ROBUSTNESS OF A BLACK-BOX WATERMARK IN A COMPACT 1D CNN UNDER QUANTISATION AND PRUNING: A REVIEW AND INITIAL MEASUREMENTS

STEFAN G. NIKOLOV

ABSTRACT: *Trained neural network models are valuable assets that are trivially copied, and embedded deployment places them physically outside the owner's control. This paper reviews watermarking schemes for proving model ownership, with a focus on compact models for embedded devices, and reports initial measurements on a compact 1D CNN classifier of power-line disturbances. A behavioural (black-box) watermark based on a trigger set was embedded during training; robustness was measured under weight quantisation (8 to 2 bits) and magnitude pruning (10–90%), with control models trained without a watermark. Embedding depends critically on the output head and the parameter budget; the watermark survives quantisation down to 6 bits and 10% pruning intact; and the bit depths that destroy it also destroy the model. Its fidelity cost ranges from 2.1 percentage points at 412 parameters to 18.7 at 294. No false positives were observed on unwatermarked controls.*

KEYWORDS: *neural network watermarking, black-box watermark, TinyML, quantisation, pruning, 1D CNN*

DOI: https://doi.org/10.46687/TDNG4859

УСТОЙЧИВОСТ НА BLACK-BOX ВОДЕН ЗНАК В КОМПАКТЕН 1D CNN ПРИ КВАНТУВАНЕ И ОРЯЗВАНЕ: ОБЗОР И НАЧАЛНИ ИЗМЕРВАНИЯ

СТЕФАН Г. НИКОЛОВ

АБСТРАКТ: Обученият невронен модел е ценен актив, който се копира тривиално, а внедряването във вградено устройство го поставя извън контрола на собственика. Докладът прави обзор на схемите за водни знаци за вградени устройства и представя начални измервания върху 1D CNN класификатор на смущения в мрежовото напрежение. Поведенчески (black-box) воден знак е вграждан по време на обучението; устойчивостта е измерена при квантуване на теглата (8 до 2 бита) и орязване (10–90%), с контролни модели без воден знак. Вграждането зависи от изходната глава и от броя на параметрите; знакът преживява непокътнат квантуване до 6 бита и 10% орязване; битовите дълбочини, които го разрушават, разрушават и модела. Цената му е от 2,1 процентни пункта при 412 параметъра до 18,7 при 294. Фалшиво положителни резултати при контролите не са наблюдавани.

1. Въведение

Обученият невронен модел е ценен актив. При компактен модел стойността му не е в изчислителния ресурс, защото обучението на мрежа с няколкостотин параметъра е евтино. Скъпи са две неща: данните, тоест събирането и етикетирането на реални сензорни записи, и инженерните решения, които правят модела възможен при такъв бюджет. Такива решения са определянето на класовете, изборът на прозорец и на честота на дискретизация, изборът на изходна глава (output head). Финалните тегла са резултат от поредицата от тях и от търсене сред много конфигурации.

Архитектурният клас – двуслойна едномерна конволюционна мрежа – е публично известен. Тайни са конкретната конфигурация, която работи при няколкостотин параметъра, и данните, върху които е намерена. Моделът обаче е файл с числа: копира се

безпроблемно и копието не се различава от оригинала. Криптографският подпис не решава проблема, защото доказва уникалност само на непроменено копие: при квантуване (quantisation) и орязване (pruning), които са задължителна стъпка от преобразуването на модела при системи с ограничени ресурси, подписът вече не съвпада, а моделът остава функционално еквивалентен.

При модел, внедрен във фърмуера на микроконтролер, авторът губи физическия контрол върху него: устройството е в ръцете на клиента. Фърмуерът може да бъде прочетен, а при блокиран достъп до паметта параметрите и структурата на мрежата могат да бъдат възстановени и през електромагнитен страничен канал [13].

Едновременно с това преобразуването на модела променя числено носителя, в който един воден знак би бил вграден. Защитата трябва да преживее орязването, квантуването и компилаторните преобразувания и да се внедри в устройството непокътната, преди изобщо да е срещнала противник.

Водните знаци, вграждани в невронни модели, са предлагани като решение от 2017 г. [1, 2], но изследванията са съсредоточени върху големи езикови модели и модели за изображения, проверявани в облака. Дали и как тези схеми работят за компактен едномерен конволюционен модел върху сензорни времеви редове, минал през преобразуването на модела, е въпрос без публикуван отговор.

Настоящият доклад има три цели. Първата е обзор на схемите за водни знаци и на атаките срещу тях. Фокусът е върху приложимостта им при вградени ограничения. Втората е проверка дали пресечната точка компактен 1D CNN, INT8, сензорни времеви редове и микроконтролер е покрита от публикувана схема. Третата е представяне на начални измервания върху компактен 1D CNN модел за класификация на смущения в мрежовото напрежение: кога водният знак успява да се вгради, какво оцелява след квантуване и орязване и каква е цената му в точността на класификацията.

2. Водни знаци в компактни невронни модели: състояние, ограничения и начални измервания

2.1 Терминология

Разглеждат се водни знаци, вграждани в самия модел – в стойностите на параметрите му или в поведението му – а не водни знаци в генерираното от модела съдържание [1, 2]. По-нататък в текста „знак“ означава воден знак, а „схема“ – схема за воден знак, тоест тройката процедури: генериране на ключ, вграждане на знака и проверка на присъствието му в заподозрян модел [2].

Внедряване наричаме поставянето на обучения модел в микроконтролера. Преобразуване на модела наричаме числените промени, през които моделът минава преди това: орязване, квантуване и компилаторни преобразувания (сгъване на BatchNorm, сливане на оператори).

Водният знак се различава от стеганографията по целта. Стеганографията крие самото съществуване на съобщение в носителя – изображение, видео или друг файл – и успехът ѝ се мери с неоткриваемост [14, 15]. При водния знак съобщението може да е известно и знакът продължава да изпълнява предназначението си; успехът се мери с това дали знакът оцелява промените в модела и доказва собствеността върху него.

Разграничението white-box / black-box се отнася до начина на проверката. При white-box проверяващият разполага с теглата на заподозрения модел [1]; при black-box – само с

двойки вход–изход [2]. Когато моделът работи върху чуждо вградено устройство, до теглата обикновено няма достъп, затова в този доклад се измерва black-box схема.

Всяка схема за воден знак трябва да балансира три изисквания, които са в противоречие помежду си. Вярност – моделът със знак да работи практически колкото модела без него. Устойчивост – знакът да оцелява промените по модела. Капацитет – колко информация носи знакът [1, 4].

При компактен модел вярността и капацитетът са ограничени от един и същ брой параметри: под определен брой знакът не се вгражда изобщо, близо до границата се вгражда срещу загуба на точност, а над определен размер на подписа моделът не го поема напълно. Устойчивостта зависи от тях косвено – знак, който не се е вградил напълно, няма какво да запази. Измерването на този компромис е предметът на точка 2.6.

2.2 Методология

Търсенето е проведено в IEEE Xplore, ACM Digital Library, arXiv и DBLP за периода 1999–2026 г. по комбинации от изразите „neural network watermarking“, „DNN IP protection“, „black-box watermark“, „TinyML security“ и „quantization watermark“. Включвани са работи, които предлагат или систематизират схеми за доказване на собственост върху самия модел. Изключвани са работите за водни знаци върху генерирано съдържание. Отделно е проверено дали съществува работа, покриваща едновременно четири условия: компактен 1D CNN, INT8, сензорни времеви редове и микроконтролер. Такава не е намерена.

За експеримента се използва браузърна среда за обучение на компактни невронни мрежи. Задачата е класификация на смущения в мрежовото напрежение с честота 50 Hz върху прозорец от 128 проби (40 ms при 3,2 kHz) в $K = 4$ класа. Класовете са: чист сигнал, пулсации от импулсни захранвания, хармоници (3-ти и 5-и) и импулсни смущения. Обучаващият набор съдържа 480 синтетично генерирани примера с фоново ниво на шума 0,04. Архитектурата е 1D CNN с два конволюционни слоя (ядро 5, пулинг, дилатация 1) и променлив брой канали $C \in \{2, 4, 8, 12\}$. Изходната част на мрежата (глава) е една от три: осредняване по времето (GAP), максимум по времето (GMP) или разгъване на картата във вектор (Flatten). Параметрите на обучението са: скорост 0,003, ReLU, размер на партидата 16, без L2 регуляризация, при фиксиран бюджет от точно 500 епохи за всяко обучение.

Водният знак е поведенчески (black-box), вграждан по време на обучението аналогично на Adi et al. [2]: $n = 40$ тригерни сигнала – синусоида с 2–3 интерхармоника, изместени на +18,5 Hz над хармоник, с фиксирана фаза, комбинация, каквато реална електрическа мрежа не произвежда – получават псевдослучайни етикети, изведени от таен ключ, и съставят 10% от всяка партида (варирано: 3% и 5%). Проверката подава тригерите и брои съвпаденията k (делът k/n се нарича тригерна точност); доказателствената сила е точната биномна вероятност $P(\mathrm{Bin}(n, 1/K) \geq k)$, с праг за установен знак $p < 10^{-6}$.

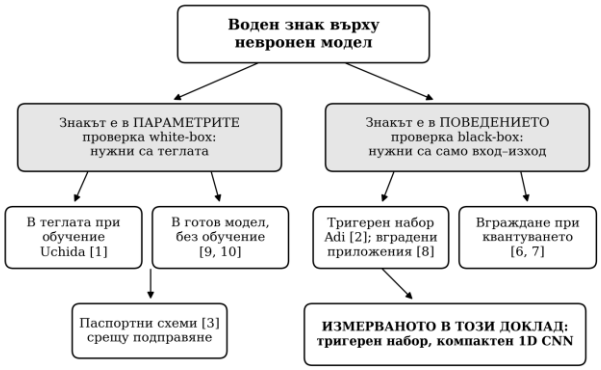
Върху всеки обучен модел се прилагат, без дообучаване (fine-tuning): симетрично квантуване по канали (per-channel) само на теглата при 8, 7, 6, 5, 4, 3 и 2 бита, и неструктурирано орязване по абсолютна стойност на 10–90% от теглата. В контролната серия са обучени две конфигурации без воден знак (GMP 8+8 и Flatten 2+2) и им е подаден същият тригерен набор (trigger set). Средата не поддържа фиксиране на началното състояние (seed), затова възпроизводими са подредбите и границите, а не точните стойности. Конфигурацията на всяко обучение е записана изцяло.

2.3 Схеми за водни знаци в литературата

Схемите за водни знаци се делят по това къде „живее“ знакът. При white-box схемите той е в параметрите: каноничният пример е схемата на Uchida et al. [1], при която теглата

на избран слой се принуждават по време на обучението да носят битове. За проверката са необходими теглата. По-късни варианти внасят знака в готов модел – чрез дообучаване или изцяло след обучението [9, 10]. Отделен клон са паспортните схеми [3], при които моделът работи пълноценно само с верен „паспорт“ – отговор на случаите, в които противник заявява собствен знак върху чужд модел.

При black-box схемите знакът е в поведението. Adi et al. [2] обучават модела да дава предварително определени отговори върху набор от специални входове – тригерен набор; собствеността се демонстрира чрез тези отговори, без достъп до теглата. Leroux et al. [8] пренасят подхода към вградени приложения. Схемата, измервана в този доклад, е от този клас. За компактни квантувани модели са предложени и схеми, при които знакът се вгражда при самото квантуване [6, 7] или се оценява върху модели от клас TinyML [9, 10]; единствената известна на автора работа върху едномерни сигнали е за ЕЕГ данни [11].



Фиг. 1. Класификация на схемите по носителя на знака и достъпа при проверката

2.4 Атаки

Целта на премахващите атаки е да заличат знака, като запазват полезността на модела [4]. Примери: дообучаване върху собствени данни, орязване на малките тегла, дестилация (distillation). Последната е обучение на нов модел, който имитира изхода на стария и не наследява теглата му. Структурните атаки не премахват знака, а променят структурата на невронната мрежа, запазвайки функцията ѝ – например чрез добавяне на фиктивни неврони, чиито изходи се компенсират взаимно, и пермутация на невроните [5]. Моделът остава функционално еквивалентен, но white-box схемите, които четат битове от определени позиции, вече не намират знака; цената за противника е нулева. Третият клас е подправянето: противникът конструира свой знак, който също се извлича убедително от същия модел, и доказателството губи стойност [3].

Систематичната оценка на Lukas et al. [4] установява, че никоя разгледана схема не издържа едновременно целия каталог от атаки. В разгледаните източници не е намерена разработка, която да оспорва това. Квантуването и орязването, измервани тук, са особен случай: те са едновременно премахваща атака и задължителна стъпка от преобразуването на модела при системи с ограничени ресурси – вредят на знака без изобщо да има противник.

2.5 Преобразуването на модела като ограничение

Преобразуването променя модела преди какъвто и да е противник, на три стъпки. Първо, орязване: премахват се параметрите с най-малка абсолютна стойност. Схема,

кодирала знака в малки тегла, го губи първа. Второ, квантуване: числата с плаваща запетая се заменят с цели по

(1) $r = S \cdot (q - Z)$,

където q е съхраняваната целочислена стойност, S е стъпката на квантуване, а Z – нулевата точка. При 8 бита диапазонът на теглата се разделя на 256 нива. Всички тегла в едно ниво стават неразличими. Колкото по-малко са битовете, толкова по-широко е нивото и толкова повече информация се губи – включително вграденият знак. Трето, компилаторни преобразувания: инструментът за преобразуване слива операции и вгражда нормализацията числено в теглата, тоест пренаписва точно носителя, за който white-box схемите са закрепени. Обзорът на Huckelberry et al. [12] разглежда целия TinyML конвейер като повърхност на атака.

По преценка на автора следствието е методологическо: за компактен модел устойчивостта трябва да се оценява и спрямо преобразуването на модела, а не само спрямо каталога от умишлени атаки. Това е предмет на 2.6.

2.6 Устойчивост на водния знак: начални измервания

Постановка. Измерени са три неща: дали знакът се вгражда изобщо, в зависимост от изходната глава и броя на параметрите; какво оцелява след квантуване и орязване; каква е цената на знака в тестова точност, чрез двойки обучения със и без знак при иначе еднаква конфигурация.

Вграждане. Вграждането зависи решаващо от изходната глава и от броя на параметрите; резултатите по конфигурации са в Таблица 1. Подредбата Flatten > GMP > GAP се запазва във всички измервания. При най-малката конфигурация (2 слоя × 4 канала) Flatten вгражда знака напълно и в трите обучения, докато GAP в две от трите обучения не се различава от случайността. С нарастване на броя на параметрите разликата между главите се свива: при 856 параметъра GMP дава пълно съвпадение, а GAP – 38 от 40 във всяко от три обучения. Долната граница е рязка: GMP с 2 × 2 канала (46 параметъра) не вгражда знака изобщо (12/40, p = 0,28).

Таблица 1

Вграждане на водния знак при 500 епохи: съвпадения от 40 тригера

Глава	Канали	Парам.	Обуч. 1	Обуч. 2	Обуч. 3	p (най-добър)
Flatten	4+4	624	40/40	40/40	40/40	$8{,}3 \cdot 10^{-25}$
GMP	4+4	128	22/40	22/40	23/40	$1{,}2 \cdot 10^{-5}$
GAP	4+4	128	12/40	14/40	21/40	$1{,}7 \cdot 10^{-4}$
GMP	8+8	412	40/40	40/40	40/40	$8{,}3 \cdot 10^{-25}$
GAP	8+8	412	33/40	25/40	30/40	$3{,}6 \cdot 10^{-14}$
GMP	12+12	856	40/40	40/40	40/40	$8{,}3 \cdot 10^{-25}$
GAP	12+12	856	38/40	38/40	38/40	$5{,}9 \cdot 10^{-21}$
GMP	2+2	46	12/40	–	–	0,28
GMP (без знак – контрола)	8+8	412	13/40	–	–	0,18

Квантуване. При всички конфигурации с пълен знак съвпадението остава 40/40 до 6 бита включително, при загуба на тестова точност под 2 процентни пункта спрямо float32. Разрушаването настъпва при 3–4 бита – но там тестовата точност пада до 25–47%, тоест до или близо до нивото на случайността. Знакът и моделът се разрушават заедно: не е намерен режим, в който квантуването премахва знака, а моделът остава използваем (Фиг. 2). По

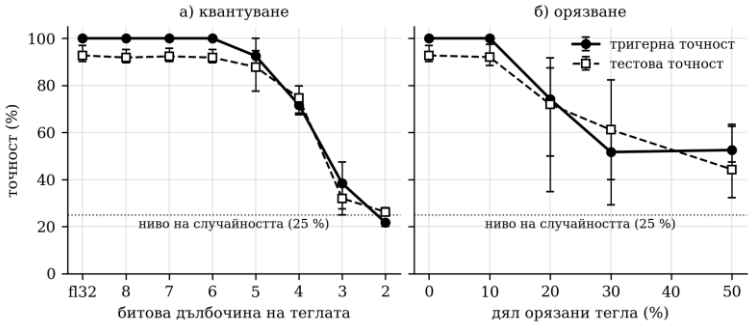
преценка на автора това означава, че за разглеждания клас модели квантуването не е практична премахваща атака, а компилаторните преобразувания и дообучаването остават откритият въпрос.

Орязване. Същата картина: при 10% орязани тегла знакът е непокътнат (40/40 при конфигурациите с пълно вграждане), при 30% падат едновременно знакът (55–75%) и точността, при 50% и нагоре моделът е негоден. Кривите на тригерната и на тестовата точност се движат заедно.

Цена на вярността. При 412 параметъра и дял 10% знакът струва 2,1 процентни пункта тестова точност (96,9% без знак срещу 94,8% със знак). Свалянето на дела до 3% прави цената неизмеримо малка, но сваля знака до ръба на прага (24/40, $p = 2{,}8 \cdot 10^{-6}$). При 294 параметъра (Flatten 2 × 2) цената скача на 18,7 пункта (90,1% срещу 71,4%) – единично наблюдение, без повторение: тригерите и задачата се конкурират за едни и същи тегла. Това е измереното съдържание на твърдението, че вграждането използва излишък, какъвто компактният модел няма. Увеличаването на подписа до n = 80 тригера надхвърля капацитета на модела: запомнят се 59/80 и полученият частичен знак се разрушава при 4 бита или 30% орязване. И тази точка е измерена еднократно.

Контроли. Контролният модел GMP 8+8 без воден знак дава 13/40 съвпадения ($p = 0{,}18$), а след квантуване на 8 бита — 14/40 ($p = 0{,}10$). Втори контролен модел (Flatten 2 × 2) също дава 13/40. Проверката не произвежда фалшиво положителни резултати върху модели, обучени върху същите данни.

Ограничения. Изводите по-горе важат само в границите на описаната постановка. Тези граници са следните. Квантува се само представянето на теглата; активациите остават float32, докато реалното INT8 преобразуване квантува и тях. Основните конфигурации в Таблица 1 са измерени с по три независими обучения; граничните точки (2 × 2 канала, вариациите на дела на тригерите и на броя им) – с по едно. Използван е един набор данни, един ключ, един тригерен механизъм. Конвейерът не включва компилаторни преобразувания (сгъване на BatchNorm, сливане на оператори) и реално устройство. Средата не позволява фиксиран seed. Дообучаването и дестилацията като премахващи атаки не са измерени.



Фиг. 2. Тригерна и тестова точност при квантуване и орязване (Flatten, 3 обучения: средно и размах)

2.7 Неизследван проблем в разгледаната литература

Обзорът установи две неща, които измерванията в 2.6 не могат да заменят. Първото е покритието: няма публикувана схема, която да покрива едновременно компактен 1D CNN, INT8, сензорни времеви редове и микроконтролер. Констатацията се опира и на търсения без резултат: заявки от вида „1D CNN watermarking“, „time series model watermarking“, „on-MCU watermark verification“ не връщат релевантни публикации към август 2026 г. Най-

близките изследвания покриват по едно от тези условия: TinyML, но изображения [9, 10]; едномерни сигнали, но без квантуване и преобразуване [11]; квантувани модели, но от мащаба на езиковите [6, 7].

Таблица 2

Представителни схеми: механизъм, показана устойчивост, известни слабости

Схема (год.)	Категория	Проверка	Показана устойчивост	Известни слабости
Uchida et al., 2017 [1]	white-box, в теглата	white-box	дообучаване, орязване	статистическо откриване; структурни атаки [5]
Adi et al., 2018 [2]	black-box, тригерен набор	black-box	дообучаване, трансфер	дообучаване с регуляризация; фалшиви претенции
Паспортни (Fan et al.), 2019 [3]	паспортна	white-box	двусмислие (по замисъл)	замяна на слоеве; поведение след INT8 – неизследвано
Li et al., 2023 [6]	в пролуката при квантуване	black-box	самото квантуване е носител	тесен клас модели (LLM)
EmMark, 2024 [7]	white-box, за квантувани модели	white-box	премахване, подправяне	само LLM; не в мащаб MCU
Leroux et al., 2024 [8]	black-box, многобитова	black-box	вградени приложения	без пълното преобразуване на модела
Adorante et al., 2025 [9]	TinyML адаптация	и двете	квантуване, орязване, шум (STM32)	само изображения; без структурни атаки
FreeMark, 2025 [10]	TinyML, post-training	white-box	квантуване, орязване	само изображения; малък капацитет
EEG схема, 2025 [11]	1D сигнали	black-box	дообучаване	без квантуване и преобразуване
Изследван тук в точка 2.6	black-box, тригерен набор	black-box	квантуване до 6 бита; орязване до 10%	една задача; без компилаторни преобразувания

Второто е разнородността: схемите в Таблица 2 са оценявани върху различни модели, набори данни, атаки с различна сила и различни определения за успех, затова публикуваните числа за устойчивост не са съпоставими. Систематичната оценка [4] и по-късните структурни атаки [5] показват, че всяко поколение схеми затваря слабостта на предишното и отваря нова; както е отбелязано в точка 2.4, не е намерена схема, устойчива едновременно на премахване, структурни трансформации и фалшиви претенции (атака чрез двусмислие).

3. Заключение

Обзорът показа, че публикуваните схеми за водни знаци не са проектирани за пресечната точка компактен 1D CNN – INT8 – сензорни времеви редове – микроконтролер, а публикуваните оценки за устойчивост не са съпоставими. Измерванията върху компактен класификатор потвърдиха, че вграждането на поведенчески воден знак зависи решаващо от изходната глава и от броя на параметрите: под определен брой параметри знакът не се вгражда изобщо, а близо до границата се вгражда за сметка на осезаем спад в точността на основната задача. Квантуването на теглата до 6 бита и орязването до 10% оставиха знака непокътнат във всички измерени конфигурации; разрушаването му настъпваше едва в режими, в които и самият модел преставаше да е използваем. Контролните модели без воден знак не дадоха фалшиво положителни резултати. Всички измервания са върху една задача, една схема на вграждане и ограничен брой повторения, при квантуване само на теглата – изброените в 2.6 ограничения важат за всеки от изводите.

Приносът на доклада е в две направления. Едното е измерването: доколкото е известно на автора, това е първото измерване на поведенчески воден знак върху компактен 1D CNN

за сензорни времеви редове, преминал през квантуване на теглата и орязване. Другото е проверката: тя е придружена от контролна серия модели без воден знак, което позволява твърдението за липса на фалшиво положителни резултати да бъде проверено, а не само заявено.

Оставащите изследователски въпроси (ИВ) са следните. ИВ1: пренася ли се поведенческият знак при дестилация (обучение на модел-ученик само върху изходите на учителя, без тригерите) и зависи ли това от набора за пренос и от вида на етикетите. ИВ2: може ли тригерният набор да се проектира така, че дестилацията да пренася знака, вместо да го премахва. ИВ3: издържа ли прагът от 6 бита при пълно INT8 квантуване, включващо и активациите. ИВ4: коя процедура за проверка е устойчива на компилаторните преобразувания (сгъване на BatchNorm, сливане на оператори), сходни със структурните атаки [5]. ИВ5: може ли знакът да се вгради в готов компактен модел чрез дообучаване и как това се сравнява с вграждането по време на обучението – по вярност, устойчивост и долна граница на параметричния бюджет. Тези въпроси са обект на следваща изследователска дейност.

ЛИТЕРАТУРА:

1. Uchida, Y., Nagai, Y., Sakazawa, S., Satoh, S., Embedding watermarks into deep neural networks. Proc. ACM Int. Conf. on Multimedia Retrieval (2017), 269-277.
2. Adi, Y., Baum, C., Cisse, M., Pinkas, B., Keshet, J., Turning your weakness into a strength: Watermarking deep neural networks by backdooring. Proc. USENIX Security Symposium (2018).
3. Fan, L., Ng, K.W., Chan, C.S., Rethinking deep neural network ownership verification: Embedding passports to defeat ambiguity attacks. Advances in Neural Information Processing Systems (2019).
4. Lukas, N., Jiang, E., Li, X., Kerschbaum, F., SoK: How robust is image classification deep neural network watermarking? Proc. IEEE Symposium on Security and Privacy (2022).
5. Yan, Y., Pan, X., Zhang, M., Yang, M., Rethinking white-box watermarks on deep learning models under neural structural obfuscation. Proc. USENIX Security Symposium (2023).
6. Li, L., Jiang, B., Wang, P., Ren, K., Yan, H., Qiu, X., Watermarking LLMs with weight quantization. Findings of EMNLP (2023).
7. Zhang, R., Koushanfar, F., EmMark: Robust watermarks for IP protection of embedded quantized large language models. Proc. ACM/IEEE Design Automation Conference (2024).
8. Leroux, S., Vanassche, S., Simoens, P., Multi-bit, black-box watermarking of deep neural networks in embedded applications. Proc. IEEE/CVF CVPR Workshops (2024).
9. Adorante, R., Carra, A., Lattuada, M., Pau, D.P., Robust watermarking of tiny neural networks by fine-tuning and post-training approaches. Symmetry, 17 (7) (2025), 1094.
10. Adorante, R., Facchinetti, T., Pau, D.P., Optimized FreeMark post-training white-box watermarking of tiny neural networks. Electronics, 14 (21) (2025), 4237.
11. Abdelaziz, A., Fathi, A., Fares, A., Protecting intellectual property of EEG-based neural networks with watermarking. arXiv:2502.05931 (2025).
12. Huckelberry, J., Zhang, Y., Sansone, A., Mickens, J., Beerel, P.A., Janapa Reddi, V., TinyML security: Exploring vulnerabilities in resource-constrained machine learning systems. arXiv:2411.07114 (2024).
13. Batina, L., Bhasin, S., Jap, D., Picek, S., CSI NN: Reverse engineering of neural network architectures through electromagnetic side channel. Proc. USENIX Security Symposium (2019), 515-532.
14. Kordov, K., Zhelezov, S., Steganography in color images with random order of pixel selection and encrypted text message embedding. PeerJ Computer Science, 7 (2021), e380.
15. Параскевов, Х., Мехмедов, Е., Стеганографска програма с графичен интерфейс за защита на данни в социални мрежи. MATTEX 2022, Сборник научни трудове, 1 (2022), 79-85.

Стефан Георгиев Николов

Енкод ЕООД

ORCID ID: 0009-0004-8897-0139

E-mail: stefangn1975@gmail.com