Journal of Computer Virology and Hacking Techniques (2026) 22:61
https://doi.org/10.1007/s11416-026-00630-3

REVIEW

A taxonomy of attacks on the exposure notification system used by contact tracing applications

Henrique Faria[1] · Paulina Rutecka[2] · Pedro Pinto[3]

Received: 13 March 2026 / Accepted: 19 May 2026

Abstract

To enhance pandemic response in COVID-19, Google and Apple introduced the Exposure Notification (EN) system - a decentralized, Bluetooth-based framework enabling contact tracing applications to evaluate and mitigate exposure risks. As the EN system is expected to evolve in upcoming versions to address future pandemic scenarios, characterizing and addressing its current attack vectors and vulnerabilities is critical to ensuring its effectiveness and robustness in future public health responses. Existing literature studies have revealed vulnerabilities, assessed their impact, and, in some cases, proposed countermeasures to improve the EN's resilience. However, the current state of the art still lacks a general taxonomy with adequate granularity that includes the most recently disclosed attacks. This paper presents a comprehensive taxonomy of attacks targeting the EN system. Building on existing categorizations, the proposed taxonomy refines and expands previous frameworks to systematically classify 47 known attacks. It adopts a granular, multilevel structure organized into three core categories – Denial of Service, False Alert Injection, and Information Disclosure – based on attack characteristics, objectives, and outcomes. This taxonomy details current EN attack vectors and procedures, highlighting areas and vulnerabilities that warrant further exploration, analysis, and revision. By systematically mapping these threats, this research aims to advance the design of resilient and secure contact tracing frameworks–a basis for strengthening future contact tracing applications.

Keywords COVID-19 · Contact tracing · Applications · Exposure notification · Taxonomy · Attacks

1 Introduction

The COVID-19 pandemic has been recognized as one of the most significant global health crises of the century [1], with a profound impact not only on physical health but also on the global economy, social interactions, and consequently, mental health [2]. During the pandemic, intensive efforts were devoted to developing strategies aimed at containing the spread of the virus. One such strategy was the use of digital contact tracing applications to assess a user's potential exposure. By detecting devices in proximity, it becomes possible to trace potential transmission chains by identifying individuals with whom an infected person had contact. Regrettably, these applications failed to achieve their intended objectives, primarily due to significant user mistrust regarding the applications, particularly concerning privacy issues [3]. Moreover, empirical analyses of official Android contact tracing apps deployed in Europe have shown that, although these apps were generally well engineered, they still exhibited weaknesses, vulnerabilities, and misconfigurations with potential security and privacy implications [4]. Consequently, ensuring the security of such systems becomes a critical factor in preparing for future health crises, where these applications could be utilized for rapid epidemic management to contain outbreaks at their very inception.

Digital contact tracing applications can employ centralized or decentralized frameworks, using various methods to log encounters between mobile devices, such as GPS location or Bluetooth identifiers [5]. The centralized approach handles the matching process of the identifiers in a central server controlled by the Health Authority [6]. The decen-

✉ Pedro Pinto
pfp@isep.ipp.pt

Henrique Faria
henriquefaria@ipvc.pt

Paulina Rutecka
paulina.rutecka@uekat.pl

[1] Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal

[2] University of Economics in Katowice, Katowice, Poland

[3] GECAD, ISEP, Polytechnic of Porto, Porto, Portugal

Published online: 13 July 2026

tralized frameworks use random identifiers and process the exposure analysis on the user's device.

The Exposure Notification (EN) is a decentralized contact tracing framework developed by Google and Apple, used by over 38 million users in Europe [7]. In Google's API documentation [8] the core features of the EN system are presented. The EN system is used by contact tracing applications and handles the exchange of anonymous identifiers using Bluetooth advertisements and is also responsible for the identifier generation and the matching process of the received identifiers with the ones marked as infected. The anonymous identifiers, named Rolling Proximity Identifiers (RPIs), rotate every 15 minutes by deriving a new identifier from a Rolling Proximity Identifier Key (RPIK). In turn, this key is derived from a key that rotates every day: the Temporary Exposure Key (TEK). The RPIs received via Bluetooth advertisement scanning are stored on the device and used to determine whether a user was in contact with an infected individual by matching infected RPIs against those stored locally. Since this process relies on Bluetooth Low Energy signal measurements, proximity estimation is inherently imperfect and can be affected by environmental conditions, device orientation, and hardware characteristics [9].

As EN is expected to evolve in upcoming versions to address new pandemic scenarios, its current attack vectors and vulnerabilities should be addressed. Since EN relies on Bluetooth advertising, it is vulnerable to known Bluetooth attacks such as jamming [10] and disabling the Bluetooth feature on a mobile device [11]. Prior work has also uncovered attacks that compromise data integrity by injecting false alerts [12–15], attacks that link a user to an anonymous identifier exchanged by the EN system [16], and attacks that may enable large-scale physical harm scenarios [17]. Authors such as Dehaye & Reardon have also analyzed deployment methods, for example, by injecting malicious code into an SDK [18]. Google's documentation [19] provides a one-level categorization of attacks on the EN system. Other studies have also proposed categorizations for these attacks [12, 16].

Although numerous attacks targeting the EN ecosystem have been identified and several classification attempts have been proposed, existing approaches exhibit important limitations. Broad reviews of contact-tracing applications, such as [20, 21], provide useful high-level threat overviews, but they are not designed as EN-specific attack taxonomies. First, prior EN-focused taxonomies are typically limited to single-level or overall categorizations [22–24], which do not capture the multi-stage nature and operational complexity of modern attacks. Second, they often focus on specific subsets of threats (e.g., privacy or integrity), rather than providing a unified view across all security dimensions [22, 25]. Third, recent attack vectors, including those exploiting SDK integrations, large-scale abuse scenarios, and advanced Bluetooth-based techniques, are not systematically incorporated into existing classifications. As a result, there is currently no unified, multi-level taxonomy that comprehensively captures the full landscape of EN attacks while preserving analytical usefulness for security evaluation, system design, and risk assessment.

To address these limitations, this paper proposes a structured and comprehensive taxonomy of attacks targeting the EN system. Unlike prior approaches, the proposed taxonomy is designed as a multi-level classification framework that systematically organizes attacks based on their operational impact and underlying mechanisms. In particular, this work first identifies and formalizes a research gap in existing EN attack classifications, highlighting the lack of multi-level, comprehensive, and up-to-date taxonomies. Building upon this observation, we introduce a three-level taxonomy of EN attacks, structured around three primary categories identified in [12, 16], namely, denial of service, false alert injection, and information disclosure, which are further refined into subcategories and concrete attack instances. The proposed taxonomy integrates and systematizes previously reported attacks with newly identified or recently discussed threats, thereby providing a unified and current view of the EN threat landscape.

Furthermore, the taxonomy establishes a structured mapping between attacks and underlying system vulnerabilities, enabling more rigorous security analysis and supporting risk assessment as well as mitigation planning. Finally, we demonstrate how the proposed taxonomy can be used by both researchers and practitioners to analyze attack surfaces, prioritize risks, and design more resilient contact tracing systems. In contrast to existing categorizations, the proposed taxonomy emphasizes analytical usability. Organizing attacks according to their outcomes and operational characteristics, enables clearer reasoning about attack impact, facilitates comparison across attack classes, and supports the identification of systemic weaknesses in EN-based systems.

This paper is structured as follows: Section 3 reviews existing taxonomies for attacks on EN and decentralized contact tracing systems. Section 4 outlines the methodology and design adopted in this study. Section 5 introduces the proposed taxonomy in detail. Section 6 critically discusses its implications. Finally, Section 7 summarizes the conclusions and suggests directions for future research.

2 Methodology

To design the proposed taxonomy, a structured review of existing literature and official documentation related to attacks on the EN system and decentralized contact tracing frameworks was performed. The overall selection and filtering process is illustrated in Fig. 1.

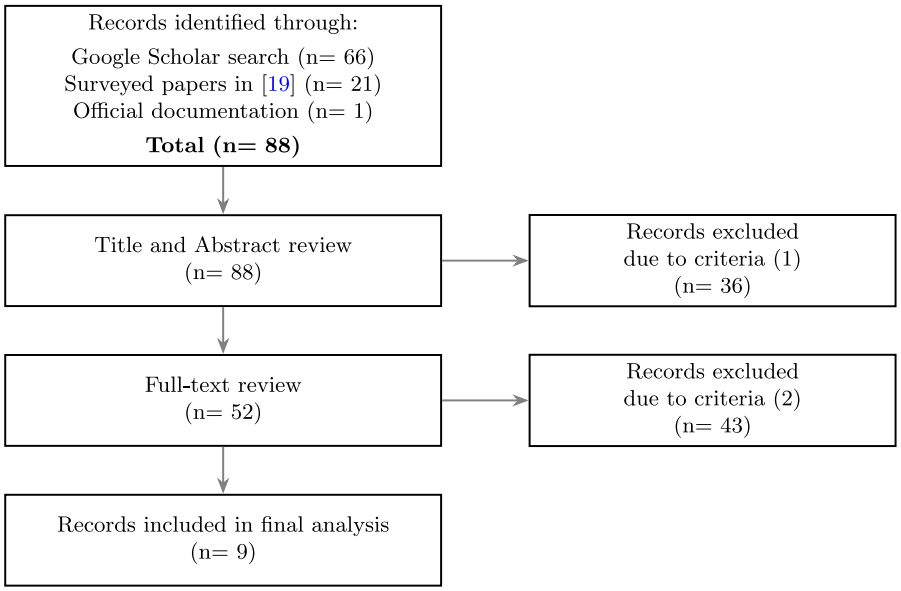

Fig. 1 Flow diagram of the literature selection process

The data collection process was based on three primary sources: Google Scholar, previous surveys/analysis of attacks reported in the literature, as summarized in [7], and publicly available official documentation provided by Google and Apple, focusing on identified risks and attack scenarios related to EN.

In Google Scholar, the following query was used: ("GAEN" OR "Google/Apple Exposure Notification" OR "Exposure Notification") AND "taxonomy" AND "attack". The time frame for initial identification included articles published from 2020, the year when Google/Apple Exposure Notification (GAEN) was launched, to the end of 2022. This time frame covers the entire COVID-19 pandemic period, which lasted from March 2020 to May 2022, when the epidemic state was lifted [26]. This resulted in 66 papers.

Regarding previous surveys and analysis, in [19], Google presents a list of security and privacy risks related to EN, together with mitigation measures incorporated into system design. This resulted in 21 reports to analyse.

Finally, the official public documentation provided by Google was also examined, which resulted in 1 report to analyse.

The initial collection process identified a total of 88 papers. After screening titles and abstracts, 36 papers were excluded based on the following criteria (1):

- Publications not written in English;
- Papers unrelated to attacks on EN or decentralized contact tracing systems.

This resulted in 52 papers selected for full-text review. At this stage, 43 papers were excluded based on the following criteria (2):

- Studies describing a single attack, without broader analysis;
- Studies that did not provide an attack characterization, classification, or taxonomy.

Finally, 9 papers were included in the qualitative synthesis. These selected works were then analyzed to extract individual attacks, identify their characteristics, and examine how they had been categorized in prior literature. This analysis served as the foundation for designing the proposed multi-level taxonomy, including the definition of categories, subcategories, and attack mappings.

3 Related work on EN-based attacks categorization

In the public document [19], Google provides a list of security and privacy risks related to EN, as well as measures adopted in the design of EN to mitigate these risks. Fig. 2 provides a graphical representation of this list. These attacks are categorized according to their primary objective: tracking, learning about social interactions, determining COVID-positive status, or disrupting the system.

The tracking category includes attacks that use features of the EN system to track a user's infection status. Variants in this category include Bluetooth-based tracking, retrieving confirmed-infected TEKs and deriving RPIs to link them to a user, and linking diagnosis keys across different days. The learning social interactions category includes attacks in which an adversary tries to infer whether a user has met an infected person by analyzing whether an exposure warning is raised. The learning COVID-positive status category

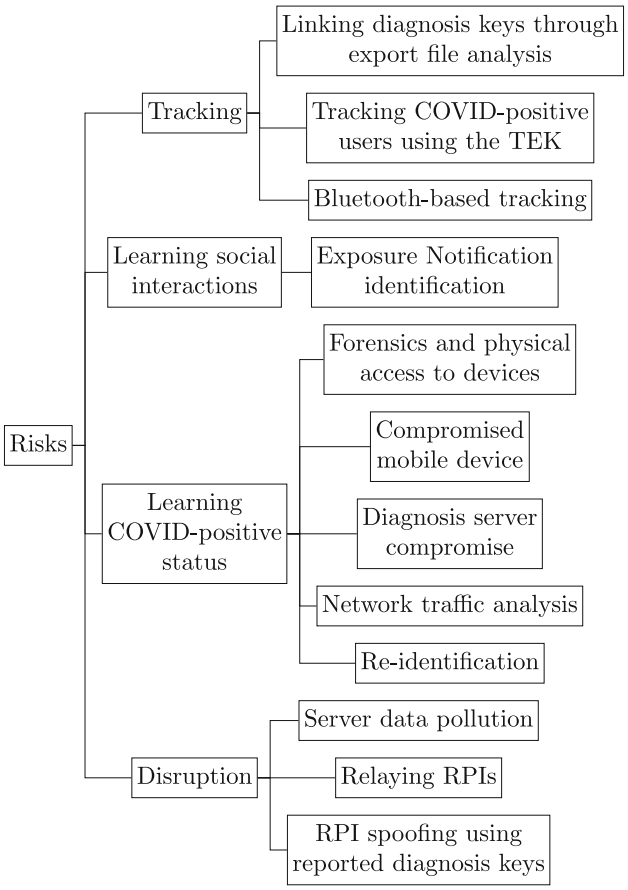

Fig. 2 Exposure Notification risks identified in [19]

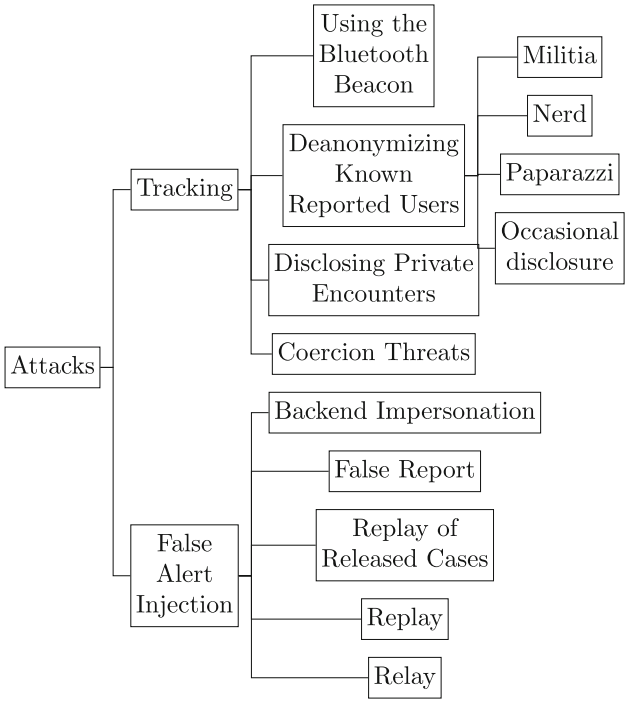

Fig. 3 Attacks and categories extracted from [12]

involves a more direct approach to ascertaining a user's infection status. By compromising the diagnosis server or the mobile device, or by performing a forensic analysis of the device, an attacker can validate the user's status. In other cases (e.g., network traffic analysis and re-identification), the attacker analyzes HTTPS communications between the user and servers or compares collected RPIs with those derived from infected keys. The disruption category involves actively injecting false RPIs and compromising the integrity of data stored on devices. These RPIs can be spoofed from a reported diagnosis key and relayed to another location (as in the relay attack in [16]). An attacker can also generate exposure warnings by falsely uploading specific TEKs , for example, via server data pollution, which is similar to the false report attack in [12].

In [12], the author introduces and categorizes attacks and vulnerabilities related to the EN system. Fig. 3 provides a graphical representation of this categorization. The author groups attacks into two main categories: False Alert Injection and Tracking. Within Tracking, there is a subcategory for Deanonymizing Known Reported Users. The Tracking attack involves obtaining information about a user and their infection status. These attacks range from using Bluetooth beacons to capture RPIs to coercion and physical threats. The subset of deanonymization attacks uses techniques to capture RPIs from a specific user and then compare them with those released as infected. False alert injection attacks involve sending false or stolen RPIs and TEKs to other users to trigger an exposure alert. This can be done via replay and relay attacks, or by falsely reporting a TEK as infected. The attacks are further detailed in [7]. This categorization provides a one-level taxonomy based on false alert injection and tracking. Given the characteristics of the analyzed attacks, this categorization could benefit from greater granularity.

In [27], the authors present a privacy threat model that includes multiple attacks on the EN system. The model associates specific goals with each identified attack; Fig. 4 provides a graphical representation of these attacks and the attacker's goals. Eavesdropping involves capturing the RPIs sent by EN. Data disclosure, coercion, and spoofing/-tracking/replay attacks have been explained previously. The attacks were analyzed in [7]. This taxonomy differs from the others discussed in this state-of-the-art review because it focuses on attack goals. It presents multiple scenarios and variants for each attack and uses a single level of categorization.

In [28], the authors present an alternative decentralized contact tracing system that they consider better prepared for mass-surveillance attacks. They describe multiple attacks (including those in [12]) and categorize them as Mass Surveillance or Other. The attacks categorized as Other are

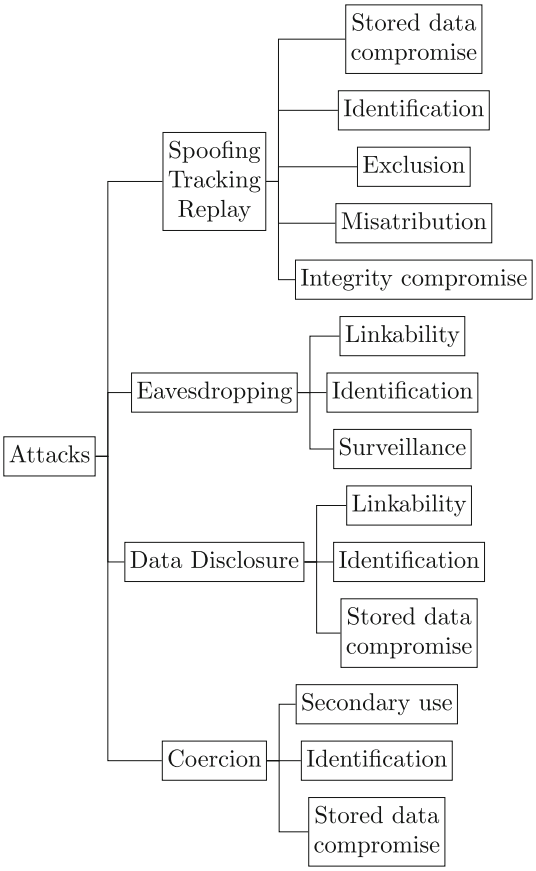

Fig. 4 Attacks and categories extracted from [27]

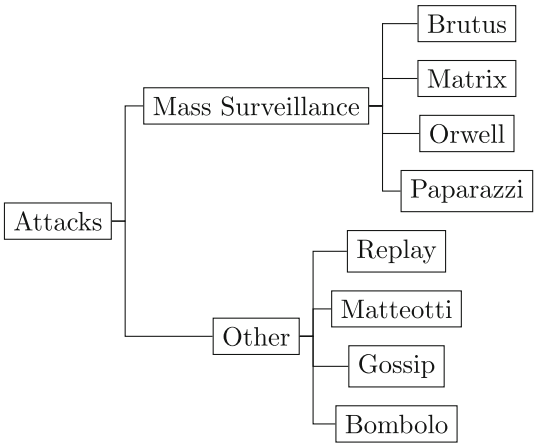

Fig. 5 Attacks and categories extracted from [28]

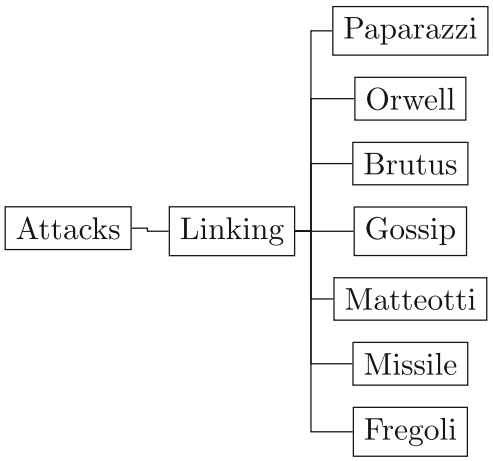

Fig. 6 Attacks identified in [22]

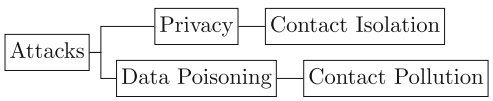

Fig. 7 Attacks and categories extracted from [23]

either variants of the replay attack in [12] or examples of singling-out attacks [7], such as the Gossip attack. Mass-surveillance attacks correspond to compromised scenarios, such as compromising the backend, the EN app, or the health authority. They focus on checking users' infection status or tracking their contacts. Further details on these attacks are provided in [7]. Fig. 5 provides a graphical representation of this classification. This categorization focuses specifically on mass-surveillance attacks and is therefore not a general classification. Additional attacks and classification levels could be included.

In [22], the authors present an alternative to the EN contact tracing system and compare vulnerabilities in the exchange of ephemeral identifiers in decentralized systems with those in their proposal. The paper discusses linkability and low-cost attacks, and although it does not define an explicit category, its attacks can be viewed as linking-related. This categorization is shown in Fig. 6. This classification focuses only on attacks targeting the identifier exchange. Since it does not discuss other attacks identified at the time (e.g., [12]), it is not considered a general taxonomy for EN attacks. All mentioned attacks are related to linkability, and each explores different parts of the EN system, including scenarios such as collusion between attackers and the Health Authority.

In [23], the authors present two attacks, Contact Pollution and Contact Isolation, and classify them as Data Poisoning and Privacy attacks, respectively. Fig. 7 illustrates this classification. This classification covers only the two attacks discovered by the authors. The categories distinguish attacks based on their impact on data integrity and user privacy. These attacks are similar to the Replay and Singling-out attacks; further details are provided in [7].

In [24], the authors analyze two attacks using the categories of Security and Privacy, and also test these attacks in real scenarios. Fig. 8 depicts the classifications provided for these attacks. These classifications are limited to the attacks tested or discovered by the authors. The Relay Worm-

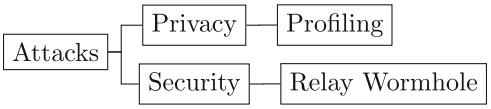

Fig. 8 Attacks and categories extracted from [24]

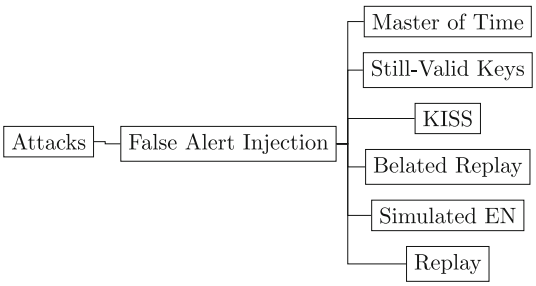

Fig. 9 Attacks and categories extracted from [29]

hole attack is similar to the Relay and Replay attacks, while the Profiling attack is similar to the Deanonymizing Known Reported Users category described in [12]. It should be noted that this work does not present a comprehensive taxonomy of all EN system attacks.

In [29], the authors explore multiple attacks that inject identifiers into a target to trigger false exposure events. As in [12], this category is referred to as False Alert Injection. Similar to previous papers, this work does not provide a general taxonomy for EN attacks. However, it provides details on different False Alert Injection attacks and how they are executed. In this paper, time-travelling attacks are introduced; in these attacks, an adversary manipulates a device's time so that EN accepts an outdated key. These attacks include Master of Time and Belated Replay. Another approach for injecting false alerts involves using an app that simulates EN's behaviour or using infected TEKs from other countries that are still valid (9).

Fig. 10 shows the taxonomy presented in [30] for general attacks on contact tracing. The authors divide attacks into three main categories, based on the technology they use: Bluetooth, GPS, or generic (applicable to both). Bluetooth-based attacks such as bluesnarfing and bluebugging involve establishing a connection with the target, which is not possible when using Bluetooth advertising. Bluejacking allows an attacker to send unsolicited messages to the target. Most generic attacks have been explained in previous classifications and in [7], except for Resource Drain and Screen Lock. The former involves sending messages to waste resources (e.g., battery), while the latter is a ransomware attack that locks the target's screen upon installation of a fake contact tracing app.

Krehling and Essex [31] propose a security- and privacy-scoring framework for contact tracing applications. Rather than presenting a comprehensive EN-specific taxonomy,

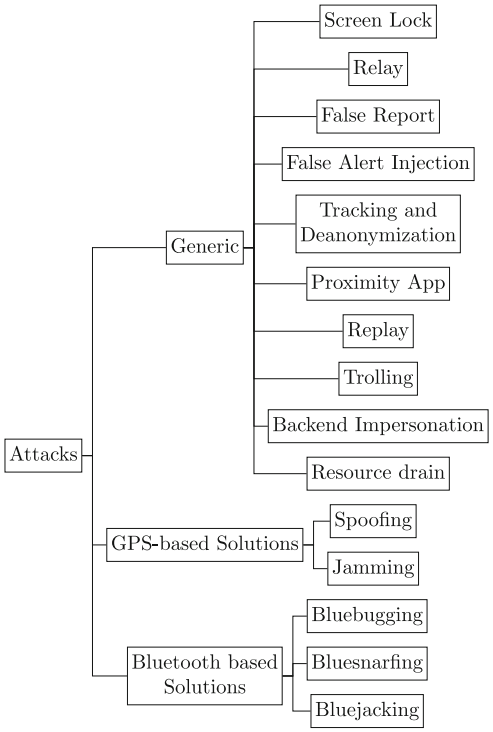

Fig. 10 Contact tracing attacks taxonomy from [30]

the paper uses a CVSS-inspired evaluation rubric and an attack tree to compare protocol families and representative applications. Fig. 11 summarizes the attack structure used in that work. Its main distinction from the taxonomies discussed above is that it organizes attacks according to high-level attacker goals—breaking privacy or injecting false information—in support of comparative scoring, rather than constructing a granular, multi-level classification of EN attacks.

As a final remark, to the best of our knowledge, there is no comprehensive and up-to-date taxonomy for EN-based attacks. Google's categorization focuses on specific attacks that are common to decentralized contact tracing systems and on whether EN is vulnerable to them. It does not cover the latest known attacks. Likewise, the taxonomy in [30] is not specific to the EN system; instead, its objective is to distinguish attacks that target GPS, Bluetooth, or both. More broadly, the reviews by Ang and Shar [20] and Hamza et al. [21] are relevant as high-level contact-tracing security references: they synthesize general threat scenarios, security and privacy concerns, and design considerations across contact-tracing applications. However, they do not propose a EN-specific, multi-level taxonomy of attacks. Similarly, Krehling and Essex [31] contribute a useful attack-modeling and scoring framework for contact tracing applications, but their emphasis is on comparative evaluation across app architectures rather than on a comprehensive taxonomy of EN attacks. The taxonomies analyzed in the papers mentioned

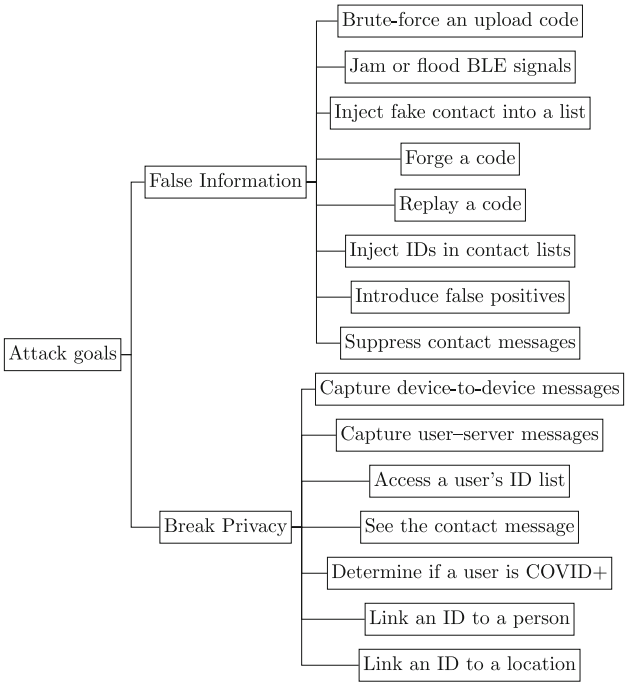

Fig. 11 Attack-tree structure summarized from [31]

in the systematic review in [7] are not general. They could be extended with more granularity, as they use only one classification level and do not include all known attacks. The reviewed categorizations typically focus on a subset of attacks discovered or tested by the authors. Some classifications, such as the one in [12], have been adopted by other authors in more recent papers.

4 Design of the proposed taxonomy

The proposed taxonomy is based on the outcome of the attack, since it directly reflects the operational impact of each attack on the EN system. Thus, we align the taxonomy with the core security objectives of the EN system: availability, integrity, and confidentiality, by defining three main categories: denial of service, false alert injection, and information disclosure. This approach is based on the initial categorization proposed in [12, 13, 16] and allows us to systematically analyze how each attack affects the system's functionality and user trust, as follows:

- Denial of Service (DoS) attacks disrupt the availability of the EN system by exhausting device resources (e.g., battery, Bluetooth interface)
- False Alert Injection attacks compromise the integrity of exposure notifications, leading to false positives or negatives
- Information Disclosure attacks violate confidentiality, exposing sensitive user data or infection status

However, this proposal expands these categorizations by introducing additional levels and greater granularity than the single-level classifications currently used. Accordingly, the first level of classification consists of these three known attack classes, from which more granular classifications are derived. The second level of the taxonomy tree includes two subcategories for each main category. This level focuses on the outcome or target of each attack and classifies attacks by their likely outcome. The definition of each subcategory takes into account an attacker's objective, how the attack is executed, and its outcome. Thus, the attacks are assigned to specific subcategories based on:

- Mechanism of Execution: how the attack is carried out (e.g., passive vs. active disruption, direct vs. inferred information disclosure)
- Outcome and Impact: the intended outcome and the impact.

The proposed taxonomy is presented as a tree structure, where each attack is referenced using unique cardinal numbers and classified under a single primary category to maintain clarity and avoid overlap. Recently, researchers in [29] uncovered a type of attack that, although it belongs to the false alert injection class, does not have the same outcome as most other attacks in that class. This taxonomy, therefore, differentiates attacks by outcome rather than grouping them under broad categories. However, we acknowledge that some attacks may be followed by other attacks. To address this, we classify attacks based on their primary objective (e.g., disrupting availability, injecting false data, or disclosing information).

The proposed taxonomy is designed to be extensible and incorporate novel attacks, allowing them to be classified as they emerge, even if they exhibit hybrid characteristics.

4.1 Denial of service category

A Denial of Service (DoS) attack is an attempt to thwart the legitimate use of a service or connection [32]. In the Bluetooth context, such attacks can disrupt communication between two devices by exhausting or otherwise affecting the resources on one of them. These resources range from the capacity of the Bluetooth chipset to the device's battery and storage. The attacks in this category are not limited to jamming and classic DoS attacks that involve high volumes of

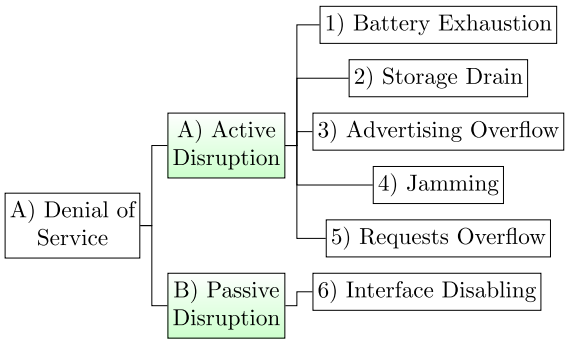

Fig. 12 Denial of Service attacks category

requests; they also include attacks launched via applications that disrupt the expected behavior of a mobile app.

The EN attacks included in the Denial of Service category of the proposed taxonomy are presented in Fig. 12.

Two subcategories are proposed for these types of attacks: passive disruption and active disruption. A passive disruption attack thwarts the EN system by disabling Operating System (OS) features that are essential to its operation (e.g., Bluetooth and GPS on an Android device). By disabling Bluetooth, either manually or programmatically, an attacker does not directly exploit EN features; instead, they disable a feature of the core system through an external measure. An active disruption attack affects the EN system by exploiting the EN features, for example by flooding an area with Bluetooth messages that carry EN Universally Unique Identifier (UUID) or draining the battery of a device by repeatedly transmitting fake EN Bluetooth messages. Such attacks target device resources, such as the battery, storage, or Bluetooth interface. Accordingly, they are divided into Internal Resources Exhaustion and External Resources Exhaustion. The former encompasses attacks that target resources not directly used in the EN communication process (e.g., battery and storage), whereas the latter targets the Bluetooth interface. The attacks categorized as external resource exhaustion are jamming, advertising overflow, and requests overflow, because they target the device's Bluetooth interface. Battery exhaustion and storage drain are classified as internal resource exhaustion. Finally, interface disabling is a passive disruption technique that involves disabling the device's Bluetooth feature.

4.2 False alert injection category

False alert injection attacks raise infection alerts on unsuspecting targets who were not exposed to an infected person. As noted in [12], a false infection contact may lead to discrimination or social accountability and may also be used to affect a target (e.g., a football player) before an important match. To trigger a false alert, an attacker must inject a specific RPI into the target's device; the means of delivery and RPI generation vary across attacks. This category of attacks is also described as fake contact events [12].

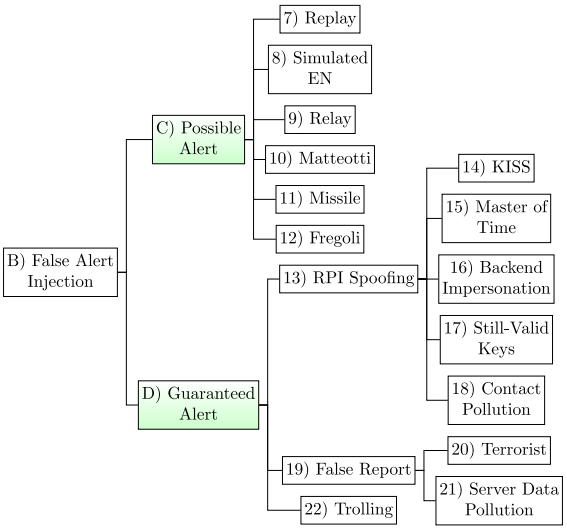

Fig. 13 False alert injection attacks

This type of attack can result in either a guaranteed alert or a possible alert. These subcategories distinguish attacks by their likelihood of raising an exposure alert on the target. Guaranteed alert attacks will always raise an alert on the target's device, whereas possible alert attacks are not guaranteed to reach the target effectively.

Replay, relay, and simulated EN attacks use randomly obtained RPIs that may or may not later be identified as infected, so they fall under the possible alert subcategory. Missile, Fregoli, and Matteotti attacks operate by spreading a randomly obtained RPI, and there is no assurance that the target receives the identifier. In contrast, RPI spoofing, false report, and trolling attacks use RPIs that are confirmed as infected. RPI spoofing attacks also include multiple variants that abuse the validity window of RPIs , for example, by deriving identifiers from released TEKs or by changing the time on the target device. Fig. 13 shows the false alert injection section of the proposed taxonomy.

4.3 Information disclosure category

EN is designed under the assumption that infected users remain anonymous and that there are no effective ways to trace RPIs to a user or device. Information disclosure attacks compromise the anonymity of the EN system to reveal whether an infected RPI originates from a particular user or whether two users have been in proximity. Information disclosure attacks are divided by the reliability and method used to obtain private data. Inferred attacks use algorithms that analyze distance, GPS positioning, or EN copycat apps (i.e., apps that mimic EN) to infer information. In contrast,

direct attacks obtain information directly from the target and do not require additional computation. Inferred attacks rely on either creating an app similar to EN to collect more information or using profiling algorithms to determine whether two RPIs may belong to the same user. For example, private encounter disclosure relies on checking whether two users receive an exposure warning within a similar time interval. Other advanced methods involve compromising the diagnosis server, tracking released TEKs , or analyzing network traffic between the target device and the server. Direct attacks are less versatile because they typically target a single individual and cannot be reused across multiple encounters. These attacks obtain information directly from the target's device, for example by capturing a single RPI or manually checking infection status in the EN-based app. Examples range from coercing the target to reveal their contact tracing app status to performing a forensic analysis of the mobile device or recording individual RPIs . Fig. 14 shows the information disclosure section of the proposed taxonomy.

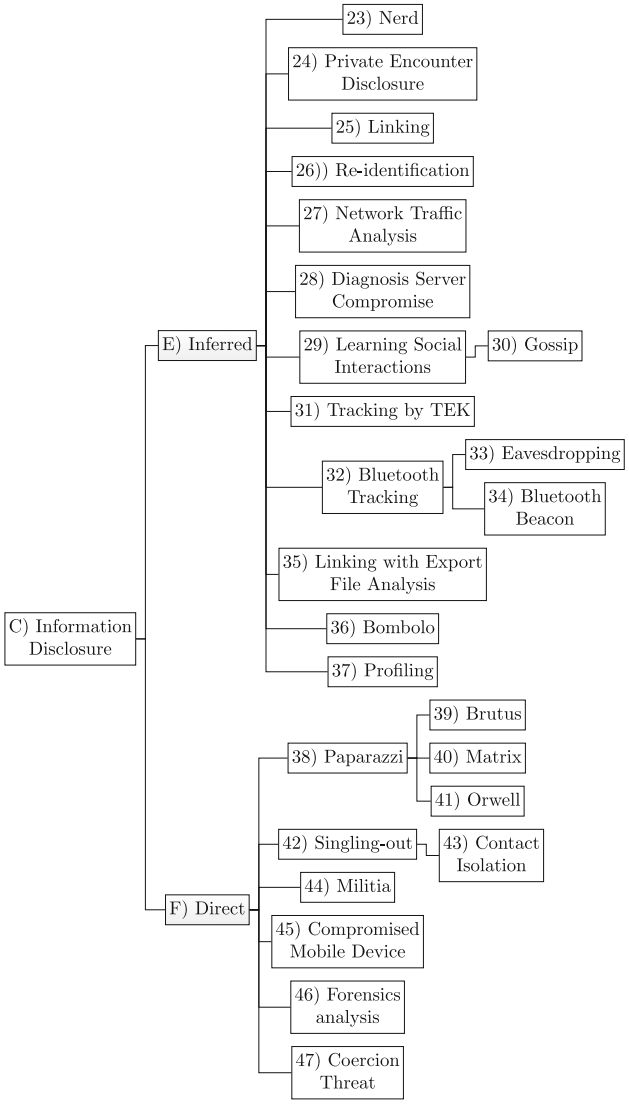

Fig. 14 Information disclosure attacks

5 Proposed taxonomy

Fig. 15 presents the proposed taxonomy, including all discussed sections and the classified attacks.

Each attack in the proposed taxonomy can be described as follows:

1. Battery Exhaustion - This attack occurs when an attacker sends large quantities of RPIs to a target to exhaust its battery.
2. Storage Drain - An attacker sends valid RPIs to the target, consuming storage resources on the device [33, 34].
3. Advertising Overflow - A malicious app sends many RPIs in a short period of time, blocking EN advertisements. This attack can be carried out by giving the malicious app's listener a higher priority for device events such as reboot or Bluetooth communication. It differs from other attacks because it directly affects EN advertisements rather than only depleting internal device resources [7].
4. Jamming - This attack targets Bluetooth data transmission between two devices by using a device that continuously transmits thousands of invalid RPIs [31].
5. Requests Overflow - A DoS attack in which an attacker sends a large number of RPIs to a target device, overloading its Bluetooth capabilities and rendering it unable to operate normally.
6. Interface Disabling - In this attack, interfaces necessary for EN to operate are intentionally disabled [11].
7. Replay - This attack consists of retransmitting RPIs that were previously received by a device [14, 31]. In [35], the authors propose improving EN by incorporating geospatial awareness when handling RPIs .
8. Simulated EN - A signal strength manipulation attack in which a malicious app simulates the EN system but manipulates signal strength during advertising to appear closer to the target than it actually is. This causes the target to receive an RPI and the EN system to falsely register the attacker as being nearby [29].
9. Relay - This attack is similar to replay attacks: an attacker receives an RPI in one location and relays it to another device in a different location.
10. Matteotti - The attacker impersonates the backend server and sends a specific TEK to the target as an infected key. This can lead the target to receive false-positive exposure notifications and may cause them to spread the virus to others.

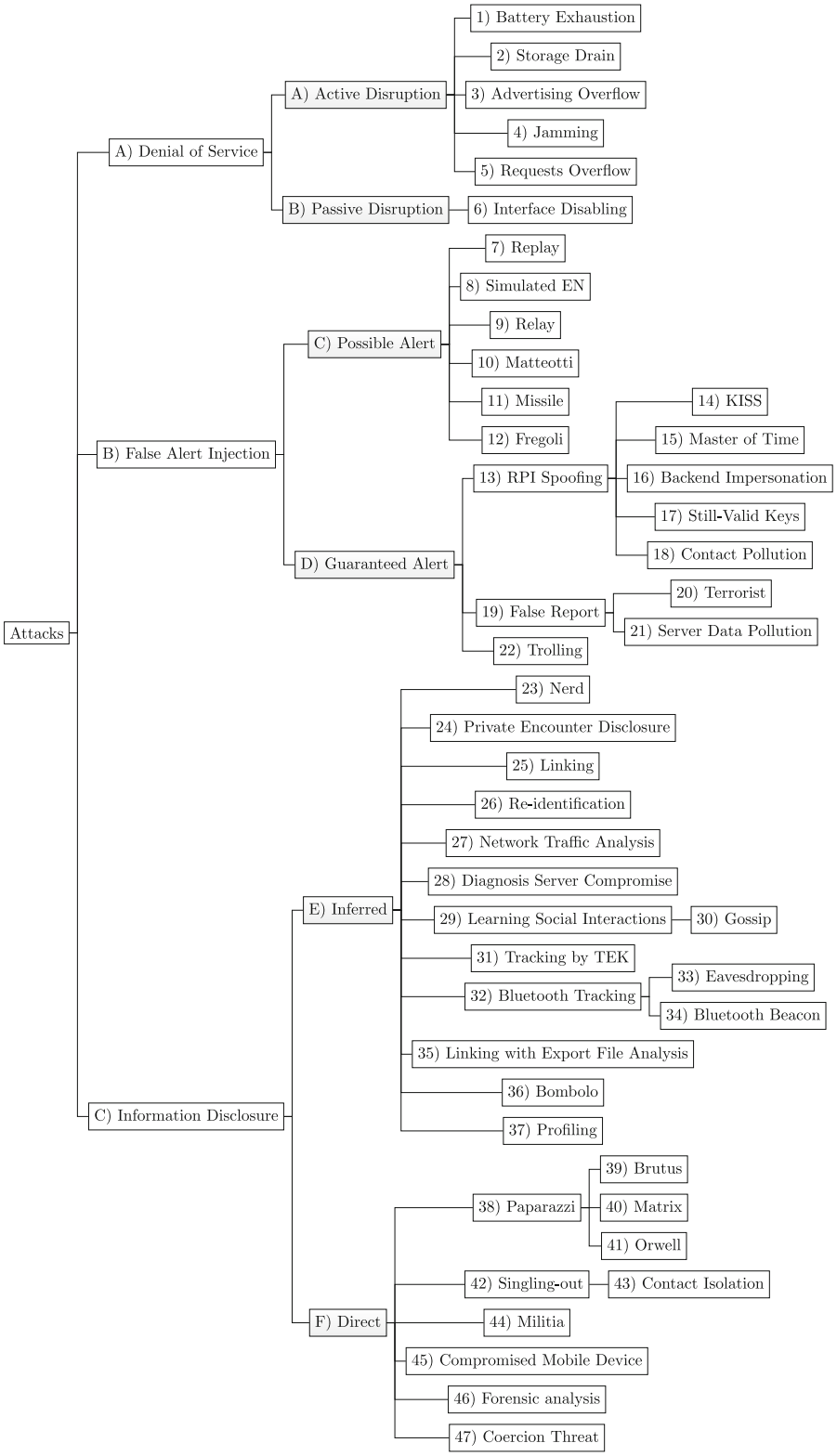

Fig. 15 Proposed taxonomy for EN attacks

11. Missile - Similar to the Relay attack, the attacker uses a long-range Bluetooth transmitter to broadcast RPIs over longer distances.
12. Fregoli - This attack is described in [22] and it is similar to the Replay attack.
13. RPI Spoofing - Involves compromising the integrity of data stored on devices by actively injecting false RPIs . This is done either by spoofing a diagnosis key or by relaying it to another location. The goal is to disrupt the normal functioning of the exposure notification system.
14. KISS - In this attack, an attacker obtains TEKs published by the health authority and replays the derived RPIs [29]. This attack is also referred to as the Belated Replay attack.
15. Master of Time - The attacker alters the target device's time (setting it earlier) and injects RPIs that were valid for that earlier time period [29].
16. Backend Impersonation - The attacker poses as the back-end server to send infected TEK to the target [12].
17. Still-Valid Keys - Published TEKs are outdated, but some remain valid due to the 2-hour RPI life tolerance. Interoperability between countries allows attackers to use valid keys from one country in another.
18. Contact Pollution - Similar to the Replay attack, the attacker sends fake RPIs to a target's device.
19. False Report - This attack is performed when an attacker falsely reports their own keys as infected, generating alerts for other users [12, 31].
20. Terrorist - This attack is an improvement on the False Report attack. In [17], the authors describe a blockchain black-market solution that allows infected keys to be sold and bought safely.
21. Server Data Pollution - Similar to the False Report attack, the attacker reports their own keys as infected to trigger exposure warnings.
22. Trolling - This attack is performed when an adversary knows they are infected and attaches their device to a carrier to spread their RPIs among unsuspecting users [16]. When their TEKs are published, multiple exposure alerts will be triggered.
23. Nerd - This attack is performed when an attacker's app collects specific information (such as GPS location, timestamps, or Wi-Fi networks) for each encounter simultaneously with EN [12]. All the collected data can be inserted into a database and then used to identify reported cases.
24. Private Encounter Disclosure - Encounters between two users can be disclosed by an attacker if the attacker can successfully raise an alert on one of their devices [12].
25. Linking - This attack is performed when an attacker associates two transmitted messages with the same device [13]. In [36], the authors evaluated the feasibility and scalability of linking attacks on EN. Through a field experiment and simulation, they found that linking attacks are highly feasible and scalable, with a probability of 86%.
26. Re-identification - In this attack, the attacker collects RPIs and links them to diagnosis keys published by the health authority.
27. Network Traffic Analysis - This attack is performed when the attacker infers the target's infection status by analyzing network traffic and communications with the health authority's server.
28. Diagnosis Server Compromise - This attack assumes that the attacker can de-anonymize users with access to the diagnosis server.
29. Learning Social Interactions - In this attack, the attacker infers whether a user has met an infected person by analyzing exposure warnings.
30. Gossip - This attack assumes that the attacker records one RPI and checks whether a user is infected when the Health Authority publishes new TEKs .
31. Tracking by TEK - The attacker links anonymous identifiers to the same device by misaligning Bluetooth MAC and RPIs rotation.
32. Eavesdropping - In this attack, the attacker captures RPIs to build user movement profiles.
33. Bluetooth Beacon - The attacker collects information about active Bluetooth devices using a Bluetooth beacon.
34. Bluetooth Tracking - This attack occurs when a user is at risk of being tracked via any Bluetooth interface.
35. Linking with Export File Analysis - In this attack, an adversary examines the Diagnosis Key batches delivered by the diagnosis server and tries to link Diagnosis Keys that were independently generated on different days and are a priori unlinkable.
36. Bombolo - In this attack, the attacker analyzes personal contacts between users using data uploaded to the Health Authority.
37. Profiling - In this attack, users' movements are profiled by capturing their RPIs [24, 27, 37].
38. Paparazzi - In this attack, an attacker uses a powerful antenna to capture RPIs from a specific user and later checks them for positive matches [12].
39. Brutus - In this attack, the attacker maps the real identity of the infected user to uploaded data using an authentication mechanism.
40. Matrix - This attack is similar to the Paparazzi attack, with a minor difference in the attacker's capabilities.
41. Orwell - Like the Matrix attack, this attack is similar to the Paparazzi attack.
42. Singling-out - This attack is described in [16] and is performed when an attacker records only one RPI. The attacker can then check whether a given user is infected when the health authority publishes new TEKs [31].

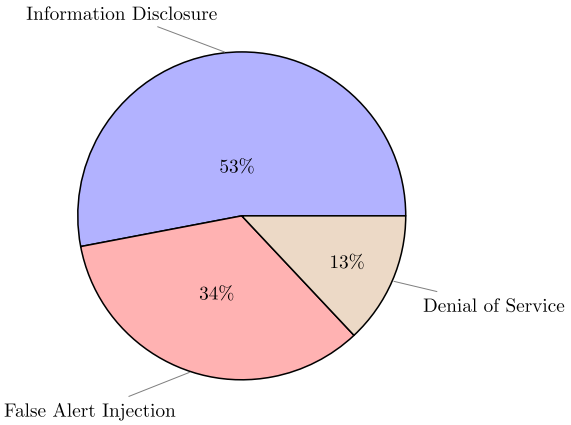

Fig. 16 Distribution of attacks by the main category

43. Contact Isolation - In this attack, the attacker links an RPI to a device by recording only one RPI and checking for positive matches when new TEKs are published [31].
44. Militia - In this attack, a group of people uses data obtained from attacks such as Nerd to threaten and isolate infected individuals.
45. Compromised Mobile Device - In this attack, an attacker verifies the user's infection status by accessing EN data stored on the device, potentially through a malicious s Software Development Kit (SDK).
46. Forensics analysis - This attack occurs when an attacker gains information about a target's infection status by accessing EN information stored on the device.
47. Coercion Threat - In this attack, the attacker threatens the target to reveal their infection status by accessing EN data stored on the device.

6 Analysis and discussion

The proposed taxonomy provides a structured and comprehensive view of the attack surface of EN systems, enabling a deeper understanding of how vulnerabilities manifest across different operational layers. The 47 attacks are categorized into three primary categories–Denial of Service, False Alert Injection, and Information Disclosure–and refined into subcategories based on attack outcomes, extending prior single-level classifications.

Fig. 16 presents the number of attacks per primary categorization. From the results, Information Disclosure accounts for the majority of identified attacks (53%), followed by False Alert Injection (34%) and Denial of Service (13%). This distribution suggests that privacy and inference-based attacks constitute the dominant threat vector, indicating that adversaries are more likely to exploit information leakage than to disrupt system availability or inject false data.

Beyond individual attack categories, the taxonomy highlights that attacks can be composed or chained across categories. For instance, inferred Information Disclosure attacks (e.g., profiling or Bluetooth tracking) may provide auxiliary data that enhance the effectiveness of direct attacks such as coercion or deanonymization. Similarly, False Alert Injection attacks can be combined with social engineering techniques to amplify their real-world impact. This reinforces the need to consider multi-stage attack scenarios, rather than isolated threats, when evaluating system security.

The taxonomy also exposes weaknesses in core EN design assumptions, particularly regarding anonymity and unlinkability. Although EN relies on rotating identifiers and decentralized processing, several attacks demonstrate that these protections can be undermined through correlation techniques, auxiliary data sources, or large-scale data collection. Consequently, the privacy guarantees offered by EN systems should be regarded as context-dependent rather than absolute.

From a defensive perspective, the implications are summarized in Table 1. This table provides a structured mapping between each attack category, its operational impact, and corresponding mitigation strategies, enabling a translation from threat analysis to system design decisions.

For Denial of Service attacks, the table emphasizes their impact on system availability and usability, primarily through resource exhaustion (e.g., battery, storage, and Bluetooth interface). Although these attacks represent a smaller portion of the taxonomy, their effect is immediate and visible to users, potentially degrading trust and reducing adoption. The recommended mitigations–such as rate limiting, adaptive Bluetooth duty cycling, and prioritization of EN-related processes–reflect the need to manage resource contention at the system level.

In contrast, False Alert Injection attacks primarily target data integrity and trustworthiness. Their implications extend beyond technical disruption to broader social effects, including misinformation, unnecessary isolation, and erosion of confidence in public health systems. The proposed countermeasures–such as multi-factor verification of diagnosis reports, geospatial plausibility checks, and cryptographic binding between identifiers–indicate that defending against these attacks requires stronger validation and authentication mechanisms across the reporting pipeline.

Information Disclosure attacks, which constitute the largest category, directly threaten user privacy and anonymity guarantees. They have the potential to lead to regulatory non-compliance and secondary misuse of sensitive data. The recommended mitigations–such as differential privacy, minimal data retention, and end-to-end encryption–highlight the importance of privacy-by-design principles and stricter control over both data exposure and metadata leakage.

Despite its contributions, this work has some limitations:

Table 1 Attack categories, impact, and mitigation strategies

Attack Category	**Operational Impact**	**Mitigation Strategies**
Denial of Service (DoS)	• Exhausts device resources (battery, storage, Bluetooth)	• Implement rate limiting for RPI transmissions
	• Disrupts EN functionality and user experience	• Use adaptive duty cycling for Bluetooth
	• Reduces public health efficacy	• Assign priority queues for EN-related events
False Alert Injection	• Undermines trust in EN systems	• Require multi-factor verification
	• Compromises public health response	• Use geospatial plausibility checks
	• Exploits cryptographic gaps	• Deploy cryptographic binding of RPIs to TEKs
Information Disclosure	• Violates user privacy	• Adopt differential privacy for RPI/TEK
	• Risks legal non-compliance	
(e.g., General Data Protection Regulation (GDPR))	• Enforce minimal data retention policies	
	• Enables secondary exploitation	• Use end-to-end encryption

- The taxonomy is derived from the publicly available literature and may not capture emerging or undisclosed threats
- Although the classification provides a qualitative understanding of the threat landscape, it does not quantify the likelihood or real-world impact of each attack
- the proposed taxonomy is specific to EN-based systems and may require adaptation for alternative or hybrid contact tracing architectures
- Continuously updating the taxonomy to reflect evolving threats will be essential to maintaining its relevance in future public health scenarios

The proposed taxonomy can serve as a foundation for future research in several directions. These include the development of quantitative risk assessment models, the design of automated detection and mitigation mechanisms, and the evaluation of next-generation contact tracing systems that integrate advanced privacy-preserving techniques.

In addition, the evolution of EN must address cross-cutting challenges that transcend individual attack categories, particularly in the areas of interoperability, usability, and transparency. Standardization of cryptographic and communication protocols is essential to close interoperability gaps that attacks such as *Still-Valid Keys* exploit, especially when users travel across borders. Federated trust models can further strengthen cross-jurisdictional validation, ensuring that TEKs and RPIs from foreign EN systems are consistently authenticated. At the same time, balancing security with usability is critical: overly intrusive measures, such as frequent verification prompts, risk alienating users, while adaptive security policies–such as relaxing verification in low-risk areas and tightening it in high-risk zones–can maintain both protection and user engagement. Transparency and accountability are equally vital, as regular security audits and clear incident response protocols not only build public trust, but also ensure compliance with evolving regulatory requirements.

Looking beyond immediate threats, long-term resilience is also important for future EN system design. Future-proofing against emerging risks, such as quantum computing, requires the adoption of post-quantum cryptographic algorithms for the generation of TEK and RPI. AI-based security could also offer another layer of defense, enabling real-time anomaly detection to identify and mitigate novel attacks as they arise. Decentralized architectures, including blockchain or distributed ledger technologies, present opportunities to enhance transparency and tamper-proofing, reducing reliance on centralized authorities, and minimizing single points of failure. These forward-looking strategies ensure that EN systems remain robust against both current vulnerabilities and those that may emerge as technology evolves.

Ultimately, the effectiveness of EN systems hinges not only on addressing specific attack vectors, but also on embedding adaptability and user-centric principles into their core architecture. By prioritizing interoperability, usability, and future readiness, designers can create systems that are not only secure and private but also scalable and trusted by the

public. This holistic approach is essential to maintain the integrity and efficacy of digital contact tracing in current and future public health crises.

7 Conclusions

The EN system has played a pivotal role in digital contact tracing during the COVID-19 pandemic. As the system continues to evolve to address emerging pandemic challenges, a thorough understanding of its vulnerabilities and attack vectors is essential to ensure its long-term effectiveness and resilience.

In this paper, a general taxonomy for EN-based attacks is presented. A set of 47 existing attacks is classified into three main categories—Denial of Service, False Alert Injection, and Information Disclosure—according to their characteristics. The proposed taxonomy introduces six new subcategories and uses a granular, multi-level, tree-like structure that includes all known EN-related attacks. These new categories enable a better understanding of EN's current attack vectors and highlight areas that can be further explored, analyzed, and corrected.

The analysis of the attacks and the taxonomy, highlight that Information Disclosure attacks dominate the attack surface, emphasizing persistent challenges in preserving user privacy despite the system's decentralized design. Additionally, the taxonomy reveals that attacks can be combined across categories, reinforcing the need to consider multi-stage threat scenarios in security evaluations. From a practical perspective, the proposed taxonomy supports more systematic security analysis, and it offers a structured foundation for both researchers and practitioners to better understand vulnerabilities and improve the resilience of EN-based systems.

This research aims to contribute to enhancing the efficiency, security, and privacy of contact tracing systems by providing a holistic view of their associated attacks. Recent research also highlights that, by integrating real-time data analytics and adaptive privacy safeguards, AI-based systems offer a scalable and ethically sound solution for future pandemic preparedness. However, further research is needed to address challenges related to widespread adoption, interoperability across regions, and the evolving landscape of cyber threats targeting these systems. The development of resilient, secure contact tracing frameworks thus remains a critical priority for global public health infrastructure.

Acknowledgements This work was supported and received funding from "CyberPRAISE - Cybersecurity research for PRivAte, Intelligent and truStablE solutions" - NORTE2030-FEDER-01820300.

Author Contributions P.P. designed the study, H.F. and P.P. collected and processed all relevant information and wrote the main manuscript text and P.R. prepared the discussion. All authors reviewed the manuscript.

Funding Open access funding provided by FCT|FCCN (b-on).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Naseer, S., Khalid, S., Parveen, S., Abbass, K., Song, H., Achim, M.V.: Covid-19 outbreak: impact on global economy. Front. Public Health (2023). https://doi.org/10.3389/fpubh.2022.1009393
2. Ozbay, G., Sariisik, M., Ceylan, V., Çakmak, M.: A comparative evaluation between the impact of previous outbreaks and covid-19 on the tourism industry. Int. Hosp. Rev. **36**(1), 65–82 (2021). https://doi.org/10.1108/IHR-05-2020-0015. (https://www.emerald.com/ihr/article-pdf/36/1/65/711826/ihr-05-2020-0015.pdf)
3. Ben Nasr, I., Kondrateva, G., Khvatova, T., Ben Arfi, W.: The role of contact-tracing mobile apps in pandemic prevention: a multidisciplinary perspective on health beliefs, social, and technological factors. Soc. Sci. Med. **358**, 117204 (2024). https://doi.org/10.1016/j.socscimed.2024.117204
4. Kouliaridis, V., Kambourakis, G., Chatzoglou, E., Geneiatakis, D., Wang, H.: Dissecting contact tracing apps in the android platform. PLoS ONE **16**(5), 0251867 (2021). https://doi.org/10.1371/journal.pone.0251867
5. Leslie, M.: COVID-19 fight enlists digital technology: contact tracing apps. Engineering (2020). https://doi.org/10.1016/j.eng.2020.09.001
6. Skoll, D., Miller, J.C., Saxon, L.A.: COVID-19 testing and infection surveillance: is a combined digital contact tracing and mass testing solution feasible in the United States? Cardiovasc. Dig. Health J. (2020). https://doi.org/10.1016/j.cvdhj.2020.09.004
7. Faria, H., Paiva, S., Pinto, P.: An advertising overflow attack against android exposure notification system impacting covid-19 contact tracing applications. IEEE Access **9**, 103365–103375 (2021). https://doi.org/10.1109/ACCESS.2021.3099017
8. Google: Exposure Notifications API: https://developers.google.com/android/exposure-notifications/exposure-notifications-api (2021). Accessed 22 Mar 2021
9. Ramamoorthy, S., Mahon, J., O'Mahony, M., Itangayenda, J.F., Mukande, T., Makati, T.: Automatic contact tracing using bluetooth

low energy signals and imu sensor readings. arXiv:2206.06033 (2022)
10. Hsiao, H., Huang, C., Hong, B., Cheng, S., Hu, H., Wu, C., Lee, J., Wang, S., Jeng, W.: An empirical evaluation of bluetooth-based decentralized contact tracing in crowds. CoRR. arxiv:2011.04322 (2020)
11. Legendre, F., Humbert, M., Mermoud, A., Lenders, V.: Contact tracing: an overview of technologies and cyber risks. arXiv:2007.02806 [cs] (2020). Accessed 31 Jan 2021
12. Vaudenay, S.: Analysis of DP3T. Cryptology ePrint Archive, Report 2020/399. https://ia.cr/2020/399 (2020)
13. Dehaye, P.-O., Reardon, J.: SwissCovid: a critical analysis of risk assessment by Swiss authorities. arXiv:2006.10719 [cs] (2020). Accessed 22 Feb 2021
14. Daw, E.: Component-based comparison of privacy-first exposure notification protocols. IACR Cryptol. ePrint Arch. **2020**, 586 (2020)
15. Pietrzak, K.: Delayed authentication: preventing replay and relay attacks in private contact tracing. In: Bhargavan, K., Oswald, E., Prabhakaran, M. (eds.) Progress in Cryptology – INDOCRYPT 2020. Lecture Notes in Computer Science, pp. 3–15. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-65277-7$_1$
16. Gvili, Y.: Security analysis of the covid-19 contact tracing specifications by apple inc. and google inc. IACR Cryptol. ePrint Arch **2020**, 428 (2020)
17. Avitabile, G., Friolo, D., Visconti, I.: Terrorist attacks for fake exposure notifications in contact tracing systems. In: Sako, K., Tippenhauer, N.O. (eds.) Applied Cryptography and Network Security, pp. 220–247. Springer, Cham (2021)
18. Dehaye, P.-O., Reardon, J.: Proximity tracing in an ecosystem of surveillance capitalism. arXiv:2009.06077 [cs] (2020). https://doi.org/10.1145/3411497.3420219
19. Google: exposure notification: risks and mitigations FAQ. Web site (2020). https://github.com/google/exposure-notifications-internals/blob/main/en-risks-and-mitigations-faq.md. Accessed 18 Feb 2021 **(accessed on 1 February 2021)**
20. Ang, V., Shar, L.K.: Covid-19 one year on - security and privacy review of contact tracing mobile apps. IEEE Pervas. Comput. **20**(4), 61–70 (2021). https://doi.org/10.1109/MPRV.2021.3115478
21. Hamza, M., Khan, A.A., Akbar, M.A.: Toward a secure global contact tracing app for COVID-19. In: Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering, pp. 453–460 (2022)
22. Buccafurri, F., Angelis, V.D., Labrini, C.: A privacy-preserving solution for proximity tracing avoiding identifier exchanging. CoRR (2020) arxiv:2005.10309
23. Huang, J., Yegneswaran, V., Porras, P., Gu, G.: On the privacy and integrity risks of contact-tracing applications. arXiv:2012.03283 [cs] (2020). Accessed 25 Feb 2021
24. Baumgärtner, L., Dmitrienko, A., Freisleben, B., Gruler, A., Höchst, J., Kühlberg, J., Mezini, M., Mitev, R., Miettinen, M., Muhamedagic, A., Nguyen, T.D., Penning, A., Pustelnik, D., Roos, F., Sadeghi, A.-R., Schwarz, M., Uhl, C.: Mind the GAP: security privacy risks of contact tracing apps. In: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 458–467 (2020). https://doi.org/10.1109/TrustCom50675.2020.00069. ISSN: 2324-9013
25. Iovino, V., Vaudenay, S., Vuagnoux, M.: On the effectiveness of time travel to inject covid-19 alerts. In: Paterson, K.G. (ed.) Topics in Cryptology - CT-RSA 2021, pp. 422–443. Springer, Cham (2021)
26. Hoepman, J.-H.: A critique of the google apple exposure notification (gaen) framework. In: Privacy Symposium: Data Protection Law International Convergence and Compliance with Innovative Technologies, pp. 41–58. Springer, Berlin (2022)
27. Gangavarapu, A., Daw, E., Singh, A., Iyer, R., Harp, G., Zimmerman, S., Raskar, R.: Target privacy threat modeling for COVID-19 exposure notification systems. arXiv:2009.13300 [cs] (2020). Accessed 3 Apr 2021
28. Avitabile, G., Botta, V., Iovino, V., Visconti, I.: Towards defeating mass surveillance and sars-cov-2: the pronto-c2 fully decentralized automatic contact tracing system. IACR Cryptol. ePrint Arch. **2020**, 493 (2020)
29. Iovino, V., Vaudenay, S., Vuagnoux, M.: On the effectiveness of time travel to inject covid-19 alerts. In: Paterson, K.G. (ed.) Topics in Cryptology - CT-RSA 2021, pp. 422–443. Springer, Cham (2021)
30. Dar, A.B., Lone, A.H., Zahoor, S., Khan, A.A., Naaz, R.: Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. Comput. Sci. Rev. **38**, 100307 (2020). https://doi.org/10.1016/j.cosrev.2020.100307
31. Krehling, L., Essex, A.: A security and privacy scoring system for contact tracing apps. J. Cybersecur. Priv. **1**(4), 597–614 (2021). https://doi.org/10.3390/jcp1040030
32. Mirkovic, J., Reiher, P.: A taxonomy of DDoS attack and DDoS defense mechanisms. ACM SIGCOMM Comput. Commun. Rev. **34**(2), 39–53 (2004). https://doi.org/10.1145/997150.997156. (**Accessed 2021-01-31**)
33. Chen, B.-R., Hu, Y.-C.: Mitigating denial-of-service attacks on digital contact tracing. In: Proceedings of the 18th Conference on Embedded Networked Sensor Systems. SenSys '20, pp. 770–771. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3384419.3430599. Accessed 24 Feb 2021
34. Hussein, M.R., Shams, A.B., Apu, E.H., Mamun, K.A.A., Rahman, M.S.: Digital surveillance systems for tracing COVID-19: privacy and security challenges with recommendations. CoRR (2020). arxiv:2007.13182
35. Ellis, C., Wen, H., Lin, Z., Arora, A.: Replay (far) away: exploiting and fixing google/apple exposure notification contact tracing. Proc. Priv. Enhanc. Technol. **4**, 727–745 (2022)
36. Nomoto, K., Akiyama, M., Eto, M., Inomata, A., Mori, T.: On the feasibility of linking attack to google/apple exposure notification framework. Proc. Priv. Enhanc. Technol. **4**, 140–161 (2022)
37. Rowe, F., Ngwenyama, O., Richet, J.-L.: Contact-tracing apps and alienation in the age of COVID-19. Eur. J. Inf. Syst. (2020). https://doi.org/10.1080/0960085X.2020.1803155