených záznamov je realizovaný podľa definovaných pravidiel, napríklad na základe hodín dňa, dní v týždni, víkendov alebo konkrétnych časových intervalov. Rozsah odstránenia je riadený parametrom *drop_fraction*.

Výsledkom útoku je vznik medzier v časovom rade, čo narúša kontinuitu dát a môže negatívne ovplyvniť štatistickú analýzu aj detekčné algoritmy.

1.5. **Data injection** (vkladanie nových záznamov)

Tento útok pridáva do datasetu nové, umelo vytvorené záznamy, ktoré simulujú falošné merania. Implementácia podporuje viacero režimov, ktoré sa líšia charakterom generovaných dát.

a) duplicate_ts. V tomto režime sú vytvorené nové záznamy s duplicitnou časovou značkou prevzatou z existujúcich dát. Hodnoty meraní sú však modifikované náhodnou odchýlkou okolo priemeru príslušných veličín. Výsledkom je vznik viacerých rozdielnych meraní pre rovnaký časový okamih.

b) out_of_range. V tomto prípade sú generované záznamy s mierne posunutou časovou značkou a extrémnymi hodnotami. Hodnoty sú vytvorené ako odchýlka od priemeru o násobok smerodajnej odchýlky riadený parametrom *outlier_sigma*. Takto vznikajú odľahlé hodnoty reprezentujúce fyzikálne nepravdepodobné merania.

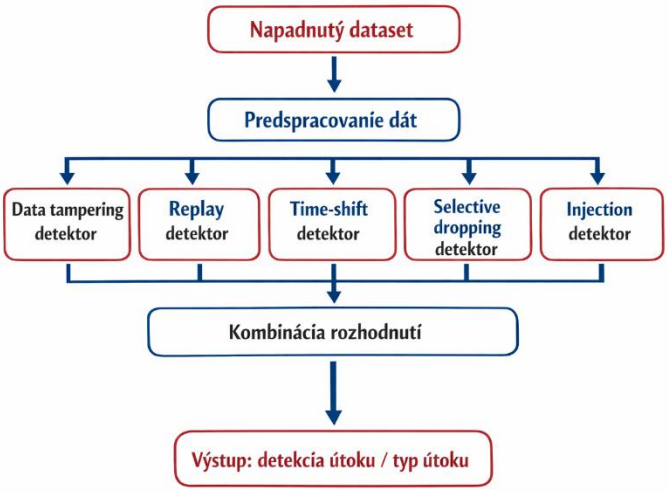
c) implausible_combo. Tento režim vytvára fyzikálne nekonzistentné kombinácie veličín. Typicky ide o situáciu, keď je nastavená vysoká hodnota výkonu, avšak prírastok energie nezodpovedá očakávanému vzťahu medzi výkonom a časovým krokom. Takéto záznamy nemusia byť extrémne v jednotlivých veličinách, ale porušujú fyzikálne vzťahy medzi nimi.

TABUĽKA II
Zhrnutie útokov

Útok	Typ zásahu	Modifikácie
Data tampering	zmena hodnôt	pridaný náhodný šum
Replay attack	zmena hodnôt v čase	nahradenie aktuálnych hodnôt historickými dátami
Time-shift	zmena časovej značky	posun časových značiek bez zmeny hodnôt
Selective dropping	odstránenie dát	systematické vynechanie časti
Data injection	pridanie dát	vloženie nových falošných záznamov

IV. NÁVRH DETEKČNÝCH MECHANIZMOV

Keďže jednotlivé scenáre útokov zasahujú do dát odlišným spôsobom, bol pre každý typ útoku navrhnutý samostatný detekčný mechanizmus. Navrhnutý prístup teda nevyužíva jeden univerzálny detektor, ale súbor špecializovaných pravidiel a kontrol, ktoré zohľadňujú charakter konkrétnej manipulácie dát. Detekčné mechanizmy sú založené najmä na analýze časových radov, kontrole fyzikálnej konzistencie meraní, overovaní časovej nadväznosti záznamov a identifikácii netypických alebo nepravdepodobných vzorov v dátach. Boli navrhnuté ako binárne rozhodovacie funkcie, ktorých výstupom je informácia o prítomnosti alebo neprítomnosti daného typu útoku.



Obr. 2. Bloková schéma navrhnutého detekčného systému

Na obr. 3 je znázornená bloková schéma navrhnutého detekčného systému, v ktorom sú vstupné dáta po predspracovaní paralelne analyzované viacerými špecializovanými detekčnými mechanizmami. Výsledky jednotlivých detektorov sú následne kombinované s cieľom určiť prítomnosť a typ útoku.

1. Detekcia útoku Data tampering

Detekcia útoku typu *data tampering* je založená na overovaní konzistencie medzi okamžitým výkonom a prírastkom energie, prípadne na analýze štatistických vlastností výkonového signálu. V prípade dostupnosti údajov o kumulatívnej energii sa porovnáva očakávaný energetický prírastok vypočítaný z výkonu a časového kroku so skutočným rozdielom energie medzi po sebe nasledujúcimi záznamami.

Narušenie dát je možné kvantifikovať pomocou zvyškovej chyby energetickej bilancie:

$$r_t = |(E_t - E_{t-1}) - P_{t-1}\,\Delta t|. \quad (1)$$

kde E_t predstavuje kumulatívnu energiu P_{t-1} výkon a Δt je časový skok. Hodnota r_t vyjadruje mieru nekonzistentnosti medzi výkonom a energiou, pričom jej zvýšené hodnoty indikujú možné narušenie integrity dát.

Ak rozdiel medzi očakávanou a skutočnou hodnotou energie prekračuje definovaný prah vo významnej časti datasetu, záznam je vyhodnotený ako podozrivý. V prípade absencie energetického stĺpca sa detekcia opiera o analýzu štatistických vlastností výkonového signálu, najmä o identifikáciu dominujúcich hodnôt, ich opakovanosti, dĺžky sekvencií a frekvencie výskytu. Cieľom je odhaliť neprirodzené správanie signálu, ktoré môže byť dôsledkom pridania šumu alebo inej formy manipulácie.

2. Detekcia útoku Replay attack

Detekcia replay útoku vychádza z predpokladu, že opakovane vložené historické dáta vytvárajú v časovom rade identické alebo veľmi podobné segmenty. Mechanizmus preto analyzuje výkonový signál v sekvenčných oknách, ktorých dĺžka je určená na základe 15 minútového vzorkovacieho intervalu. Jednotlivé segmenty časového radu je možné formálne vyjadriť ako:

$$s_i = [x_i,\ x_{i+1}, \ldots, x_{i+\omega-1}]. \quad (2)$$

kde s_i predstavuje segment signálu začínajúci v čase i a ω je dĺžka analyzovaného okna. Replay útok je indikovaný v prípade, že existujú dva segment s_i a s_j, ktoré sa vyskytujú v rôznych častiach časového radu:

$$s_i = s_j, |i - j| \geq \omega. \quad (3)$$

Prirodzene konštantné úseky, napríklad nočné minimá spotreby, sú z analýzy vyradené na základe nízkej variability signálu, aby sa zabránilo falošným detekciám. Ak sa identifikuje opakovanie dostatočne dlhého nekonštantného segmentu v inom časovom intervale, dataset je vyhodnotený ako napadnutý replay útokom.

3. Detekcia útoku Time-shift

Detekcia útoku typu *time-shift* je založená na analýze časových značiek a ich vzájomných rozdielov. Ak sú v dátach priamo dostupné príznaky časového posunu, môžu byť využité priamo. V opačnom prípade sa vyhodnocuje pravidelnosť a monotónnosť časového radu. Rozdiel medzi po sebe nasledujúcimi časovými značkami je definovaný ako:

$$d_i = t_i - t_{i-1}. \quad (4)$$

kde $t_i$predstavuje časovú značku i-teho záznamu. Referenčný časový krok je určený ako medián časových hodnôt d_{med} = *median*(d_i). Mechanizmus následne vyhodnocuje odchýlky od očakávaného časového kroku, pričom anomália je indikovaná, ak:

$$\left|\frac{d_i}{d_{med}} - round\left(\frac{d_i}{d_{med}}\right)\right| > \varepsilon. \quad (5)$$

kde ε predstavuje tolerančný parameter.

Okrem toho sa kontroluje výskyt záporných časových krokov, neprimerane veľkých časových skokov a takmer duplicitných časových značiek, ktoré môžu vznikať pri lokálnom posune alebo cyklickom mapovaní dát. Ak sa v časovej štruktúre záznamov objaví viacero takýchto anomálií, prítomnosť útoku je vyhodnotená ako pravdepodobná.

Pri experimentálnom overovaní bol tolerančný parameter nastavený na hodnotu $\varepsilon = 0{,}1$, ktorá bola stanovená empiricky na základe analýzy rozloženia časových odchýlok v trénovacích datasetoch.

4. Detekcia útoku Selective dropping

Detekčný mechanizmus pre selective dropping je zameraný na identifikáciu narušenej kontinuity časového radu. Základným príznakom útoku je výskyt medzier medzi časovými značkami, ktoré sú výrazne väčšie než očakávaný vzorkovací interval. Rozdiel medzi po sebe nasledujúcimi časovými značkami je definovaný rovnako ako v rovnici (4), kde útok je indikovaný v prípade, že sa v časovom rade vyskytujú výrazne väčšie medzery.

$$d_i > \beta d_{med}. \quad (6)$$

kde β predstavuje prahový koeficient určujúci prípustnú odchýlku od štandardného časového kroku.

Okrem samotných medzier sa kontroluje aj prítomnosť chýbajúcich, prázdnych alebo nepašovateľných časových značiek. V prípade dostupnosti meraných hodnôt sa vyhodnocuje aj podiel chýbajúcich numerických hodnôt. Ak sa v datasete vyskytne viacero neštandardných medzier alebo zvýšený podiel neplatných záznamov, dáta sú označené ako ovplyvnené selektívnym vynechávaním.

Prahový koeficient bol nastavený na hodnotu $\beta = 1{,}5$, čo zodpovedá prípustnej odchýlke 50 % od štandardného vzorkovacieho intervalu; táto hodnota bola zvolená empiricky s cieľom minimalizovať počet falošných detekcií.

5. Detekcia útoku Data injection

Detekcia data injection je založená na kombinácii kontroly duplicít a overovania fyzikálnej prípustnosti vybraných meraných veličín. Prvým príznakom útoku je výskyt duplicitných časových indexov, ktoré môžu indikovať vloženie nových záznamov s rovnakou časovou značkou. Ďalej sa kontrolujú fyzikálne limity dostupných meraní, napríklad rozsah napätia, frekvencie a účinníka. Formálne možno túto podmienku vyjadriť ako:

$$x_t \notin [x_{min}, x_{max}]. \quad (7)$$

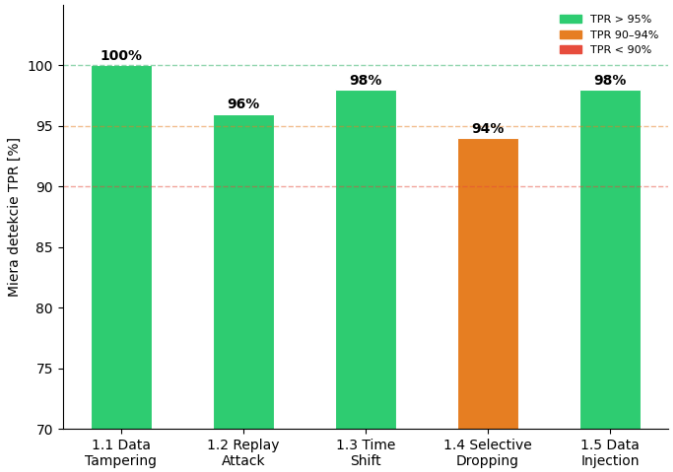
kde x_t predstavuje meranú veličinu v čase t a interval [x_{min}, x_{max}] definuje jej fyzikálne prípustný rozsah.

Ak niektorá hodnota prekročí tieto medze, záznam je vyhodnotený ako anomálny. Pri externých datasetoch bez dostatočného kontextu o význame a jednotkách meraných veličín sa detekcia zámerne obmedzuje, aby sa predišlo nespoľahlivým záverom.

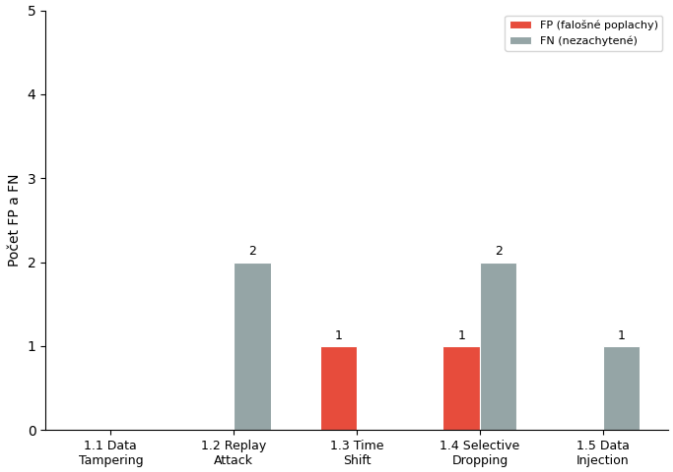
V. VÝSTUP DETEKČNÝCH MECHANIZMOV

Výsledky navrhnutých detekčných mechanizmov boli vyhodnotené na synteticky generovaných datasetoch. Celkovo bolo analyzovaných 50 datasetov reprezentujúcich rôzne profily spotreby a scenáre správania odberateľov. Na každý dataset bol aplikovaný kompletný súbor navrhnutých útočných algoritmov, čím sa zabezpečilo jednotné a konzistentné testovanie všetkých detekčných mechanizmov naprieč rôznymi typmi útokov. Vyhodnotenie bolo realizované pomocou základných metrík kvality detekcie, konkrétne miery správnej detekcie

(TPR – True Positive Rate), ako aj počtu falošne pozitívnych (FP) a falošne negatívnych (FN) detekcií.



Obr. 3. Skutočná miera detekcie podľa TPR



Obr. 4. Falošné a nezachytené poplachy

Výsledky detekcie jednotlivých typov útokov sú znázornené na Obr. 4 a Obr. 5. Prvý graf zobrazuje mieru detekcie vyjadrenú pomocou metriky TPR, zatiaľ čo druhý graf počet falošne pozitívnych (FP) a falošne negatívnych (FN) detekcií pre jednotlivé scenáre útokov,

Z grafov je možné pozorovať, že pre väčšinu typov útokov dosahuje miera detekcie hodnoty nad 95 %. Útok *data tampering* dosahuje najvyššiu hodnotu TPR na úrovni 100 %. Útoky *time-shift* a *data injection* dosahujú hodnotu približne 98 %, zatiaľ čo *replay attack* dosahuje hodnotu 96 %. Najnižšia hodnota TPR je zaznamenaná pri útoku *selective dropping* na úrovni 94 %.

Z pohľadu počtu chýb detekcie je možné vidieť, že falošne pozitívne detekcie sa vyskytujú len pri vybraných typoch útokov a ich počet je nízky. Falošne negatívne detekcie sa vyskytujú vo viacerých scenároch, pričom ich počet sa líši v závislosti od typu útoku.

Mierne nižšia úspešnosť detekcie pri útokoch selective dropping a replay attack je spôsobená ich charakterom. Selektívne vynechávanie krátkych úsekov dát môže vytvárať medzery, ktoré sa pohybujú na hranici nastaveného prahového koeficientu β, čo vedie k niektorým nedetekovaným prípadom. V prípade replay útoku môžu krátke opakované segmenty zodpovedať prirodzeným periodicitám v spotrebnom profile (napr. nočné minimá), čo sťažuje jednoznačnú detekciu bez falošných poplachov.

VI. ZÁVER

Cieľom tejto práce bolo navrhnúť a implementovať systém pre generovanie syntetických dát inteligentných meradiel a zároveň overiť možnosti detekcie kybernetických útokov na takto vytvorených dátach. Navrhnutý generátor umožňuje vytvárať časové rady spotreby elektrickej energie s nastaviteľnými parametrami, ako sú typ profilu odberateľa, periodicita, náhodnosť či úroveň šumu. Vytvorené datasety vykazujú charakteristiky blízke reálnym meraniam, najmä z hľadiska dennej variability a dlhodobej spotrebnej dynamiky.

Experimentálne overenie bolo realizované na 50 synteticky generovaných datasetoch, na ktoré boli aplikované rôzne scenáre útokov. Dosiahnuté výsledky potvrdili vysokú účinnosť navrhnutých detekčných mechanizmov, pričom väčšina útokov bola detegovaná s mierou úspešnosti nad 95 %. Najvyššia úspešnosť bola dosiahnutá pri útoku typu *data tampering*, kde detekčný mechanizmus spoľahlivo identifikoval narušenie hodnôt merania. Vysokú účinnosť dosiahli aj detektory pre útoky *time-shift* a *data injection*, ktoré boli schopné identifikovať anomálie v časových značkách a fyzikálne nekonzistentné hodnoty. Naopak, mierne nižšia úspešnosť bola zaznamenaná pri útokoch typu *replay attack* a *selective dropping*. Tieto útoky sa vyznačujú tým, že zachovávajú realistický priebeh dát alebo spôsobujú len čiastočné narušenie časového radu, čo sťažuje ich jednoznačnú identifikáciu. Napriek tomu aj v týchto prípadoch dosahovali detekčné mechanizmy vysokú mieru úspešnosti,

Napriek dosiahnutým výsledkom má navrhnutý prístup aj určité obmedzenia. Syntetické dáta, hoci sa snažia aproximovať reálne správanie odberateľov, nemusia vždy zachytiť všetky komplexné vzory a neštandardné situácie v reálnych distribučných sieťach. Detekčné mechanizmy sú navyše navrhnuté ako špecifické pre jednotlivé typy útokov, čo môže obmedziť ich schopnosť generalizácie na nové alebo kombinované formy útokov. Možným smerom ďalšieho výskumu je rozšírenie generátora o realistickejšie modely správania používateľov, napríklad na základe reálnych dát alebo pokročilých generatívnych modelov. V oblasti detekcie je možné uvažovať o využití metód strojového učenia, ktoré by umožnili automatické učenie vzorov anomálií a zlepšenie robustnosti voči neznámym typom útokov. Zaujímavým rozšírením je aj kombinácia viacerých detekčných prístupov do jednotného hybridného systému.

Navrhnutý prístup má potenciál praktického využitia najmä v oblasti testovania a validácie detekčných algoritmov v prostredí inteligentných sietí. Syntetické dáta umožňujú bezpečne simulovať rôzne scenáre útokov bez potreby zásahu do reálnych systémov, čím predstavujú vhodný nástroj pre vývoj, testovanie a porovnávanie bezpečnostných riešení v energetike.

LITERATÚRA

[1] JIN, S.: False Data Injection Attack Against Smart Power Grid Based on Incomplete Network Information. Electric Power Systems Research, 2024, vol. 230, 110294. DOI: 10.1016/j.epsr.2024.110294.

[2] REDA, H. T. – ANWAR, A. – MAHMOOD, A.: Comprehensive Survey and Taxonomies of False Data Injection Attacks in Smart Grids: Attack Models, Targets, and Impacts. Renewable and Sustainable Energy Reviews, 2022, vol. 163, 112423. DOI: 10.1016/j.rser.2022.112423.

[3] FAROOQ, A. – SHAHID, K. – OLSEN, R. L.: Securing the Green Grid: A Data Anomaly Detection Method for Mitigating Cyberattacks on Smart Meter Measurements. International Journal of Critical Infrastructure Protection, 2024, vol. 46, 100694, DOI: 10.1016/j.ijcip.2024.100694.

[4] LIU, Z. – LIU, M. – WANG, Q. – TANG, Y.: False Data Injection Attacks on Data-Driven Algorithms in Smart Grids Utilizing Distributed Power Supplies. Engineering, 2025, vol. 51, s. 62–74. DOI: 10.1016/j.eng.2024.11.025.

[5] AOUFI, S. – DERHAB, A. – GUERROUMI, M.: Survey of False Data Injection in Smart Power Grid: Attacks, Countermeasures and Challenges. Journal of Information Security and Applications, 2020, vol. 54, 102518. DOI: 10.1016/j.jisa.2020.102518.

[6] DRAYER, E. – ROUTTENBERG, T.: Detection of False Data Injection Attacks in Smart Grids Based on Graph Signal Processing. IEEE Transactions, 2019. DOI: 10.1109/JSYST.2019.2927469.

[7] ZHANG, G. – GAO, W. – LI, Y. – GUO, X. – HU, P. – ZHU, J.: Detection of False Data Injection Attacks in a Smart Grid Based on WLS and an Adaptive Interpolation Extended Kalman Filter. Energies, 2023, vol. 16, 7203. DOI: 10.3390/en16207203.

[8] PAUDEL, S.: An Evaluation of Methods for Detecting False Data Injection Attacks in the Smart Grid. Frontiers in Computer Science, 2024, vol. 6, 1504548. DOI: 10.3389/fcomp.2024.1504548.

[9] ZHU, Y. – LIU, R. – CHANG, D. – GUO, H.: Detection of False Data Injection Attacks on Power Systems Based on Measurement-Eigenvalue Residual Similarity Test. Frontiers in Energy Research, 2023, vol. 11, 1285317. DOI: 10.3389/fenrg.2023.1285317.

[10] SHAHID, M. A. – AHMAD, F. – ALBOGAMY, F. R. – HAFEEZ, G. – ULLAH, Z.: Detection and Prevention of False Data Injection Attacks in the Measurement Infrastructure of Smart Grids. Sustainability, 2022, vol. 14, 6407. DOI: 10.3390/su14116407.

ADRESY AUTOROV

Ľubomír Pallaj, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, lubomir.pallaj@tuke.sk
Jaroslav Petráš, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, jaroslav.petras@tuke.sk
Ardian Hyseni, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, ardian.hyseni@tuke.sk
František Margita, Technická Univerzita Košice, Katedra elektroenergetiky, Mäsiarska 74, Košice, SK 04210, Slovenská Republika, frantisek.margita@tuke.sk