

CAPABILITY-VALIDATED INBOUND DESCRIPTOR (CVID)

Infrastructure-Level Privacy and Security for Digital Communications (Online Telephony Fraud and SPAM Control) in AI ERA

Technical Research Note

Author: Sangam Das

Independent Researcher, Electronics & Telecommunications Engineering

EXECUTIVE SUMMARY

Current digital communication infrastructure—including telephony, email, and messaging—treats contact identifiers such as **telephone numbers, email addresses, messaging IDs, and device identifiers** as persistently reachable endpoints. Once these identifiers are exposed, communication attempts can generally be initiated by any party that possesses them.

This architectural model creates structural vulnerabilities, including:

- spam and robocalls
- phishing and social-engineering attacks
- identifier harvesting
- long-term fraud exposure
- large-scale automated targeting
- construction of persistent tracking or surveillance datasets

A central practical problem is that most digital fraud, unsolicited communication, and spam campaigns begin **when malicious actors obtain a user's real contact identifiers**—most commonly a **telephone number or email address**.

These identifiers are frequently collected through:

- data breaches
- illegal data brokerage markets

- scraping of public services and websites
- application data leakage
- secondary disclosure across platforms

Once acquired, these identifiers become **long-lived entry points for repeated abuse**.

In many cases, the fraud itself is not the first event; **the first event is the exposure of the identifier**, which then enables:

- phishing campaigns
- impersonation attacks
- harassment and spam
- account takeover attempts
- large-scale automated targeting

As long as **raw identifiers remain widely exposed across digital systems**, these abuse pathways remain structurally difficult to prevent.

The Growing Role of AI in Scaling Fraud

Recent advances in artificial intelligence are dramatically increasing the effectiveness and scale of such attacks.

Only a few years ago, large-scale robocalling campaigns were limited by relatively simple prerecorded messages and manual fraud operations. Today, AI systems enable **millions of highly convincing calls, messages, and phishing attempts to be generated within a single day**, often using realistic **human-like synthetic voices and conversational agents**.

AI systems can now:

- generate human-like voice calls at massive scale
- personalize phishing messages automatically
- analyze harvested data to target victims more precisely
- conduct automated conversational fraud using voice or chat systems

Because these capabilities can be deployed at **very low operational cost**, the economic barrier to large-scale fraud is rapidly decreasing.

If the underlying architecture continues to expose raw identifiers such as **telephone numbers, email addresses, and device identifiers**, AI-driven fraud campaigns may scale dramatically,

allowing attackers to target **millions of individuals simultaneously** with highly convincing automated interactions.

Reduced Human Requirements for Large-Scale Fraud

A particularly concerning aspect of AI-enabled fraud is that **large numbers of human operators are no longer required to conduct large-scale scams.**

Historically, fraud operations such as call-center scams required **large teams of human callers**, physical infrastructure, and manual coordination. Modern AI systems significantly change this model. A small group—or even a single individual—can deploy automated systems that generate voice calls, respond to victims conversationally, and adapt scripts dynamically based on the victim's responses.

Illustrative Example

Consider a scenario in which a malicious actor obtains a dataset of **one million phone numbers from a data broker**. With modern AI tools, the attacker can deploy an automated voice system capable of:

- generating realistic human-like voice calls
- conducting interactive conversations with victims
- adapting responses using AI language models
- operating continuously without human intervention

In such a scenario, the attacker does not need thousands of human callers. **A single automated system can simultaneously place tens of thousands of calls**, conducting conversations with multiple victims at the same time.

This dramatically reduces the operational cost of fraud and increases its potential scale.

As a result, if **raw identifiers such as phone numbers, email addresses, or device IDs remain widely accessible**, AI-enabled fraud systems may allow attackers to **scale scams with unprecedented efficiency**.

For this reason, addressing the **root structural issue—exposure of raw identifiers—becomes increasingly important**. Preventing large-scale harvesting of these identifiers can significantly reduce the ability of automated systems to target individuals at scale.

The Root Structural Problem

Current defensive approaches—including:

- spam filters
- caller authentication frameworks
- blocking tools
- privacy policies
- platform moderation

are largely **reactive mechanisms**.

They attempt to mitigate abuse **after identifiers have already been exposed** or **after communication attempts have already occurred**.

However, these approaches do not address the **root structural vulnerability**, which is the widespread exposure and reuse of **raw user identifiers across digital systems**.

A Root-Level Mitigation Approach

A more fundamental approach is to reduce or eliminate the routine exposure of **raw identifiers** in operational digital workflows.

This can be achieved through **identifier virtualization**, in which systems use **temporary or purpose-bound identifiers** instead of directly exposing persistent contact identifiers such as:

- email addresses
- telephone numbers
- messaging handles
- device identifiers

Under such architectures, communication systems operate through **controlled virtual identifiers**, while the underlying real identifiers remain protected within trusted identity management systems.

By limiting the circulation of raw identifiers across platforms and services, this approach can significantly reduce the ability of **data brokers, scraping systems, and malicious actors to harvest and reuse user identifiers at scale**.

In simple terms, the most effective structural defense against large-scale targeting is:

do not expose raw identifiers in the first place.

Note on AI-Enabled Population-Scale Surveillance Risks

The rapid development of artificial intelligence introduces a growing concern: the potential for **population-scale surveillance through automated fusion of large digital datasets**.

It is important to note that modern AI systems **do not rely solely on a single identifier such as a telephone number or email address**. Contemporary data analytics techniques increasingly combine **multiple identifiers and datasets simultaneously**, including:

- phone numbers
- email addresses
- device identifiers
- IP addresses
- location signals
- social-network interactions
- browsing or application activity
- public records or commercial datasets

Through AI-based data fusion, these different signals can be **correlated and analyzed together**, allowing systems to infer behavioral patterns, relationships, and activity profiles at very large scale.

This capability is significantly more powerful today than it was only a few years ago. AI systems can automatically process massive datasets, identify patterns, and link data points across multiple services or platforms. When persistent identifiers circulate widely across digital ecosystems, they become **reference anchors that enable cross-system correlation**.

The risk therefore does not arise from a single identifier alone, but from the **ability of AI systems to combine many signals and datasets simultaneously**.

This phenomenon has already been discussed in various European policy and research contexts. For example:

- The **European Union Agency for Cybersecurity (ENISA)** has published reports on large-scale data analytics and emerging cyber risks, highlighting how automated analysis of large datasets can create new security and privacy challenges.

- The **European Data Protection Board (EDPB)** and **European Commission policy discussions around the AI Act** have also referenced risks related to large-scale data processing, automated profiling, and cross-system data aggregation.
- The **Joint Research Centre (JRC) of the European Commission** has examined how AI-driven data analysis can enable new forms of digital risk and behavioral inference.

Although I am not a cybersecurity specialist, the underlying issue described here—**large-scale automated data correlation enabled by AI**—is widely recognized in the cybersecurity and data-protection community. Independent experts in cybersecurity, telecommunications infrastructure, or AI governance would be able to validate the technical plausibility of these concerns.

For this reason, limiting the widespread exposure of persistent identifiers may help reduce the ability of AI systems to assemble large-scale correlation datasets.

The architecture proposed in this submission addresses this challenge through:

- **Capability-Validated Inbound Descriptor (CVID)** to control inbound communication reachability
- **Virtual Identity (VI)** to avoid exposing persistent identifiers during operational processing
- **Compliance Jurisdiction Token (CJT)** to enforce purpose-bound authorization for data use

By reducing the circulation of raw identifiers and replacing them with controlled, purpose-bound identifiers, such architectures may make it more difficult for external actors to accumulate the stable identifier anchors required for **AI-driven population-scale data correlation and surveillance**.

A Structural Security Principle

If raw identifiers are not widely distributed across digital systems, the ability of malicious actors to collect and reuse them for spam, fraud, and surveillance is substantially reduced.

In this sense, **preventing identifier exposure addresses the problem at its root**, rather than attempting to mitigate abuse after identifiers have already been leaked.

As artificial intelligence continues to reduce the cost and increase the scale of automated fraud, architectural approaches that **limit the exposure of persistent identifiers** may become an increasingly important component of digital communication security.

Final Note: Data Minimization and Limiting the Export of Raw Identifiers

A key principle emerging from the risks described above is the importance of **data minimization and limiting the export of raw identifiers across digital systems**.

Modern AI systems are increasingly capable of combining and analyzing multiple datasets simultaneously. When persistent identifiers—such as **telephone numbers, email addresses, device identifiers, or other stable identifiers**—are **widely shared across services**, they create stable reference points that enable large-scale data correlation and automated profiling.

For this reason, reducing the routine exposure and transmission of such identifiers can play an important role in limiting both **AI-enabled fraud operations and large-scale automated surveillance risks**.

In practical terms, a simple structural principle can be considered:

Do not export raw identifiers unless strictly necessary for the intended service.

Instead, communication and data-processing systems can adopt approaches consistent with established European data protection principles, including:

- **data minimization**
- **purpose limitation**
- **privacy by design**

This may include the use of **virtualized identifiers, purpose-bound tokens, or controlled communication capabilities**, such as the **CVID + VI + CJT architecture described in this submission**.

By ensuring that operational systems rely on **temporary or purpose-bound identifiers rather than widely distributed persistent identifiers**, it becomes more difficult for external actors to harvest identifiers, construct large correlation datasets, or conduct large-scale automated targeting.

Such architectural approaches complement existing regulatory frameworks by addressing risks **at the infrastructure level**, helping reduce exposure of sensitive identifiers before they can be misused.

Capability-Validated Inbound Descriptor (CVID)

(Not an email alias or Virtual number)

Capability-Validated Inbound Descriptor (CVID) proposes a different model. Instead of treating raw contact identifiers as permanently reachable public endpoints, CVID replaces them with cryptographically **purpose-bound, time-limited, quota-enforced communication capabilities**.

Under this approach, inbound communication is permitted only when:

- a valid communication capability exists,
- its scope covers the intended interaction, and
- its quota, duration, and channel constraints remain valid.

A key architectural property of CVID is that the raw contact identifier need not be disclosed to the communicating party. The communicating entity holds only the authorized capability descriptor, while routing to the real contact endpoint occurs through protected infrastructure under controlled conditions.

This produces several technical effects:

- raw contact identifiers are no longer routinely exposed across external systems;
- communication can be limited to defined purpose, time, quantity, and channel constraints;
- misuse of compromised capabilities is bounded; and
- large-scale harvesting and persistent contact-based profiling become substantially more difficult.

In this sense, CVID is not an aliasing or forwarding mechanism. It is an **infrastructure-level communication authorization model**.

From a regulatory perspective, CVID is relevant to:

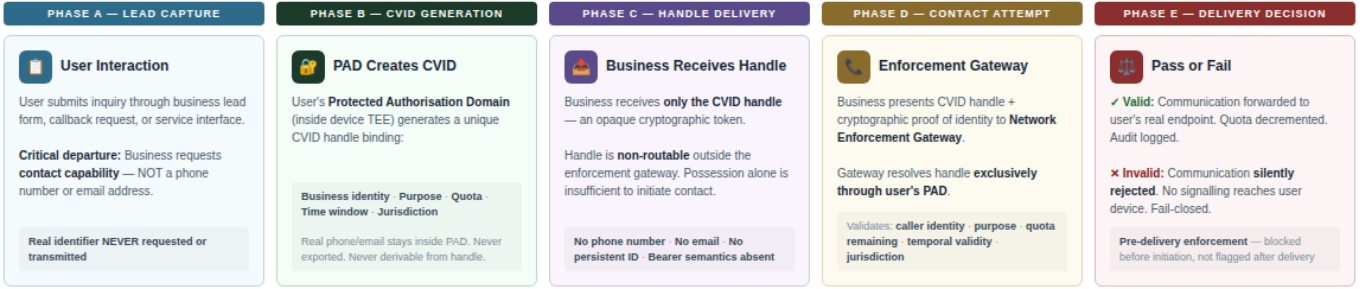
- **GDPR principles**, particularly data minimization, purpose limitation, storage limitation, and security of processing; and
- **NIS2 objectives**, particularly attack-surface reduction, fraud prevention, and incident prevention.

CVID is therefore presented as a technical mechanism for exploring how communication-layer privacy and security can be strengthened without requiring a redesign of the legal framework itself.

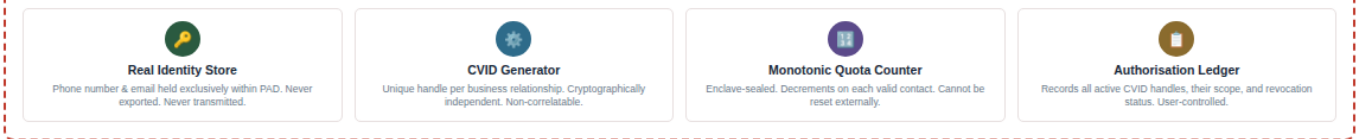
CVID Architecture — Capability-Validated Inbound Communication

PRE-DELIVERY ENFORCEMENT · IDENTITY PROTECTION ACROSS TELEPHONY, EMAIL & PAYMENT CHANNELS

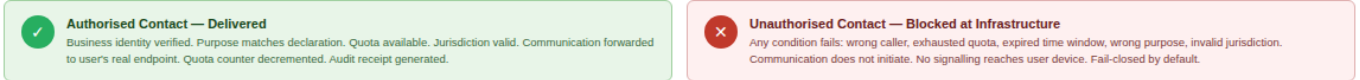
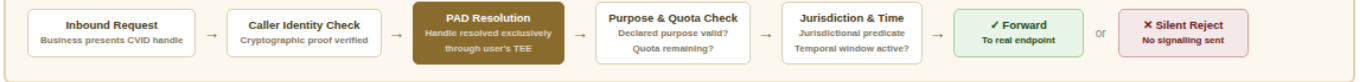
OPERATIONAL FLOW — FIVE PHASES



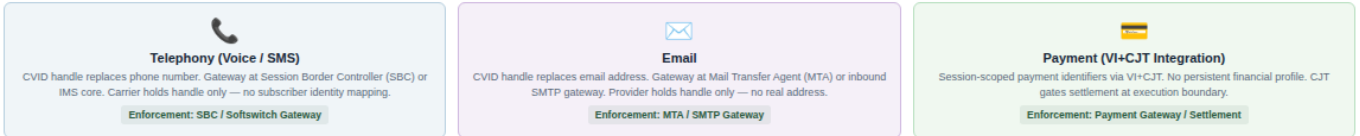
PROTECTED AUTHORISATION DOMAIN (PAD) — INSIDE USER'S TEE / SECURE ENCLAVE



NETWORK ENFORCEMENT GATEWAY — PRE-DELIVERY VALIDATION SEQUENCE



MULTI-CHANNEL ENFORCEMENT — SAME ARCHITECTURE, ALL COMMUNICATION LAYERS



SURVEILLANCE RESISTANCE COMPARISON

SURVEILLANCE VECTOR	CURRENT ARCHITECTURE	UNDER CVID
Persistent identifier as identity anchor	Present — phone/email is stable, globally unique, lifelong	Absent — real identifier exists only within user's PAD
Cross-service correlation	Trivial — same identifier used across all services	Impossible — each service holds independent CVID handle
Compelled disclosure from carrier/provider	Yields real identity — carrier holds subscriber mapping	Yields CVID handles only — no identity mapping exists outside PAD
Relationship graph construction	Straightforward — call/email metadata reveals full network	Structurally prevented — unique handle per relationship
Bulk population surveillance	Effective — list of numbers/emails identifies population	Ineffective — list of handles is computationally inert
Spam, vishing, and unsolicited contact	Possible — identifier possession confers contact right	Blocked — contact requires cryptographic authorisation at gateway

A · LEAD CAPTURE ▶

B · CVID GENERATION ▶

C · HANDLE DELIVERY ▶

D · ENFORCEMENT ▶

E · DECISION

Enforcement Points: Session Border Controller (SBC) · Mail Transfer Agent (MTA) · Payment Gateway · API Boundary

Architecture Properties: Fail-closed · Non-bearer · Purpose-bound · Quota-limited · Vendor-neutral · Post-quantum forward-compatible

CVID Framework
 Anti-Fraud Action Plan Submission
 Ref. Ares(2025)783547
 Independent peer review invited

PART 1: THE STRUCTURAL PROBLEM IN CURRENT COMMUNICATION SYSTEMS

Permanent Reachability as an Architectural Weakness

Most modern digital communication systems operate on a simple assumption:

if a party possesses a reachable identifier, it can attempt communication.

Examples include:

- a phone number that can be dialed,
- an email address that can be messaged,
- a messaging identifier that can be contacted.

These identifiers function as **globally or broadly reachable routing handles**. As a result, once exposed, they can become persistent targets for abuse.

Consequences of Persistent Identifier Exposure

This design leads to a number of recurring vulnerabilities:

Vulnerability	Typical Mechanism	Practical Consequence
Spam and robocalls	Harvested phone numbers; automated dialing systems	User nuisance; infrastructure load; trust erosion
Phishing and fraud	Email or SMS used for deceptive communications	Financial loss; identity compromise; social engineering
Identifier harvesting	Data breaches, scraping, data brokerage	Long-term exposure of contact points
Cross-platform linkage	Reuse of same identifier across multiple services	Behavioral profiling; loss of privacy
Account takeover vectors	Reuse of exposed phone/email in recovery channels	Authentication abuse; financial risk
Surveillance graph construction	Contact identifiers linked across systems and time	Persistent tracking and profiling

Why Existing Defenses Are Not Sufficient

Current protective measures generally do not alter the underlying architectural assumption of permanent reachability. For example:

- **spam filters** detect or classify messages after transmission,
- **caller authentication systems** help address spoofing but not reachability,

- **user-side blocking** shifts burden to recipients,
- **policy controls** are contractual rather than technical, and
- **network obfuscation tools** may hide traffic origin without preventing identifier exposure.

These mechanisms are useful, but they do not prevent raw contact identifiers from becoming durable attack surfaces.

PART 2: CVID — PURPOSE-BOUND COMMUNICATION CAPABILITIES

Core Concept

CVID replaces permanently reachable raw identifiers with **validated communication capabilities**.

A CVID descriptor authorizes a defined inbound interaction according to parameters such as:

- permitted purpose,
- allowed channel,
- quota or frequency,
- validity duration, and
- contextual scope.

The critical difference is that the entity holding the CVID does **not need to possess the recipient's raw telephone number or email address**.

Functional Model

Phase 1: Capability Issuance

A legitimate interaction occurs—for example, a user submits a support request, places an order, or requests an account notification.

The system then issues a CVID descriptor specifying:

- purpose,
- communication channel,
- maximum number of allowed interactions, and

- validity period.

The user's raw contact data remains in a protected resolver or identity vault controlled by the originating system.

Phase 2: Capability-Based Communication

When the communicating party initiates contact, it presents the CVID rather than the raw identifier.

Infrastructure verifies:

- whether the capability is valid,
- whether the proposed use matches the authorized purpose,
- whether quota remains available, and
- whether the capability has expired.

If verification succeeds, protected infrastructure routes the communication internally to the real contact endpoint.

Phase 3: Enforcement

If any condition fails, the communication attempt is blocked before completion.

This means that misuse is constrained not by policy promises alone, but by technical enforcement of:

- purpose,
- duration,
- channel, and
- quantity.

PART 3: WHY CVID IS DISTINCT FROM ALIASES OR VIRTUAL NUMBERS

CVID is **not** equivalent to:

- virtual phone numbers,
- masked caller identifiers,
- secondary email aliases, or
- forwarding addresses.

These existing mechanisms may hide a raw identifier temporarily, but they usually preserve **reachability**. Anyone who obtains the alias can still attempt repeated contact.

CVID differs in that it does not merely substitute one reachable address for another. Instead, it makes communication dependent on **possession of a valid, bounded, and verifiable capability**.

In simplified terms:

- **aliasing hides identity,**
- **CVID governs whether communication is permitted at all.**

This distinction is particularly important for spam prevention, fraud reduction, and privacy protection.

PART 4: SECURITY AND PRIVACY EFFECTS

1. Reduced Identifier Exposure

Because raw contact identifiers are not routinely disclosed to communicating parties, the opportunities for:

- harvesting,
- resale,
- linkage, and
- reuse across contexts

are reduced.

2. Bounded Misuse

Even if a CVID capability were intercepted or compromised, its misuse would remain constrained by:

- fixed quota,
- fixed purpose,
- fixed channel, and
- fixed validity window.

This is materially different from compromise of a raw phone number or email address, which may remain useful indefinitely.

3. Stronger Communication-Layer Privacy

CVID supports a model in which communication takes place without turning the contact identifier itself into a public-facing resource.

4. Upstream Prevention

Because enforcement may occur at gateway, messaging, or network-control layers, unauthorized attempts can be stopped before they become persistent nuisance or fraud vectors.

PART 5: REGULATORY RELEVANCE

GDPR Relevance

CVID is relevant to several GDPR principles, including:

- **Article 5(1)(b) – Purpose limitation**
Communication is restricted to authorized purposes.
- **Article 5(1)(c) – Data minimization**
Raw identifiers need not be exposed to each communicating entity.
- **Article 5(1)(e) – Storage limitation**
Capabilities can expire automatically rather than persisting indefinitely.
- **Article 5(1)(f) and Article 32 – Integrity, confidentiality, and security of processing**
Attack surface associated with contact identifiers is reduced.
- **Article 21 – Right to object**
Revocation of communication capability can help operationalize objection rights.

NIS2 Relevance

CVID also supports broader cybersecurity and resilience objectives by:

- reducing unauthorized inbound attack vectors,
- limiting robocall and phishing exposure, and

- reducing the value of harvested contact datasets.

PART 6: CONCLUSION

The core problem addressed by CVID is architectural:

current communication systems treat contact identifiers as permanently reachable endpoints.

This model makes spam, fraud, and long-term identifier exploitation difficult to avoid once exposure occurs.

CVID proposes an alternative approach in which communication is authorized through **purpose-bound, time-limited, quota-enforced capabilities**, while routing to the real identifier remains protected within controlled infrastructure.

The result is a communication model in which:

- legitimate communication remains possible,
- misuse is technically constrained,
- raw identifiers are better protected, and
- privacy and security objectives can be supported at the infrastructure layer rather than only through post-hoc filtering or policy controls.

CVID is therefore best understood as a **communication-layer privacy and security mechanism** that complements broader digital governance goals under the GDPR, NIS2, and related European regulatory frameworks.