EU CyberNet

A Digital Emblem for Cyber Conflicts: Making Humanitarian Protection Machine-Readable[i]

- Jersain Zadamig Llamas Covarrubias

EU CyberNet Expert Series
No 6
2025

about

Jersain Zadamig Llamas Covarrubias

Jersain Llamas is a Mexican lawyer and technology strategist who bridges law, policy and engineering to address contemporary cyber risks. He holds an LL.M., a PhD in Data Science and professional certifications including CISO, CEH (EC-Council) and CIPP/E (IAPP). He is a research professor at the University of Guadalajara, Founder and CEO of OBEX Cybersecurity.

He has been part of the EU CyberNet Expert Pool since March 2024.

summary

Cyber operations now reach hospitals and humanitarian NGOs during armed conflicts. This peer-reviewed article argues that a digital emblem, a verifiable, machine-readable marker, can translate the protected status that International Humanitarian Law (IHL) already grants into a signal that operators and automated tools can detect early, reducing the risk of wrongful targeting and strengthening accountability when violations occur. [1][2] The idea is not a silver bullet, it is a practical layer that complements security controls and legal prohibitions, while addressing real risks such as exposure and misuse through authentication and governance measures. [1]

Why a Digital Emblem and Why Now?

Armed conflicts have expanded into the digital domain. Threat actors probe, map and sometimes disrupt networks that host medical services and humanitarian logistics. IHL already protects these entities, just as it does in the kinetic world. However, in cyberspace, the traditional visible emblems (Red Cross/Red Crescent/Red Crystal) are not reliably perceived by the people and machines making time-critical decisions. The digital emblem fills that practical gap: a discoverable, verifiable status signal that can be recognised during routine reconnaissance and pre-exploitation checks, so ethical operators and policy-constrained tooling can stand down before harm occurs. [1][2]

This proposal builds on an established legal foundation. The International Committee of the Red Cross (ICRC) has clarified how IHL principles (distinction, proportionality and precaution) apply to cyber operations, including the protection of medical units and humanitarian activities. [2] Parallel to the Red Cross emblems, the 1954 Hague Convention for the Protection of Cultural Property in Armed Conflict established the Blue Shield emblem to protect cultural heritage sites and institutions. [9] Similar to how cultural institutions display physical Blue Shield emblems on buildings to signal protection, the digital emblem would extend this proven protection model to cyberspace, marking the digital infrastructure of protected entities. Scholarly guidance, such as the Tallinn Manual 2.0, addresses the misuse of protective emblems and offers a doctrinal anchor for sanctioning digital analogues, reinforcing that perfidy and emblem abuse are prohibited in armed conflict, online as well as offline. [8] Politically, the 34th International Conference of the Red Cross and Red Crescent adopted a 2024 resolution on the human cost of ICT activities in war, signalling momentum for technical measures that operationalise protection in practice. [3][4]

What the Emblem Is and What It Is Not?

The digital emblem is not an alternative to firewalls, segmentation or incident response plans. It also does not create new rights, it operationalises the protection IHL already provides. Think of the digital emblem like a digital ID badge that computers can automatically verify. Just as a physical Red Cross emblem on a hospital roof is visible to pilots, the digital version makes protection visible to computer systems. In practical terms: imagine visiting a hospital's website, while you might not see anything different, automated systems scanning the network would immediately detect special markers (like a digital badge file at a known location, similar to how websites have a 'favicon' icon). In technical terms, it is a layered signal that can sit in standard discovery paths: a cryptographically-signed JSON file at a well-known endpoint [ii], DNS-based indicators that tag the domain or structured HTTP headers sent with every web request. All of these allow verification without exposing sensitive infrastructure details. [1]

Not everyone can claim a digital emblem, only entities already entitled to use the physical Red Cross, Red Crescent or Red Crystal emblems under IHL. This includes military medical services, authorised civilian hospitals and medical units, the ICRC and National Red Cross/Red Crescent Societies and humanitarian organisations operating under explicit state authorisation during armed conflicts. [1][2] The authorisation process would mirror the physical emblem's governance – a multi-tiered authentication system. Organisations like ICANN or Regional Internet Registries could issue dedicated resources (special domain names or IP ranges) only to verified entities, creating a verification chain where fraudulent use would be easily detectable because the cryptographic signature wouldn't trace back to an authorised issuer, similar to how fake physical emblems are prosecuted as war crimes under domestic law. [1][8]

Research prototypes, most notably ADEM (Authentic Digital EMblem), show how cryptographic design can enable 'covert inspection' – verification with freshness and authenticity guarantees, but minimal disclosure. [6] To be clear, ADEM does not use blockchain technology. Instead, it employs standard public-key cryptography and certificate chains, the same battle-tested technology securing HTTPS websites, because blockchain's distributed ledger approach would be too slow and resource-intensive for humanitarian operations in conflict zones with limited connectivity. [6] Addressing concerns about automated attacks: the digital emblem is fundamentally different from robots.txt, which relies on voluntary compliance without enforcement. The emblem operates at multiple levels: *technical* detection (reconnaissance tools can be programmed to check for and respect the emblem), *legal* deterrence (violating it constitutes a war crime under IHL) and *accountability* mechanisms (creating technical logs and evidence trails for post-attack prosecution). [1][8] While determined malicious actors may ignore it, just as some ignore physical Red Cross emblems, the emblem reduces preventable errors by law-abiding operators and creates clear evidence when violations occur. Evidence shows that even some cybercriminal groups have demonstrated sensitivity to humanitarian norms, voluntarily de-escalating in conflicts or apologising for hospital attacks. [15][16] In parallel, the IETF DIEM effort tracks discovery and validation profiles to support interoperable deployment across vendors and networks. [5]

In scope, the emblem should mark specific services, applications or datasets relevant to humanitarian or medical functions, not entire cloud regions or mixed-use infrastructure. Precision matters: over-marking would dilute credibility, while under-marking would miss the point. And because humanitarian networks are heterogeneous, the emblem must be easy to deploy, revoke and maintain at scale, with low cost and minimal operational burden, especially where skills and connectivity are constrained. [1]

Where It Helps Most: The Early Kill-Chain

The digital emblem faces real-world constraints that merit candid discussion. First, its legal protection activates only during armed conflict (though entities can deploy it preparatorily to ensure immediate protection once conflict starts). [1][2] Second, dual-use infrastructure presents challenges: a hospital running both patient care systems and general administrative networks would need precise marking – only the medical systems qualify for emblem protection, not the payroll servers. [1] Third, the emblem cannot prevent all reconnaissance. Rather, it identifies protected assets early so ethical operators can exclude them from target lists. Fourth, and perhaps most importantly, recent conflicts in Ukraine and Gaza demonstrate that not all actors respect IHL, whether in kinetic or cyber domains. These are serious limitations.

However, three factors support pursuing the emblem despite this scepticism: (1) Even partial compliance reduces harm, while we lack quantitative data on effectiveness rates, empirical evidence shows that some non-state cyber actors have demonstrated sensitivity to humanitarian norms. During the Ukraine conflict, hacker collectives voluntarily de-escalated attacks on certain targets, and the LockBit ransomware group publicly apologised after inadvertently attacking a children's hospital. [15][16] This suggests that clear, machine-readable humanitarian markers can influence behaviour even among actors not traditionally bound by IHL. Any reduction in preventable targeting errors by law-abiding operators represents lives and critical infrastructure saved. (2) It creates accountability, technical logs and cryptographic evidence enable post-conflict prosecution of emblem violations, strengthening deterrence and enforcement over time. [1][8] (3) Historical precedent physical emblems remain valuable despite

violations; the digital version extends this proven (if imperfect) protection model to cyberspace. The emblem is not a panacea, but a practical harm-reduction tool that turns 'do not target' obligations into machine-checkable constraints for operators and automated systems acting in good faith. [6]

Many preventable incidents begin with mis-targeting in reconnaissance, the phase when operators and crawlers map potential victims. Placing a recognizable, verifiable signal at this earliest step allows ethical actors and compliant tools to abort or escalate before delivery and exploitation and gives defensive controls a policy hook for stricter handling of emblem-tagged assets. [6] Framed against the well-known kill-chain model (reconnaissance, weaponization, delivery, exploitation, installation, C2, actions on objectives), the emblem's chief value is upstream – it turns 'do not target' from a moral afterthought into a machine-checkable constraint. [6]

Source: https://en.wikipedia.org/wiki/Cyber_kill_chain#/media/File:Intrusion_Kill_Chain_-_v2.png

Benefits, Risks and the Paper's Risk Treatment

Benefits. A digital emblem improves targeting clarity for responsible operators and automation, reduces 'we didn't know' claims and produces evidentiary telemetry (verification attempts, policy aborts) that matters when conduct is assessed against IHL. [1][2] It also gives CISOs and counsel a concrete governance control to reference with cloud providers, MSSPs and testing teams.

Risks. The chief concern is exposure: a visible marker must not become a ready-made target list. The article addresses this by favouring authenticated discovery, rate-limiting and monitored verification flows over noisy banners, alongside precise scoping (marking the humanitarian service, not the whole environment). [1] Spoofing is the second concern: an attacker might fake a mark to shield offensive infrastructure. Here the answer is cryptographic authenticity (PKI-anchored tokens, revocation mechanisms and

transparency logs), so consumers can reject unauthenticated or stale signals. [1][6] A third risk is false complacency, treating 'no emblem' as 'no protection'. IHL protection does not depend on the presence of a digital mark, so doctrine, training and tooling must reflect that baseline. [1][2][8]

Finally, there is the reality of legacy and regulated technology. Some medical devices cannot be modified for technical or safety reasons. The article proposes network-level signalling (e.g., at gateways or reverse proxies) and collaboration with manufacturers and standards bodies, so support can be embedded over time, including in IoT and IPv6 contexts. [1]

Law, Policy and Pathways to Adoption

Legally, the emblem should be integrated into the IHL framework that already regulates protective emblems in kinetic conflict. The article outlines plausible routes: a new Additional Protocol to the Geneva Conventions, targeted revisions (e.g., Annex I of Additional Protocol I) or *ad hoc* agreements between belligerents to recognise and respect the signal during hostilities. [1] None of these diminish the immediate applicability of IHL, instead, they clarify and harden obligations around a digital marker, particularly misuse and perfidy.

As states and organisations explore governance options, the article points to useful reference regimes from the EU cybersecurity acquis, not because they are IHL instruments, but because they offer implementation patterns. The EU Cybersecurity Act suggests approaches for certification and assurance, the EU Cyber Resilience Act codifies security-by-design obligations relevant to connected products, and the EU Cyber Solidarity Act emphasises cross-border detection and response capacity, elements that can inform emblem validation, vendor responsibilities and oversight without redefining humanitarian law. [11][12][13]

Complementary policy documents (e.g., the EU Cybersecurity Strategy and the draft 'Cyber Blueprint') illustrate how civilian protection goals can be translated into operational cooperation and crisis playbooks. [10][14]

Evidence from Practice and Why Signals Matter

Signals alone do not stop war crimes. Yet behaviour in cyberspace shows that norms and constraints can shape actor choices at the margins. During the war in Ukraine, parts of the hacktivist ecosystem publicly embraced a set of 'eight rules for civilian hackers', echoing core IHL protections and some ransomware groups temporarily declared moratoria on hospitals during COVID-19 and a notorious crew issued a decryption key after hitting a children's hospital and facing public backlash. [7][15][16] These episodes do not prove compliance, nor do they absolve offenders, but they suggest that clear, verifiable signals can alter calculations, reduce mistakes and strengthen attribution narratives when violations occur. The digital emblem is designed to be one such signal, grounded in law and consumable by machines.

What Cyber Professionals Can Do (Within Today's Bounds)

The operational chapter of this article's extended version (linked above) is intentionally pragmatic. It recommends naming a custodian (CISO + Legal + humanitarian liaison) to control issuance and revocation, scoping what is markable (clinical systems, relief-logistics platforms, priority datasets), and starting with low-friction discovery, such as a signed token at a well-known endpoint or DNS, then moving to authenticated verification as profiles mature. [1][5][6] SOCs can teach scanners to tag emblem-marked assets in the CMDB and create structured alerts for verification failures or suspicious scraping. Red-team toolchains can add a pre-flight emblem check that aborts by default when a valid mark is present,

documenting exceptions behind a two-person approval. Where devices cannot be modified, boundary controls (proxies, gateways) become the emblem's enforcement point.

Crucially, none of these steps imply that unmarked assets lose IHL protection. Training materials and playbooks must state the opposite. The emblem is a practical aid to distinction and precaution, not a gatekeeper to humanitarian status. [1][2][8] Similarly, while national legal positions on cyber operations and IHL (like those from the Netherlands, France or Australia) clarify state practice and guide military doctrine, the absence of such published positions does not diminish a state's obligations or the protection its entities enjoy, the fundamental rules apply universally, deriving from treaty law and customary IHL rather than policy papers. [2][8]

Limits and the Road Ahead

The paper is candid about limitations. High visibility can attract threat actors, which is why authentication, throttling and careful telemetry stewardship are integral to the design. [1] Heterogeneous stacks and low-resource settings demand low-cost, low-friction deployment. And emergent technologies, especially AI-enabled operations and information manipulation, raise fresh concerns. Here, the study notes techniques like watermarking and robust verification as part of the future agenda, alongside international cooperation to harmonise misuse sanctions and verification practices. [1][17] None of this replaces the need for resilient infrastructure and sound security hygiene, rather, the emblem becomes a bridge between legal protection and technical control.

Conclusion: Turning 'Do No Harm' into Code Paths

The digital emblem is 'a practical aid to distinction and precaution, not a gatekeeper to humanitarian status', but what does this mean in practice?

For *technical teams* in hospitals and humanitarian organisations, deploy the emblem on medical systems (patient databases, telemedicine platforms) but not administrative infrastructure, create standardised well-known/digital-emblem files on your servers, maintain clear inventories distinguishing protected from dual-use systems and test verification tools using emerging IETF DIEM specifications.

For *military and cyber operators*, integrate emblem checks into reconnaissance tools, just as physical emblems are verified before kinetic strikes, update Rules of Engagement to include emblem verification as a mandatory precaution step and train personnel on IHL protection obligations and emblem recognition in cyberspace.

For *policymakers and legal advisors*, publish national cyber-IHL positions that explicitly address digital emblem recognition, strengthen domestic criminal law to cover digital emblem misuse as it does physical violations and support standardisation efforts through IETF, ICANN and Regional Internet Registries.

For *everyone*, advocate for adoption by understanding the emblem's purpose and limitations, raise awareness about cyber threats to medical and humanitarian infrastructure and demand accountability when violations occur - prosecution strengthens compliance over time.

A digital emblem will not make a determined attacker ethical. It will, however, make the law legible to systems, reduce preventable errors and produce the evidence trails that matter when lines are crossed. That is already a humanitarian win. The legal logic exists, the operational need is real. What remains is disciplined implementation, starting small, verifying well and integrating the signal into everyday tooling, so that, even amid the fog of cyber conflict, the protected remain visible.

References

[i] https://revistas.flacsoandes.edu.ec/urvio/article/view/6332

[ii] www.hospital.org/.well-known/digital-emblem

[1] ICRC (2022), *Digitalizing the Red Cross, Red Crescent and Red Crystal Emblems: Benefits, Risks and Possible Solutions* - https://bit.ly/3AdwstG

[2] ICRC (2019), *International Humanitarian Law and Cyber Operations During Armed Conflicts: Position Paper* - https://bit.ly/3FA9mzZ

[3] Red Cross and Red Crescent, Resolution on protecting civilians and other protected persons and objects against the potential human cost of ICT activities during armed conflict (34IC/24/R2, Oct. 2024), *34th International Conference of the Red Cross and Red Crescent* - https://bit.ly/4bUFxpP

[4] Gisel, L. & Rodenhäuser, T. (2025), '¿Primer paso de otros muchos? Un avance para proteger a las poblaciones civiles contra el uso de tecnologías de la información y las comunicaciones durante conflictos armados' *ICRC Law & Policy Blog* - https://bit.ly/4iP5fyi

[5] IETF DIEM (Digital Emblems) Working Group-about & drafts - https://datatracker.ietf.org/group/diem/about/

[6] Linker, F. & Basin, D. (2023), 'ADEM: An Authentic Digital Emblem', *ACM CCS*.

[7] Rodenhäuser, T. & Vignati, M. (2023), '8 rules for civilian hackers during war, and 4 obligations for states to restrain them' *ICRC Law & Policy Blog* - https://bit.ly/3DObUtR

[8] Schmitt, M. (ed.) (2017), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Cambridge University Press.

[9] UNESCO (2025), 'Distinctive Emblems of the 1954 Hague Convention and of its 1999 Second Protocol' - https://www.unesco.org/en/heritage-armed-conflicts/emblem

[10] European Commission (2020), *The EU's Cybersecurity Strategy for the Digital Decade* -https://bit.ly/3FwzMmd

[11] European Union, *Regulation (EU) 2019/881 – Cybersecurity Act* (ENISA; certification) -https://bit.ly/4kVc6In

[12] European Union, *Regulation (EU) 2024/2847 – Cyber Resilience Act* (security-by-design) - https://bit.ly/41V0IDx

[13] European Union, *Regulation (EU) 2025/38 – Cyber Solidarity Act* (detection & response) - https://bit.ly/4kWecrm

[14] European Commission, *Cyber Blueprint – Draft Council Recommendation* (COM(2025) 66) - https://bit.ly/4kvjjyK

[15] Tidy, J. (2023), 'Ukraine Cyber-Conflict: Hacking Gangs Vow to De-Escalate', *BBC* -https://bit.ly/41W7DfO

[16] Lyons, J. (2023), 'LockBit: Sorry about the SickKids ransomware, not sorry about the rest' *The Register* - https://bit.ly/3SGMQJw

[17] Afina, Y. & Persi Paoli, G. (2024), *Governance of AI in the Military Domain* (UNIDIR-RAISE - https://bit.ly/4kusfEx

A Digital Emblem for Cyber Conflicts: Making Humanitarian Protection Machine-Readable
Jersain Zadamig Llamas Covarrubias

EU CybetNet Expert Series, No 6, 2025

EU CyberNet

Funded by the European Union