International Review of the Red Cross (2024), **106** (926), 703–731.
Selected Articles
doi:10.1017/S1816383124000018

IRRC_

A next-generation protective emblem: Cross-frequency protective options for non-combatants in the context of (fully) autonomous warfare

Daniel C. Hinck[1]*, Jonas J. Schöttler[2]*, Maria Krantz[3]*, Niklas Widulle[2]*, Katharina-Sophie Isleif[4] and Oliver Niggemann[5]†**
[1]Faculty of Medical Service and Health Sciences, Command and Staff College of the Federal Armed Forces, Hamburg, Germany
[2]Research Assistant, Chair of Computer Science in Mechanical Engineering, University of the Federal Armed Forces, Hamburg, Germany
[3]Post-Doctoral Researcher, Chair of Computer Science in Mechanical Engineering, University of the Federal Armed Forces, Hamburg, Germany
[4]Chair of Metrology, University of the Federal Armed Forces, Hamburg, Germany
[5]Chair of Computer Science in Mechanical Engineering, University of the Federal Armed Forces, Hamburg, Germany
*Equal contribution.
**Corresponding email: next_gen_protective_emblem@hsu-hh.de

Abstract

The protection of non-combatants in times of autonomous warfare raises the question of the timeliness of the international protective emblem. (Fully) Autonomous weapon systems are often launched from a great distance, and there may be no possibility for the operators to notice protective emblems at the point of impact; therefore, such weapon systems will need to have a way to detect protective emblems and react accordingly. In this regard, the present contribution suggests a cross-frequency protective emblem. Technical deployment is considered, as well as interpretation by methods of machine learning. Approaches are explored as to how software can recognize protective emblems under the influence of various boundary conditions. Since a new protective emblem could also be misused, methods of distribution are considered, including encryption and authentication of the received signal. Finally, ethical aspects are examined.

Keywords: Geneva Conventions, protective emblem, artificial intelligence, autonomous warfare, loitering weapons, drone warfare, non-combatants, war crimes.

Introduction

The red cross symbol was first recognized as an internationally accepted distinctive emblem for the protection of wounded military personnel in armed conflicts in the 1864 First Geneva Convention for the Amelioration of the Condition of the Wounded in Armies in the Field. The precise usage of this international emblem was clarified in Article 18 of Additional Protocol I to the four Geneva Conventions of 1949 (AP I).[1] In summary, it involves identification through a flag/patch that carries one of the four distinctive emblems, which may be visible with infrared (IR) devices. In addition to this passive representation, active light or radio signals, or electronic markings such as radar beacons, can be used to identify a protected facility or means of transportation. As a notable "digital" marking, the transmission of Global Positioning System (GPS) data from protected facilities to parties involved in the conflict is worth mentioning. While efforts have been made in recent decades to develop autonomous and fully

†This work has been funded by the German Federal Ministry of Defence. Niklas Widulle is funded by dtec.bw, the Digitalization and Technology Research Center of the Bundeswehr; dtec.bw is funded by NextGenerationEU. Maria Krantz is funded by the Federal Ministry for Economic Affairs and Climate Action through the ZIM programme.

The advice, opinions and statements contained in this article are those of the author/s and do not necessarily reflect the views of the ICRC. The ICRC does not necessarily represent or endorse the accuracy or reliability of any advice, opinion, statement or other information provided in this article.

[1] Protocol Additional (I) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 1125 UNTS 3, 8 June 1977 (entered into force 7 December 1978) (AP I). See also Jean S. Pictet, *Commentary on the Geneva Conventions of 12 August 1949*, Vol. 4: *Geneva Convention relative to the Protection of Civilian Persons in Time of War*, ICRC, Geneva, 1958

autonomous weapon systems for more precise and effective operations, the visibility of the protective emblem has not kept pace with technological advancements.

After cyber attacks on hospitals as part of "critical infrastructure",[2] discussions have arisen regarding the possibilities for protecting designated facilities (e.g. protected domains in cyberspace). However, the "visible" protective emblem and its protection have been overlooked, despite instances in recent history where designated facilities have been directly targeted for attacks, even with the establishment of protected zones through transmitted GPS data.[3] In addition to these events, the use of grenades dropped by a man-in-the-loop drone on an apparently injured soldier in the Ukrainian–Russian conflict has led to an imperative for the representation and respect of the protective emblem in the electromagnetic spectrum.

This article focuses on the most commonly used sensors and their modes of operation in autonomous weapon systems and derives a possible implementation of the protective distinctive emblem that is recognizable by these sensors. Another focus of the article is the potential perception of protective emblems and non-combatants by fully autonomous systems.

The need for a next-generation protective emblem

Long-range weapons are military projectiles or rockets that are launched from various delivery systems such as aircraft, watercraft, land vehicles or handheld tube weapon systems (e.g. the RGW 90 LRMP) at an unspecified distance from the target. The projectile or rocket is guided remotely or self-guided to the target. Large-calibre weapons (e.g. the Panzerhaubitze 2000) and rocket weapons (e.g. the MARS II Multiple Launch Rocket System) are classified under the term "artillery" and are typically land- or water-based weapon systems. Modern fully autonomous long-range or artillery weapon systems have a higher effectiveness in terms of impact and accuracy compared to "conventional" weapon systems. As such, they are designed to minimize or ideally prevent collateral damage to civilian infrastructure and non-combatants.

In the 1980s, the US military in particular recognized the advantage of precision-guided munitions, specifically in the form of smart munitions, for artillery systems. This type of ammunition has the ability to autonomously search for, identify and attack targets. The warheads are ejected as submunitions from a projectile casing, and a multitude of these submunitions engage multiple targets in a defined area using autonomous target-seeking capabilities (such as

[2] Tamara Gurschler, Sebastian Dännart and Ulrike Lechner, *Monitor IT-Sicherheit Kritischer Infrastrukturen*, 2017, pp. 170–180.

[3] Médecins Sans Frontières, "On 3 October 2015, US Airstrikes Destroyed Our Trauma Hospital in Kunduz, Afghanistan, Killing 42 People", 2015, available at: www.msf.org/kunduz-hospital-attack-depth (all internet references were accessed in January 2024); Physicians for Human Rights, "No Place Is Safe for Health Care: The Attack on Syria's Al-Atareb Hospital", 2021, available at: https://phr.org/issues/health-under-attack/attacks-in-syria/al-atareb-surgical-hospital-no-place-in-syria-is-safe-for-health-care/.

proximity-fused or guided munitions).[4] In the 1990s, this type of munition was referred to as artificial intelligence-based or autonomous munitions.[5]

During the 1990–91 Gulf War, approximately 90,000 tons of bombs were dropped by US aircraft on Iraq and Kuwait. Some 7% of these were air-to-ground precision-guided or long-range weapons (stand-off attack weapons), of which 90% hit their designated targets.[6] In contrast, only 25% of conventional air-to-ground weapons hit their targets. The decision cycle for target engagement with precision-guided or long-range weapon systems still remains under human control, meaning that at a defined point in time, a human consciously decides to engage a target (see loitering weapons). This decision cycle is divided into the decision points find, fix, track, target, engage, and assess (F2T2EA). In contrast, fully autonomous weapon systems undergo this decision cycle without further human control/decision-making after their activation.[7]

In consequence, warfare is increasingly being conducted at a distance using stand-off weapons without supplementary human visual on-site target designation or visual target verification, either fully autonomously or semi-autonomously. This can result in the (un)intentional engagement of facilities marked only by analogue-visible protective emblems such as red cross flags or colour-coded markings on buildings. Particularly in the case of fully autonomous weapon systems, the question arises as to how the protective emblem should be evaluated in the F2T2EA decision cycle.

Modern warheads of different systems already use different sensors to increase the accuracy of their hits. This often involves some form of system intelligence. However, it is questionable whether the warhead is able to distinguish between very similar vehicles or targets if one of them is protected by a painted, conventional, protective emblem (see Figure 1), when the same type of vehicle is also used for military and paramedical approaches (e.g. the GTK Boxer, FV432, TPz Fuchs, APC M113). If the conventional protective emblem is not recognized due to the low resolution of optical systems or is not perceived due to other influences (noise, weather conditions, partial occlusions etc.), the protection of these units is at risk. The conventional protective emblem can only be perceived in the visual spectrum, and there is no marking in other spectra (infrared, laser etc.). It is also questionable to what extent system intelligence is able to distinguish protected units from others. A distinction based only on predefined rules in a highly dynamic and variable scenario such as a battlefield is only feasible to a limited extent. Here, the use of AI could help to realize

[4] N. J. Mangus, R. B. Allen, J. C. Sentell, M. A. Cash and M. C. Zari, *Smart Munitions: An Introduction to the Concepts, the Technologies and the Systems – Primer and Briefing Manual*, GACIAC SR-87-08, GACIAC and IIT Research Institute, Chicago, IL, 1987.

[5] Maurice Zahnd, "Kampfwertgesteigerte Munition für Minenwerfer und Artillerie", *ASMZ: Sicherheit Schweiz: Allgemeine Schweizerische Militärzeitschrift*, Vol. 161, No. 10, 1995.

[6] Advisory Group for Aerospace Research and Development Neuilly-sur-Seine, *Precision Terminal Guidance for Munitions*, 1997, available at: https://apps.dtic.mil/sti/citations/ADA324120.

[7] Frank Sauer *Artificial Intelligence in the Armed Forces On the Need for Regulation Regarding Autonomy in Weapon Systems*, Security Policy Working Paper No. 26/2018, Federal Academy for Security Policy, 2018, available at: www.jstor.org/stable/pdf/resrep22189.pdf.

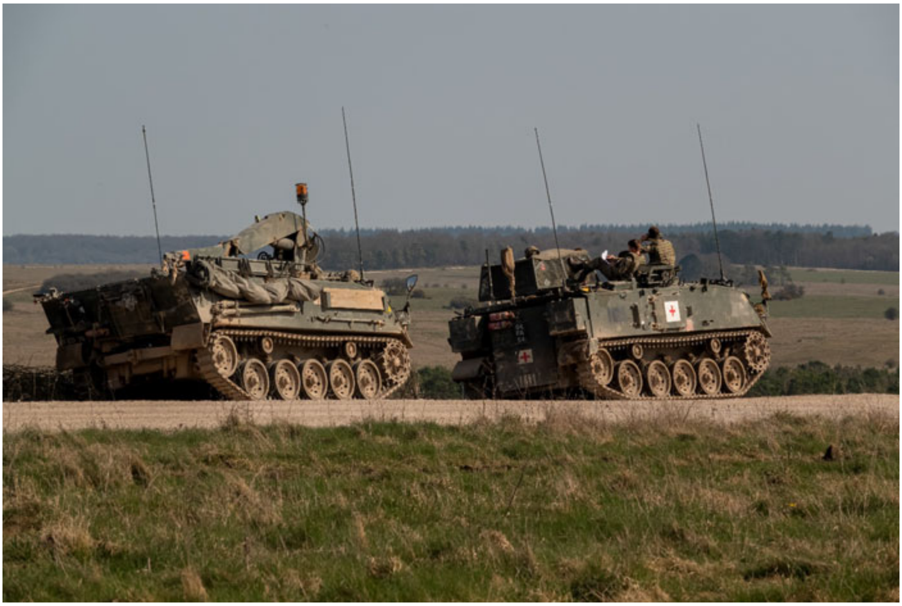


Figure 1. Depicted are two British FV434 tanks. The left one is used by the pioneers, the right one by the paramedics. They are difficult to distinguish, especially as the armoured medical vehicle only presents itself as an object to be protected by small, attached protective emblems. If there are no other protective emblems than the optical, conventional ones, the protective character may be difficult to determine for aggressors, especially if their optics do not recognize the emblem or other spectra than the optical one are used. It also remains questionable whether the intelligence currently used in drones and missiles is capable of distinguishing between these tanks. Source: Martin/stock.adobe.com.

generalized discrimination capabilities. The proposed next-generation protective emblem offers a comprehensive solution by leveraging passive and active signs, image recognition, machine learning and signal exchange (e.g. radar). This integrated approach enhances communication, security and compatibility, thereby strengthening the protection of entities in autonomous warfare scenarios.

Technical background

Weapon systems

Long-range weapon systems are of interest for the cross-frequency protective emblem because of two main aspects. On the one hand, those systems operate at a distance, which makes direct visual verification of the target hard or even impossible, in case of firing on given coordinates or if there is no imaging communication from the system back to the operator. On the other hand, those systems are equipped, normally, with different sensors such as radar, electro-optical (EO) sensors or thermal cameras. Those components can be used to detect and react to active or passive protective emblems in the spectrum.

Cruise missiles

Cruise missiles like the Roketsan Çakir[8] or the Raytheon BGM-109 Tomahawk[9] nowadays navigate with GPS/IPS or the Russian Global Navigation Satellite System (Global'naya Navigatsionnaya Sputnikovaya Sistema, GLONASS), also using inertial navigation as well as terrain contour matching (TERCOM) systems, though cheaper versions might only use a radar and barometer altimeter in combination with a clock. Automatic target recognition (ATR) can be used to identify targets without manual input and increases the accuracy of the missiles. ATR might use EO or thermal (IR) cameras, as well as, if possible, a signal-receiving sensor. In most cases those systems contain a bi-directional communication channel and intelligent algorithms to determine flight parameters and verify the environment. Cruise missiles are divided into the following categories: hypersonic (>Mach 5), supersonic (>Mach 1) and subsonic (<Mach 1) according to speed, as well as long-range (>1,000 km), medium-range (500–1,000 km) and short-range (50–500 km) according to flight range.

Drones

The presence of drones is rapidly increasing on battlefields around the world. These unmanned aerial vehicles (UAVs) are able to operate in a (fully) autonomous fashion or as a master–slave system where a human operator controls the UAV from a remote command centre. Different types of drones have to be distinguished, categorized into four major classes:

1. fixed-wing drones;
2. multi-rotor drones;
3. single-rotor drones; and
4. fixed-wing hybrid drones.

UAVs also differ in their size and loading capacity. Larger drones, like the Global Hawk RQ-4[10], General Atomics MQ-9 Reaper[11] or IAI Heron,[12] can fly more than 32 hours and carry, depending on their size, up to 1,360 kg. These UAVs fly at a height of up to 20 km. Smaller UAVs, like the Mavic 3, R18 or

[8] Roketsan, "Çakir Cruise Missile", available at: www.roketsan.com.tr/uploads/docs/kataloglar/ENG/1676879938_cakir-eng.pdf.

[9] CSIS Missile Defense Project, "Tomahawk", Center for Strategic and International Studies, 2023, available at: https://missilethreat.csis.org/missile/tomahawk/.

[10] US Air Force, "RQ-4 Global Hawk", 2014, available at: www.af.mil/About-Us/Fact-Sheets/Display/Article/104516/rq-4-global-hawk/.

[11] US Air Force, "MQ-9 Reaper", 2021, available at: www.af.mil/About-Us/Fact-Sheets/Display/Article/104470/mq-9-reaper/.

[12] Israel Aerospace Industries, "Heron: Multi-Role Male RPAS", 2021, available at: www.iai.co.il/sites/default/files/2023-05/HERON_0.pdf.

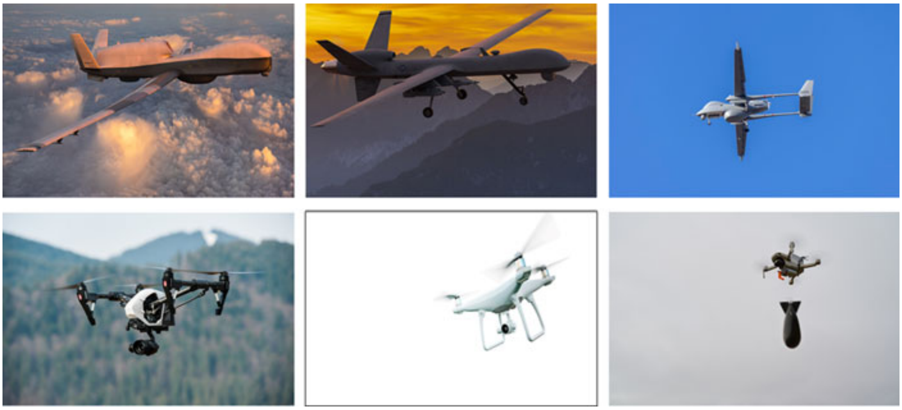


Figure 2. Examples of different UAVs. The upper row are larger drones, used for military applications, while the lower row are smaller, dual-use drones for civil and military applications. The bottom-right image shows a primary civilian drone that has been equipped with a bomb. Source: Mike_Mareen, gordzam, nesterenko_max, SFIO_CRACHO, sandsun/stock.adobe.com.

Demon, can fly up to 45 minutes, carry a maximum of 5 kg and fly up to a height of 6 km. See Figure 2 for example images. Different categories carry different gimbals and payloads, which will be specified below. These are the main features which are of interest for implementing a protective emblem that can be picked up by UAVs.

Gimbals designate the part of a UAV where (most of) the sensors are located. Usually, EO and IR cameras, as well as a laser, are stored in the housing, mounted beneath the UAV. Larger drones might carry radar. Most UAVs also use one or more global navigation satellite system services (Beidou/Compass, Galileo, GLONASS or GPS) for navigation.

The payload can be used to carry additional reconnaissance technology like radar, EO and IR cameras, and laser rangefinders, as well as armaments like missiles or bombs. Those armaments might also be equipped with sensors, but usually only when mounted to a larger UAV, whereas smaller UAVs might carry grenades that do not have any kind of sensors. When drones are used as kamikaze-style aircraft, they might carry some sort of explosive that is not to be dropped separately.

Artillery

Artillery uses precision-guided munitions nowadays, which differs in many different ways form the older, better-known standard munitions without any sensors or intelligence on board. These new shells have various sensors for steering and aiming to increase the accuracy. The M982 Excalibur, for example, uses GPS to navigate from the launch platform to the target.[13] Munitions like the 2K25 Kransopol or M712 Copperhead are laser-guided. For this purpose, the target

[13] Anthony Williams and Jayesh Dhingra (eds), *Jane's Weapons: Ammunition (2019–2020)*, IHS Markit, Coulsdon, 2019.

must be illuminated with a laser by a forward observer during the aiming approach. The laser diode in the head of the projectile receives this coded radiation and aligns the target approach accordingly. EO or IR cameras are also used during the target approach. Munitions like the Strix are able to find their targets autonomously using the 3–5 μm infrared spectrum to distinguish different targets,[14] whereas the AGM-62 Walleye uses a TV camera in its head. Radar is also used to navigate, and TERCOM is used to compare the surroundings with a stored route and thus keep to the intended flight path.[15] When approaching the target, this map of the surroundings is particularly detailed and helps the projectile to find its target. Smart ammunition, like the SMArt 155,[16] a fire-and-forget artillery projectile, combines different sensors to autonomously find and destroy its target. The SMArt 155 uses IR, millimetre-wave radar and millimetre-wave radiometer sensors to analyse the battlefield in a radius of around 150 metres and independently initiates target detection and engagement. No confirmation by a human operator is required. The SMArt 155 uses three independent criteria, two of which must be met in order to attack a target: these are the shape, the material properties and the temperature of the object being targeted.

(Fully) Autonomous systems

According to the SAE On-Road Automated Driving Committee,[17] there are six levels of automation, ranging from 0 to 5. The first three levels involve human operators, while the last two levels are crucial for autonomous systems. At level 4 (high automation), a system can operate within specific conditions and adjust mission parameters, but human intervention is still possible. In contrast, at level 5 (full automation), the system operates independently without requiring human interaction. It possesses the capability to make autonomous decisions, unrestricted by mission or environmental constraints. This level of autonomy qualifies the system as truly autonomous, but it is important to note that level 4 systems relying on communication with external entities may be referred to as automatic systems.

Signal-based options for protection

Different options for transferring signals exist. Table 1 summarizes the ranges of different signal emitters.

[14] Richard D. Jones and Leland Ness (eds), *Jane's Infantry Weapons 2011–2012*, Jane's Information Group, 2011.

[15] Joe Golden, "Terrain Contour Matching (TERCOM): A Cruise Missile Guidance Aid", *SPIE Proceedings*, Vol. 238: *Image Processing for Missile Guidance*, 1980.

[16] General Dynamics, "SMArt 155: 155mm Sensor Fuzed Munition (SFM) for the Artillery", available at: www.gd-ots.com/wp-content/uploads/2017/11/SMArt155.pdf.

[17] SAE On-Road Automated Driving Committee, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE International, 2021.

Table 1. *Signals and technologies for active and passive communication and recognition*

Technology	Signal frequency	Range	Advantages	Disadvantages
Radio waves* e.g. Automatic Identification System (AIS)	30 Hz–300 MHz 156.025 MHz–162.025 MHz	≈10–1,000 km	Good for long-range communication and sensing, used for radio and TV, AIS is already used in the military to communicate between ships and land	Limited data rate
Microwave,* radar	1–100 GHz	≈1–100 km	Good for long-range communication and sensing, can penetrate some solid objects	Limited penetration of dense objects such as walls or buildings, large antenna size, affected by environmental disturbances
L-band* e.g. Identification Friend or Foe (IFF)	1–2 GHz 1–8 GHz	≈1,000 km	Good penetration, low attenuation, suitable for long-range communication and sensing, IFF is already used in the military to identify airplanes, ships and vehicles	Lower data rates compared to higher frequency bands
X-band*	8–12 GHz	≈100 km	High data rates, good for long-range communication and sensing, better resistance to interference	Limited penetration of solid objects such as buildings or trees, can be affected by atmospheric conditions

Continued

Table 1.
Continued

Technology	Signal frequency	Range	Advantages	Disadvantages
Infrared	1–100 THz	≈100 m	Can detect heat signatures, good for night vision, passive detection	Limited range, susceptible to interference from other heat sources
Optical	430–750 THz	≈kilometres	High resolution, good for visual imaging, active or passive sensing	Limited range in adverse weather and at night, requires line-of-sight
Thermal camera	9–14 μm	≈kilometres	Can detect heat signatures, good for night vision, can penetrate some materials	Limited resolution, susceptible to interference from other heat sources
Radio-frequency identification (RFID)	kHz GHz	≈metres ≈kilometres	Low cost, small form factor, suitable for inventory tracking and asset management	Limited range, susceptibility to interference from other RFID devices
Wi-Fi	2.4 GHz, 5 GHz	≈100 m	High data rates, widespread availability, suitable for local area network communication	Limited range, susceptible to interference from physical obstructions and other Wi-Fi devices
Passive signs, symbols	N/A	≈100 m	Can be used for detection and recognition of predefined signs or symbols	Limited to predefined signs, susceptible to false positives or negatives
Night vision devices	N/A	≈100m	Can amplify low levels of light	Limited range, susceptible to interference from bright light sources and atmosphere

*Communication via satellite possible, bands used for radar.

In the military, two identification systems are already implemented in different frequency ranges. Identification Friend or Foe (IFF),[18] which operates in the L-band (1–2 GHz) and sometimes the S-band (2–4 GHz), is an essential technology used by the military to distinguish between friendly and hostile aircraft, ships or vehicles over distances of up to approximately 1,000 km. IFF systems use modulated signals to transmit identification information, ensuring that friendly units are recognized correctly. Automatic Identification System (AIS)[19] devices operate between 156.025 MHz and 162.025 MHz, and are widely utilized in the maritime domain to facilitate communication between ships and land-based stations. These signals can provide a baseline for the development of the digital protection emblem, to embed, for example, protective information into these established signals, creating a distinctive red cross pattern. The selection of different frequencies for the protective emblem will guarantee that the already established frequencies, such as those used in IFF or AIS, remain unaffected.

Radio-frequency identification (RFID)[20] uses radio waves to identify and track objects. The technology is already used in a military setting, for example to track the location and status of military equipment[21] and to control access to important military areas such as control centres or locations of weapon systems. It consists of three basic components: a reader, an antenna and a tag.

The reader, the purpose of which is to identify objects, emits radio waves in a certain frequency range, typically 125kHz, 13.56MHz or 900MHz.[22] The tag, which is positioned on the object to be tracked or identified, contains the antenna and a microchip. On this microchip, a unique identification number and other data can be stored. Once a signal from the reader is captured by the tag, it sends back the information on the chip to the reader on the same frequency, which can process the information using a computer system.

While passive RFID tags do not have a power source and need the energy from the reader's signal to transmit their data, active RFID systems contain an energy source and can therefore actively transmit data.[23] Active RFID systems also have a longer range and can transmit their signal over a range of a few metres to hundreds of metres, or even several kilometres. Active RFID systems

[18] Sviatoslav Starokozhev, Ivan Shevtsov, Oleksandr Datsenko, Valeriia Chumak and Anton Sierikov, "Signal Provision of Address Systems Identification Friend or Foe", *Proceedings of the 2022 IEEE 9th International Conference on Problems of Infocommunications, Science and Technology*, Kharkiv, 2022.

[19] Abbas Harati-Mokhtari, Alan Wall, Philip Brooks and Jin Wang, "Automatic Identification System (AIS): Data Reliability and Human Error Implications", *Journal of Navigation*, Vol. 60, No. 3, 2007.

[20] Qinghan Xiao, Cam Boulet and Thomas Gibbons, "RFID Security Issues in Military Supply Chains", *Proceedings of the Second International Conference on Availability, Reliability and Security (ARES'07)*, IEEE, 2007.

[21] Mark Buckner, Richard Crutcher, Michael R. Moore and Bobby Whitus, "MICLOG RFID Tag Program Enables Total Asset Visibility", *MILCOM 2002 Proceedings*, Vol. 2, 2002.

[22] Q. Xiao, C. Boulet and T. Gibbons, above note 20.

[23] Ricardo Tesoriero, Jose A. Gallud, Manuel Lozano and Victor M. Ruiz Penichet, "Using Active and Passive RFID Technology to Support Indoor Location-Aware Systems", *IEEE Transactions on Consumer Electronics*, Vol. 54, No. 2, 2008.

can also be used for real-time location tracking.[24] However, the range of an RFID signal is heavily dependent on its environment. Radio signals in the surrounding area, like cellular or Wi-Fi signals, can interfere with the RFID signal. Furthermore, physical barriers like walls can also block the signal.[25]

Once a signal is received by the reader, it first converts it from analogue to digital using an analogue-to-digital converter, before extracting the information contained in the signal. Since the signal could be distorted by interference, error correction algorithms are used; thereby, RFID systems can improve the reliability and accuracy of the data being transmitted. Possible algorithms are a cyclic redundancy check,[26] which adds a checksum to the transmitted data, or a forward error correction,[27] which adds redundant information.

A problem with detection of RFID tags arises when whole units are equipped with such tags. Especially when multiple tags are present within the range of a single reader, anti-collision protocols become necessary. Without an anti-collision protocol, the reader might not be able to distinguish between multiple tags and might read them all simultaneously or fail to read any of them. Several different anti-collision protocols can be used with RFID systems, such as the ALOHA protocol, tree-based protocols, binary search algorithms or bitwise arbitration algorithms.[28] The specific protocol used will depend on the application and the requirements of the system. The goal of these protocols is to ensure that each tag is identified in a timely and efficient way, while minimizing collisions and other sources of interference.

Machine learning (ML) approaches have been suggested to improve collision-free reading of RFID tags. One proposed solution is an anti-collision protocol called DMLAR that uses feed-forward artificial neural network methodology to predict collisions and ensure efficient resource allocation in RFID networks.[29] Such an approach might be necessary when a multitude of tags need to be read to identify armed military personnel. In situations where many RFID tags need to be read simultaneously, ML algorithms can also be applied to process and analyze the data. ML approaches have been applied to a variety of RFID data.[30]

[24] Daqiang Zhang *et al.*, "Real-Time Locating Systems Using Active RFID for Internet of Things", *IEEE Systems Journal*, Vol. 10, No. 3, 2014.

[25] Xin Li, Yimin Zhang and Moeness G. Amin, "Multifrequency-Based Range Estimation of RFID Tags", *Proceedings of the IEEE International Conference on RFID*, 2009.

[26] Yan-Fei Li, Hong-Jun Wang and Hua Li, "A RFID Algorithm Based on Cyclic Redundancy Check", *Proceedings of the 3rd International Conference on Anti-counterfeiting, Security and Identification in Communication*, IEEE, 2009.

[27] Andreas Schantin, "Forward Error Correction in Long-Range RFID Systems", *Proceedings of Smart SysTech 2012: European Conference on Smart Objects, Systems and Technologies*, IEEE, 2012.

[28] Dheeraj K. Klair, Kwan-Wu Chin and Raad Raad, "A Survey and Tutorial of RFID Anti-Collision Protocols", *IEEE Communications Surveys and Tutorials*, Vol. 12, No. 3, 2010.

[29] Rachid Mafamane, Mourad Ouadou, Hajar Sahbani, Nisrine Ibadah and Khalid Minaoui, "DMLAR: Distributed Machine Learning-Based Anti-Collision Algorithm for RFID Readers in the Internet of Things", *Computers*, Vol. 11, No. 7, 2022.

[30] Osama Mohsen, Yasser Mohamed and Mohamed Al-Hussein, "A Machine Learning Approach to Predict Production Time Using Real-Time RFID Data in Industrialized Building Construction", *Advanced Engineering Informatics*, Vol. 52, 2022; Xiaohui Tao, Thanveer Basha Shaik, Niall Higgins, Ray

Radar is a technology that uses radio waves to detect the presence, location and velocity of objects. It works by emitting a radio wave signal which travels through the air and reflects off an object in its path. The reflected signal, also called an echo, is then detected by a receiver and analyzed to determine the distance, angle and speed of the object.[31]

In terms of detection, there are several factors that can affect a radar system's ability to detect objects. These include the range of the radar, the power and frequency of the signal, and the size and reflectivity of the object being detected. Additionally, factors such as weather, interference from other sources and the presence of other objects in the signal path can also affect detection.

A radar signal can also be emitted by a stationary sender such as a coastal radar station. These often have fixed antennas to cover a given area. Radar can also be employed to transmit information to the receiver; this is known as radar communication or radar data communication.[32] Radar communication is robust against interference, and would be the use envisioned for our present purposes. Since many facilities are equipped with radar emitters, many radar signals will be received by the weapon system simultaneously. It is therefore important to ensure that the signals emitted by a protected facility are given priority. This could be hard-coded into the code – e.g., once a signal associated with a protected facility is received, it overrides all others. Better detection and discrimination of radar signals could also be achieved through the use of artificial intelligence (AI).[33]

GPS is a satellite-based navigation system that provides location and time information anywhere on or near the Earth. The GPS system consists of a network of satellites orbiting the Earth, as well as ground control stations and GPS receivers. Each satellite transmits a signal that contains information about its location and time.[34] To be able to use GPS, a weapon system must be able to receive signals from at least four GPS satellites. It then compares the time stamps of the signals received from the satellites and uses the differences in the time stamps to calculate the distance between itself and the satellites. Through trilateration, the system can then determine its own position.

In addition to the choice of technology and frequency band, signal distribution is an important consideration for communication and sensing

Gururajan and Xujuan Zhou, "Remote Patient Monitoring Using Radio Frequency Identification (RFID) Technology and Machine Learning for Early Detection of Suicidal Behaviour in Mental Health Facilities", *Sensors*, Vol. 21, No. 3, 2021; Yosuke Senta, Yoshihiko Kimuro, Syuhei Takarabe and Tsutomu Hasegawa, "Machine Learning Approach to Self-Localization of Mobile Robots Using RFID Tag", *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2007.

[31] Albrecht K. Ludloff, *Praxiswissen Radar und Radarsignalverarbeitung*, Vieweg+Teubner Verlag, Wiesbaden, 2013.

[32] Siji Quan, Weiping Qian, Junhai Guq and Van Zhang, "Radar-Communication Integration: An Overview", *Proceedings of the 7th IEEE/International Conference on Advanced Infocomm Technology*, 2014.

[33] Gerard T. Capraro, Alfonso Farina, Hugh Griffiths and Michael C. Wicks, "Knowledge-Based Radar Signal and Data Processing: A Tutorial Review", *IEEE Signal Processing Magazine*, Vol. 23, No. 1, 2006.

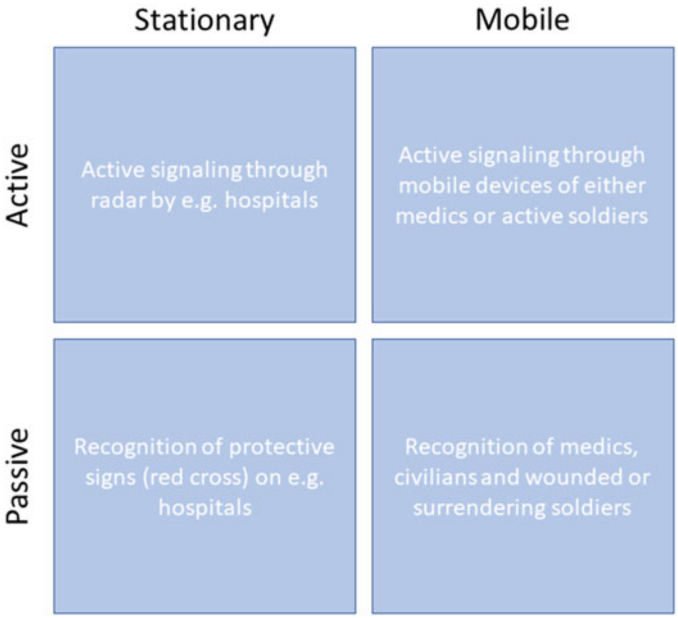
[34] Christopher J. Hegarty, "The Global Positioning System (GPS)", in Peter J. G. Teunissen and Oliver Montenbruck (eds), *Springer Handbook of Global Navigation Satellite Systems*, Springer, Cham, 2017.

systems. Bidirectional signal distribution allows for two-way communication, with signals being transmitted and received between the sender and receiver.

Signal distribution should incorporate encryption and authentication mechanisms to mitigate unauthorized access and interference; this will be discussed in more detail below.

The next-generation protective emblem

A next-generation protective emblem should protect every non-combatant according to the Geneva Conventions. While the conventional protective emblem has its limitations in various ways in modern warfare as described above, the next-generation protective emblem has to deal with cross-frequency visibility. This visibility on a broad spectrum could inform various systems about protected entities. We distinguish between active and passive recognition as shown in Figure 3. Active recognition covers those cases where the protected entity is able to send out a signal into the spectrum, in order to draw the attention of attacking systems and actively inform those systems about the entity's protected status. Passive recognition applies to all non-combatants who are not able to send out



	Stationary	Mobile
Active	Active signaling through radar by e.g. hospitals	Active signaling through mobile devices of either medics or active soldiers
Passive	Recognition of protective signs (red cross) on e.g. hospitals	Recognition of medics, civilians and wounded or surrendering soldiers

Figure 3. The recognition of a protective emblem can be divided into active and passive recognition. Active recognition here refers to the active sending of a signal which marks an entity as being protected, e.g. by sending out a radar signal. Passive protective means a right for protective needs to be recognized by the attacking system, e.g. the conventional red cross emblem or wounded soldiers. Both stationary and mobile entities can be protected.

an active signal and therefore have to be identified in a passive way – e.g., using image recognition of the conventional protective emblem or the status of a non-combatant. Attacking systems should comprehend the status of the entity by themselves and should react in a corresponding way.

Active protective emblem

Active detection is always possible when the protected entity actively emits a signal, for example using radar, RFID or GPS. This is possible for stationary facilities (hospitals, civilian buildings) but could also be installed on tents used to tend wounded soldiers and to a certain extent on mobile units like tanks. Active detection of mobile troops would likewise be possible, as will be discussed below.

Detection of stationary facilities

Stationary facilities are the easiest to protect when it comes to the emission of active signals, since emitters can be installed on-site. To ensure safe detection of these facilities, a fusion of different signals would be the best option to allow for far (hundreds of kilometres) and near (metres to kilometres) communication range and to provide redundancy in case of impaired communication capability caused by adverse weather conditions or limited visibility. Therefore, the facility would have to be equipped with several emitters. Table 1 above summarizes the ranges of different emitters.

Autonomous weapon systems are often launched from long distances, making it challenging for operators to notice protective passive emblems, such as the red cross, at the point of impact. Therefore, it is crucial for these systems to autonomously detect protective active emblems in a certain range and either discontinue their operation or request an abort through human intervention. For the design of the next-generation active emblem, we propose to use a combination of low-frequency and high-frequency signals (e.g. electromagnetic waves), which allows for a wide communication range spanning from metres to hundreds of kilometres between weapon and target.

The proposed next-generation protective emblem design incorporates radar beacons or alternative methods, as listed in Table 1, to effectively target diverse sensors and transmit robust signals over long distances. The utilization of radar or other forms of electromagnetic radiation enables communication at the speed of light, facilitating the exchange of information up to 30 seconds (depending on the distance and communication range) before the arrival of a missile. This provides an opportunity to either abort the attack or potentially evacuate buildings, thereby safeguarding human lives.

The utilization of low- and high-frequency bands or diverse technologies enhances the robustness of communication. This approach ensures increased resilience, particularly in the face of challenging environmental conditions such as inclement weather or limited visibility. It is crucial to note that the values presented in Table 1 provide approximate ranges, as they are subject to various factors including atmospheric conditions, antenna gain and transmitter power, which can significantly affect the actual communication range in real-world scenarios.

Safe detection of a protected facility can be achieved by the combination of several types of signals. This would, for example, be possible by combining radar, GPS and RFID. A protected facility could send out a radar signal, transmitting information about its position and a secure code for the identification of protected facilities (allocation of these codes is discussed later in the article). Once this is received by a weapon system, it could use GPS to determine its own position relative to that of the emitting facility. It could then send a request to a database to check whether a protected facility has been assigned to the relevant position and whether the received code is authentic. Once this is checked, the weapon system could send its own signal to the facility, based on RFID technology, and wait for the signal to be sent back by the facility's RFID tag. Once it has received this signal, it would have to stop its attack and safely manoeuvre out of the protected zone. Figure 4 gives an overview of the proposed method for protecting stationary facilities.

A problem arises when one of these signals is not available – for example, when it is not possible to send a request to the database. It could be possible to add a human-in-the-loop, who could intervene in unclear cases. Furthermore, it would be possible to always require a human to be involved in the decision to abort an attack when a protective emblem is received.

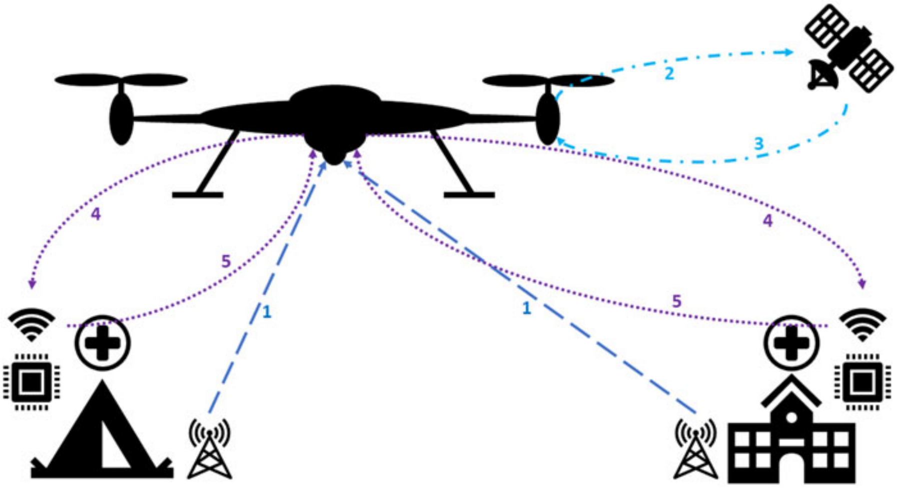


Figure 4. Possible scenario for recognition of stationary protected facilities. First (1), the weapon system receives a radar signal, which is constantly sent out by the protected entity. The weapon system then checks with a database to determine whether a known protected entity at this location is on record (2, 3). Next, the weapon system sends out a signal to the RFID system in the protected entity (4) and waits for an answer (5). Once this answer is received, the entity has been safely identified as protected and the attack is abandoned.

The above approach would necessitate the installation of several sensors on a weapon system, as well as the ability to safely communicate with a database. For new weapon systems, these features could be demanded by international standards. Older systems would have to be retrofitted, or in cases where this is not possible, decommissioned.

Detection of mobile units

Detecting moving protected entities, like wounded or surrendering soldiers or mobile paramedic units, can be problematic. Larger mobile units like tanks could still be equipped with radar and RFID emitters, but the approach described above would not be applicable here since they are not stationary and can therefore not be assigned a position in a GPS database. Smaller mobile units like paramedics cannot be equipped with a radar signal and would therefore not be detected by the system described in the previous section.

A possible option would be to equip all weapons with (active) RFID tags which could be read by an attacking weapon system. Units not carrying an RFID tag would be deemed protected and would not be attacked. RFID technology is already used to keep track of military equipment. RFID tags can be attached to vehicles, weapons and other equipment to track their location and status. This can help military commanders to quickly locate and deploy assets, as well as monitor the maintenance and repair status of equipment.[35] RFID tags could also be attached to personnel to track their location and movements, as well as to monitor their health and well-being.[36]

The received data would have to be processed and analyzed. It would also be possible to use analysis methods to detect protected entities even when they are marked with an RFID tag (for example, surrendering or retreating soldiers). It is possible to use data collected from RFID tags to train ML algorithms for anomaly detection. One study aimed to develop a system that detects abnormal behaviour in elderly people at home using active RFID tags:[37] movement data was collected through the RFID reader's signals, clustering techniques were used to build a personalized model of normal behaviour, and any incoming data outside the model was viewed as abnormal and triggered an alarm. Similarly, algorithms could be trained to recognize the movement patterns of active soldiers as compared to injured, surrendering or retreating soldiers, based on the data received from RFID tags.

The feasibility of this approach to recognizing protected individuals depends on the willingness of all nations to participate in such an approach. Two major points need to be addressed:

[35] M. Buckner *et al.*, above note 21.

[36] Rob Nicholls, "Implanting Military RFID: Rights and Wrongs", *IEEE Technology and Society Magazine*, Vol. 36, No. 1, 2017.

[37] Hui-Huang Hsu and Chien-Chen Chen, "RFID-Based Human Behavior Modeling and Anomaly Detection for Elderly Care", *Mobile Information Systems*, Vol. 6, No. 4, 2010.

1. All weapon manufactures that produce handheld, portable weapon systems would have to equip their weapons with RFID tags working on an internationally assigned frequency.
2. All existing handheld, portable weapons would have to be retrofitted with RFID tags operating on the internationally assigned frequency.

Passive protective emblem

According to the Geneva Conventions, every non-combatant is to be protected from any act of violence and war. As most non-combatants do not carry any active signal or a conventional protective emblem, they have to be identified in a different, passive way. A (fully) autonomous weapon system has to recognize these non-combatants by itself and if possible, inform the operator about the recognition.

Figure 5 shows example situations and communication between a drone and an operator. Medic units who wear an emblem (flag or patch) are to be protected during active service in the field; furthermore, a protective emblem can be derived from the situation in which a person finds him/herself.

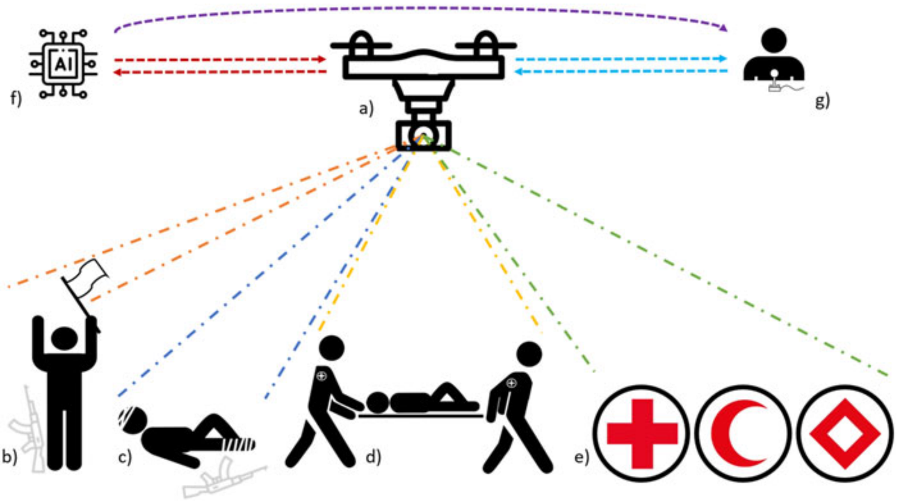


Figure 5. Example scenarios for passive protective indicators that a drone should be able to detect and react to. (a) shows a drone equipped with a daylight camera and optionally a thermal camera. (b) to (d) show examples of non-combatants or situations where attack is prohibited by the Geneva Conventions – a surrendering person (b), a wounded person (c) and paramedics carrying a sick person (d). (e) shows the internationally recognized protective emblems, which have to be detected in various situations and should guarantee that the bearer will not be attacked. (f) and (g) are possible operators on the drone. A fully autonomous drone, controlled by AI, should automatically abort any sort of attack when recognizing a protective emblem ((b) to (e)), or, in case of semi-automation, inform the human operator (g), who might then abort the action.

The passive protective emblem includes the condition and status of a person. This comprises people who are incapacitated, surrendering or laying down their arms. This means that people are protected in all situations unless they are combatants. Therefore, the passive protective emblem should serve not only for the recognition of non-combatants but rather the detection of combatants and the exclusion of all other people "on the field". These situation-based passive protective emblems, as well as the conventional protective emblem, have to be determined and evaluated on the basis of image and video material recorded by EO or IR sensors, for example from drones or the warheads of missiles. Figure 6 shows some examples of images captured by drones just before the dropping of a payload on (wounded) soldiers during recent conflicts.

The current use of the protective emblem does not, in respect to new developments in weapons technology, cover all the conditions mentioned above. The protective emblem therefore has to be enhanced and developed further to ensure that the protection of non-combatants is still guaranteed as best as possible under the given circumstances. To this end, the protective emblem could make use of modern sensor technologies and take the step from the marking of protected entities to the recognition of those entities, hence shifting the responsibility from the non-combatants to the attackers. In the following sections we will describe how this could be done for (fully) autonomous systems using sensors to recognize passive protective emblems.

EO/IR-based recognition using AI

Most modern weapon systems, especially drones, are equipped with different kinds of cameras and some sort of intelligence. This so-called intelligence encompasses a wide field of different definitions used by a broad field of actors. Some are based on threshold values, empiric algorithms or AI. (Fully) Autonomous systems operate on their own, without the need for communication; therefore, every decision is made on the platform itself. Systems needing communication with an operator can still be equipped with AI-based services, like object detection or signals intelligence evaluation, to assist the operator. The combination of existing sensors and intelligence or services can be used to build in some routine to evaluate the war scene and respect the rules of engagement and the Geneva Conventions. Based on imagery intelligence techniques, an AI model can be created which derives the context from the scene and enables a system to cancel and stop unethical actions. Therefore, in a first step, rules need to be formalized, data needs to be collected, and both need to be fitted to an AI model.

A main aspect of the passive protective emblem is the definition of situations when a person has to be recognized as a protected individual. Therefore, rules have to be created to distinguish such situations from those where combatants might falsely appear to be in a similar situation, such as when they are hiding, sneaking, ambushing etc. These rules can then be used in a context-informed neural network. However, these rules can only represent a part of the passive protective emblem. They transfer the rules of the Geneva

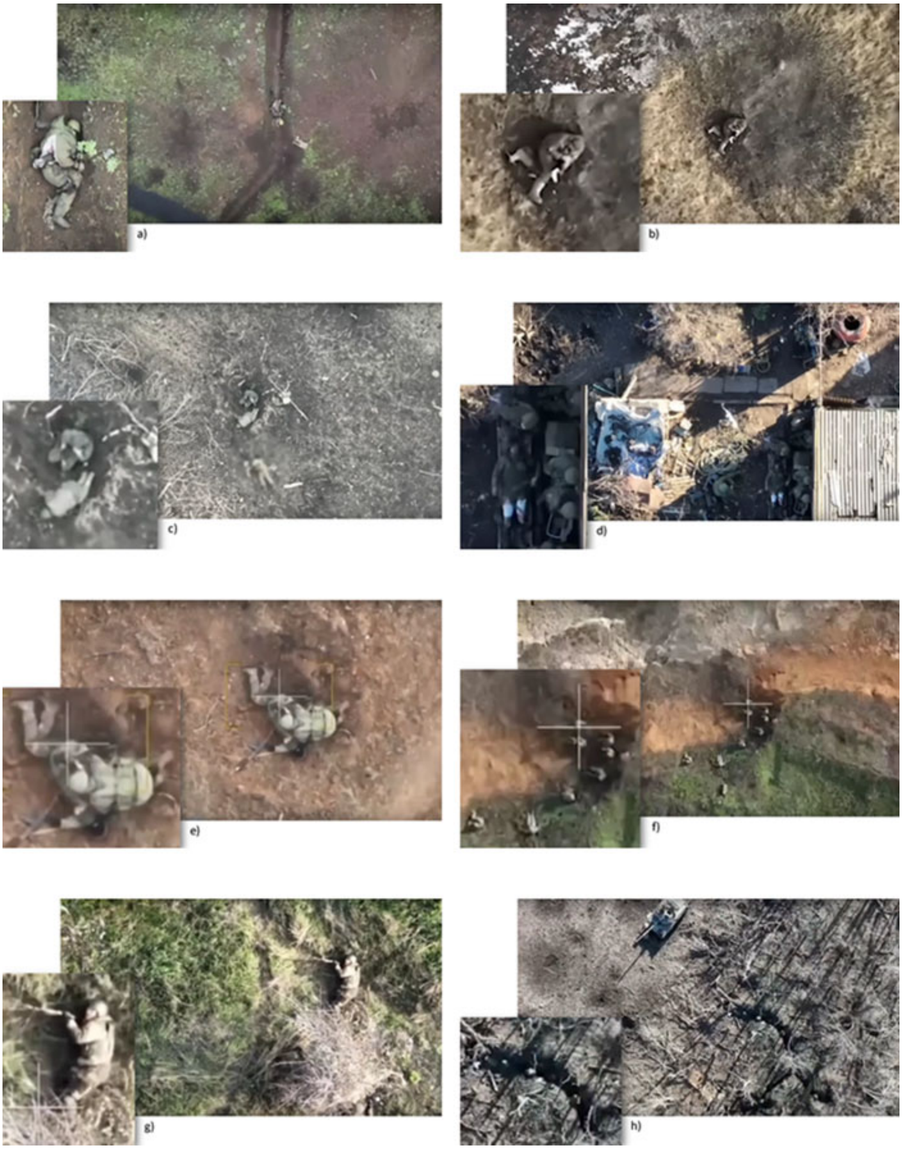


Figure 6. Images of soldiers in different situations in the field taken from drones. Images (a) to (e) show different non-combatants, whereas (f) to (h) show combatants. (a), (b) and (c) depict wounded soldiers, (d) depicts two paramedics and a solider on a stretcher, and (e) depicts one solider providing first aid to a companion. Images (f) and (h) show a group of soldiers hiding in an ambush, whereas (g) shows a solider lurking. All soldiers are about to be attacked by payloads dropped from drones. In cases (a) to (e), this means a violation of the Geneva Conventions. Source: youtube.com.

Conventions into the AI model and should be kept as basic values, but many protections are determined from the context of the situation, and these are harder to formalize. Therefore, in addition to the rules, a large dataset is needed to abstract and generalize knowledge onto an AI model.

To ensure a wide applicability, one should use data from various platforms, including both EO and IR sensors. The homogeneous distribution of all scenarios which show combatants and non-combatants in different contexts should also be ensured, as well as an equal share of day and night vision, so that the algorithm can be used twenty-four hours a day.[38] Data can be taken from various sources, like videos from the internet, video games or movies.[39]

To create a comprehensive dataset covering diverse scenarios, attention should be paid to a wide variation of the recorded circumstances and at the same time to a good augmentation of the data. The different scenarios include, for example, different camera resolutions, environmental conditions such as weather, geographical location, flight altitude during the recording, and possible camouflage patterns, as well as the states of the persons to be recorded. While videos from recent conflicts are a benchmark for the realistic creation of further training data, thanks to synthetic data it is also possible to train for crisis areas and situations that are not covered by real data. Taking care of well-distributed data will lead to a smaller bias in the data,[40] and therefore to a more objective AI service.

A major aspect while collecting data for the training of a passive protective emblem is the risk of a data poisoning attack, a subcategory of adversarial attacks.[41] Attackers on the AI system inject corrupted or manipulated data into the training data set in order to gain advantages for one party or disadvantages for opposing parties. In this way, so-called "backdoors" would be created in the passive protective emblem, which would cause gaps in the protective effect. It must therefore be ensured that the data set is not contaminated.

The method suggested above has been derived from already applied AI systems. In recent studies, automatic traffic sign recognition systems were explored as a means of improving road safety. In these systems, cameras are mounted on the vehicle to capture video feeds of the road and recognize traffic signs, providing the driver with timely warnings, nowadays even in real time, based on an embedded platform that employs digital image processing algorithms.[42] Convolutional neural networks have shown promising results in

[38] Bartosz Krawczyk, "Learning from Imbalanced Data: Open Challenges and Future Directions", *Progress in Artificial Intelligence*, Vol. 5, No. 4, 2016.

[39] Stephan R. Richter, Vibhav Vineet, Stefan Roth and Vladlen Koltun, "Playing for Data: Ground Truth from Computer Games", *Proceedings of Computer Vision – ECCV 2016: 14th European Conference*, Part II, Springer, Cham, 2016.

[40] Antonio Torralba and Alexei A. Efros, "Unbiased Look at Dataset Bias", *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, 2011.

[41] Jacob Steinhardt, Wei W. Koh Pang and Percy S. Liang, "Certified Defenses for Data Poisoning Attacks", *Advances in Neural Information Processing Systems 30: Proceedings of the 31st Annual Conference on Neural Information Processing Systems*, 2017.

[42] Enis Bilgin and Stefan Robila, "Road Sign Recognition System on Raspberry Pi", *Proceedings of the IEEE Long Island Systems, Applications and Technology Conference*, 2016; M. Sridevi, N. Sankaranarayanan, Ankit Jyothish, Aditya Vats and Milind Lalwani, "Automatic Traffic Sign Recognition System Using

improving the efficiency and robustness of these techniques.[43] A similar approach should be taken for training an AI for the recognition of protected individuals.

The detection of the passive protective emblem, both day and night, requires the use of EO and IR data. In order to gain information from the combination of the two sensors, a network that can handle the fusion of these two sensors should be used. Since models for image recognition and classification, or pose estimation (see below), may not be able to perform sensor fusion, a state-of-the art model can be trained by transfer learning[44] and empowered to perform sensor fusion through additional layers. Since time can be a critical factor and a very generative and abstract model is needed to capture complex conditions of the real environment, sub-symbolic models are particularly suitable here.[45] A symbolic approach would not be useful due to the high complexity of the modelling and the time required for calculation.

Pose estimation[46] is a technique in ML where the actual state of humans is classified by analyzing their posture and sometimes their facial expressions. This widely researched field could be adapted to a new dimension, where it could be used to classify the state of combatants and non-combatants. Therefore, new postures can be added to the models which can be transfer-learned on the data to be collected. This data could include emerging, fighting, lurking, uninvolved, injured and other states of persons, as well as the representation of the classic protective emblem (see Figure 6). The algorithm would then learn to distinguish combatants from non-combatants based on postures and conditions, such as injuries or body temperature. The differentiation can be challenging, but it is important to ensuring the success of the next-generation protective emblem. In this regard, the integration of both sensors, as well as context information (the course of the war, position etc.), is of great importance.

An ML model has to be hardened against adversarial attacks, in order to prevent deliberate manipulation of an input image/video feed. In this way, a better distinction can be made between non-combatants and combatants who pretend to be such by impersonating the status of an undefended person, e.g. by pretending to be injured. The network for distinguishing between combatants and non-combatants can be preceded by a second network that generates manipulated images and tries to convince the actual model that the content being

Fast Normalized Cross Correlation and Parallel Processing", *Proceedings of the International Conference on Intelligent Communication and Computational Techniques*, IEEE, 2017.

[43] Seokwoo Jung, Unghui Lee, Ji-Won Jung and David Hyunchul Shim, "Real-Time Traffic Sign Recognition System with Deep Convolutional Neural Network", *Proceedings of the 13th International Conference on Ubiquitous Robots and Ambient Intelligence*, IEEE, 2016; Yuga Hatolkar, Poorva Agarwal and Seema Patil, "A Survey on Road Traffic Sign Recognition System Using Convolution Neural Network", *International Journal of Current Engineering and Technology*, Vol. 8, No. 1, 2008.

[44] Transfer learning is an ML technique whereby a previously trained model is re-used to work on a different (but related) task.

[45] On sub-symbolic AI, see Orhan G. Yalçın, "Symbolic vs. Subsymbolic AI Paradigms for AI Explainability", *Towards Data Science*, 21 June 2021, available at: https://towardsdatascience.com/symbolic-vs-subsymbolic-ai-paradigms-for-ai-explainability-6e3982c6948a.

[46] Qi Dang, Jianqin Yin, Bin Wang and Wenqing Zheng, "Deep Learning Based 2D Human Pose Estimation: A Survey", *Tsinghua Science and Technology*, Vol. 24, No. 6, 2019.

shown belongs to a different class than is actually correct. By learning to distinguish between ambiguous images, the first network is ultimately less susceptible to adversarial attacks.

It should be kept in mind that the hardware installed in drones or missiles does not have much computing capacity, working mostly on the basis of field-programmable gate arrays. The available computing power further restricts the choice of networks, or requires the pruning of the models after successful training. Thus, a network should be chosen that can execute time-critical decisions with the given hardware resources.

Using TERCOM as a passive protective emblem

TERCOM uses a radar altimeter to compare the structure of the geographical surroundings with predefined parameters – i.e., a surface map of the operation area. Therefore, it could be used to recognize objects of a specific shape that are erected in the landscape. While the flight route is usually captured in lower resolution, the target area is of high resolution to improve precision. Large medical entities like field hospitals, which would be erected in a unique shape, could be detected by TERCOM and an attack could be stopped. Field hospitals are usually constructed from tents or containers, so the unique shape could be a cross of specific dimensions. An arriving missile using TERCOM would detect the shape and compare the structure with a database of objects that are not to be attacked.

Possibilities of allocation

A digital, next-generation protective emblem should be centrally defined and used by everyone, as is the current emblem. Using a digital emblem does, however, bring novel possibilities, as each emblem could be individually issued and revoked. It also brings novel challenges, since preventing and fighting misuse is even more important.

Encoding and certification of authenticity

While the protective emblem has been universally adopted since its inception and can be seen as a wide success, it has also always struggled with misuse. This can range from good-faith out-of-context usage by civilian entities to deliberate misuse to obfuscate legitimate military targets. With the traditional emblem, such misuse can be documented by pictures or verbal accounts. This can then lead to public backlash, as well as limited observance of the protective sign.[47]

A digital, next-generation protective emblem comes with additional problems regarding misuse. Documenting cases of misuse is more difficult, as they may not be visible to humans or cannot be photographed. Trust that the

[47] Baptiste Rolle and Edith Lafontaine, "The Emblem that Cried Wolf: ICRC Study on the Use of the Emblems", *International Review of the Red Cross*, Vol. 91, No. 876, 2009.

digital emblem is only used correctly is therefore more difficult to establish: if there is no way to recognize or punish misuse, why trust the emblem in the first place? The issue can be further highlighted by the use case in autonomous warfare. A drone may decide not to strike a target based on the presence of the emblem, and depending on the (usually limited) communication of the drone, even the party that is operating it may never be notified that a target was avoided due to the emblem. Even if the operators were notified, it might prove difficult for them to establish whether the usage of the protective emblem was justified. As such, there are valid concerns against the avoidance of targets marked by a digital emblem. Therefore, it is necessary to increase trust in the proper usage of the emblem in order to encourage its introduction, usage and observance.

Centralized and decentralized systems of trust

Problems of trust and misuse are not unique to protective emblems but are ubiquitous throughout the digital domain. Fortunately, this means that there are already widely used solutions in place. The most common are public key certificates, also known simply as digital certificates, used in authenticating the validity of websites and emails. It should be noted here that a simple database lookup system, as is used e.g. for aeroplane tickets, is not sufficient since such a system does not provide a method of identification. Digital certificates do provide such a method; they are the basis for Transport Layer Security (TLS), which is the basis for Hypertext Transfer Protocol Secure (HTTPS), a secure encryption standard for web browsing.[48] For the protective emblem, unlike for HTTPS, encryption of messages is not the goal, only the authentication. This is therefore not in violation of the requirements for unencrypted communication for certain non-combatants, e.g. hospital ships. Digital certificates have to be signed by trusted organizations in order to be valid. This results in a centralized structure as both sides have to trust the same trusted organization which guarantees the authenticity. This in turn results in a hierarchy of trusted institutions known as the chain of trust. In the use case of web browsing, each browser comes with a list of trusted certification authorities and each website requires a certificate that is signed by one of these authorities. Notably, it is also possible for the certification authority to revoke a certificate by issuing a signed statement. This process is very important for a protective emblem, since it allows the system to revoke misused emblems or even revoke all emblems used by an offending faction.

More recently, decentralized systems which do not rely on trusted third parties have been developed which are colloquially known as blockchains. A blockchain is an append-only database in which each data package, or block, can be appended to the chain after validating that it fits the requirements defined by a common protocol and the previous blocks. Each block contains a cryptographic hash of all previous blocks, which makes it very difficult to manipulate the content of the blocks. Multiple copies of the chain exist, and if a newly added

[48] Tim Dierks and Eric Rescorla, *The Transport Layer Security (TLS) Protocol Version 1.2*, 2008.

block is deemed invalid it will not be reproduced on the other copies. Mechanisms exist to prevent simultaneous appending by multiple parties; this is known as the double spend problem, as blockchains are most used for digital currencies. These mechanisms rely on solving complex computations (proof of work, or PoW) or ownership of portions of the digital currency (proof of stake, or PoS). Overall this leads to a design in which trust is not relegated to a singular entity but rather is guaranteed by a consensus of the largest part of the network, defined as either those with most computational resources for PoW or the most tokens in PoS.[49] In our use case, each protective emblem could be added to the chain inside a new block and be validated by the other users of the same blockchain. This is similar to the current blockchain use case of non-fungible tokens, or NFTs. Verification of authenticity is straightforward as only the token's existence in the chain needs to be checked. Ownership of the token can also be established, which for protective emblems would involve the identification of the party fielding the protective emblem.

Both digital certificates and blockchain come with a degree of complexity, but blockchain-based systems are notoriously complex to implement and maintain. The consensus needed to coordinate the authorization and revocation of protective emblems will be difficult to achieve in general. This is a political problem as parties with naturally opposing interests will need to reach a consensus, with some parties overruling others, but this can be achieved in a central organization such as the United Nations. A decentralized system of trust does not solve this fundamental issue; instead, it tries to represent the rules for consensus algorithmically, which is difficult, complex and error-prone.[50] The perceived benefit of a decentralized system also does not hold up in practice, as both the development and maintenance (i.e., governance) of the system often results in a degree of centralization.[51] For our use case, a central, neutral authority is more practical, as it already exists for the traditional emblem in the form of the International Committee of the Red Cross.

Discussion of the feasibility of digital authentication

The standard used for digital certificates by TLS is X.509. A certificate in this format usually requires 1–2 kilobytes of storage. While this is not a lot for modern communication systems, it does represent a significant size if the certificate needs to be encoded in a cross-frequency emblem. For comparison, a QR code can hold

[49] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen and Huaimin Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends", *Proceedings of the IEEE International Congress on Big Data*, 2017.

[50] Iyolita Islam, Kazi Md Munim, Shahrima Jannat Oishwee, A. K. M. Najmul Islam and Muhammad Nazrul Islam, "A Critical Review of Concepts, Benefits, and Pitfalls of Blockchain Technology Using Concept Map", *IEEE Access*, Vol. 8, No. 1, 2020.

[51] Ashish Rajendra Sai, Jim Buckley, Brian Fitzgerald and Andrew Le Gear, "Taxonomy of Centralization in Public Blockchain Systems: A Systematic Literature Review", *Information Processing and Management*, Vol. 58, No. 4, 2021.

2 kilobytes of information, and passive RFID tags 1 kilobyte.[52] Depending on the method of communication it would therefore be necessary to further compress or reduce the information in the certificate.

In summary, digital certification of authenticity provides a way to both install and remove trust in a specific emblem, or a party that uses the emblem. This is paramount for the acceptance of protective emblems as it enables the punishment of misuse, which is of even greater importance for digital emblems. Developing and implementing such a system will require a significant effort but should prove to be well worth it given the advantages over legacy solutions.

Possible use

The use of such an algorithm to recognize active or passive protective emblems is the enhanced, modern realization of the original Geneva Convention. In times of (fully) autonomous warfare, where the location of impact is a great distance away from the point of launch, operators might not be able to check for non-combatants at the area of effect. Therefore, taking the example of a fully autonomous missile that would possibly make use of active protective emblems, such a system could itself recognize the presence of protected facilities and disintegrate on its own. An example for passive emblems could be the use of autonomous drones – these still interact with a human operator and could therefore either ask for confirmation if a protective emblem is present, or refuse the use of weapons or the dropping of a payload automatically. A human-in-the-loop brings the advantage of manual control.

The next-generation protective emblem is a logical consequence of the ongoing development of modern weapons, which are able to autonomously fight an alleged target. In particular, non-combatants, who have been difficult to protect in the past (see Figure 6), can be better protected in the future through active or passive measures when using the next-generation protective emblem. Thus, the Geneva Conventions can also be implemented in modern warfare.

International efforts are needed to introduce this further development. Similar to the banning of certain weapons, the international community must agree on the use of the next-generation protective emblem. On the protected side, e.g. a military hospital, the appropriate technology must be installed, and medical units must be retrofitted with it. The weapons industry must also be obliged to implement routines in weapon systems to check the environment for the presence of protective signs. These regulations must be imposed by government on the industry.

[52] Tan Jin Soon, "QR Code", *Synthesis Journal*, 2008; David Chadwick, Alexander Otenko and Edward Ball, "Role-Based Access Control with X.509 Attribute Certificates", *IEEE Internet Computing*, Vol. 7, No. 2, 2003; Ron Weinstein, "RFID: A Technical Overview and Its Application to the Enterprise", *IT Professional*, Vol. 7, No. 3, 2005.

Vulnerabilities and ethical considerations

As with the introduction of every new technological advance, new ethical challenges arise with the introduction of a new protective emblem and the ability of weapon systems to recognize protected facilities by this emblem.

One of the most obvious ethical problems is the misuse of the protective emblem to protect facilities which would not officially have protected status. The digital protective emblem envisioned here would have the power to disable autonomous weapon systems and abort ongoing attacks, and this would potentially happen without any human interference. Automatically disabling or halting enemy attacks would confer a huge advantage to the attacked entity. Even adding a human-in-the-loop, who could assist in unsure cases and override a received protective emblem in cases where it turns out to be faulty, would have the effect of halting the attack until a final decision is made. This would give the attacked entity time to defend itself and possibly destroy the attacking weapon system while it waits for a final decision. Furthermore, the allure of misusing the digital protective emblem could be greater than that of misusing the traditional protective emblem, since the effect of a successful deception in the former case would be much more powerful.

Another problem is the opposite case, when digital protective emblems are ignored. This is, of course, possible with the traditional emblems as well, but in case of the digital protective emblems described here the behaviour would need to be coded into the (autonomous) weapon system. This can of course also be ignored, or an override can be added into the code.

This also relates to the next problem, which is that all weapon systems would have to be equipped with the technology needed to receive and process the necessary signals. All nations would have to participate in this effort, and an international agreement about the usage and processing of the digital protective emblem would have to be signed. This would mean that all newly built weapon systems would have to be equipped with the necessary sensors, while older weapon systems would have to be retrofitted with them or taken out of usage. It is, however, questionable whether all weapon manufacturers would participate in this and whether all governments would enforce these rules.

Another aspect of the vulnerability of the digital protective emblem is the intentional blocking or spamming of signals, such as radar signals. For the active protective emblem as described in this article, it would be necessary to emit and receive radar signals. Using other senders to intentionally block the radar signals emitted by the protected facility would make it vulnerable to attacks, as attacking weapon systems would not be able to recognize a facility as protected.

These vulnerabilities of the digital protective emblem could all potentially be exploited during an ongoing armed conflict. As with the traditional protective emblem, it is a matter of ethical considerations to respect the digital protective emblem and not to misuse it.

As discussed above, it would also be possible to equip all handheld weapons with emitters, such as RFID tags, to enable the recognition of active military personnel as opposed to e.g. paramedics. Also here, it would be necessary to reach an international agreement that all weapons need to be equipped with an RFID tag and that these tags are only to be used for recognition of active combatants. In this case, it would be possible to misuse the system to locate and track opposing units in order to coordinate an attack – however, since multiple technologies to achieve this already exist, it is questionable whether RFID tags would actually make soldiers easier to locate. Similarly, equipping larger mobile units like tanks with radar emitters would make them detectable from a greater distance and would allow the misuse of this information for the detection of active combatants. To some extent this is already possible today by using visual cues, but radar would allow longer-range detection of these units than is possible using visual cues. Again, this would necessitate the signing of an international agreement not to misuse such information. Furthermore, a balancing of the need for protection versus the need to keep radio silence would be necessary in such a situation. Generally, mobile units should not be equipped with long-range emitters and instead should use short-range emitters or visual protective emblems. A possibility would be to equip mobile units such as armoured medical evacuation vehicles with RFID tags, which would only be detectable at short range. On the other hand, it would also be possible to use AI and situational awareness to detect these units as non-combatants.

The use of RFID tags leads to another problem, however – namely, how to discern active, wounded and surrendering combatants. As described above, ML methods could be employed to identify these protected individuals in images or videos. However, the use of AI has its own problems and vulnerabilities. Ensuring that the algorithm has been trained properly and without an inherent bias would be vital. Even with such training, it would still be possible to use adversarial attacks on such a weapon system, which could provoke attacks on protected individuals. It would also be possible to decipher the behaviour that the algorithm has learned to identify protected individuals, for example a specific posture or way of moving. This would make it possible to train a certain behaviour which would lead the algorithm to believe that an active combatant is a protected individual, therefore protecting that person from attacks by the weapon system.

Finally, this new protective emblem would also come with a huge potential for constant surveillance of soldiers and civilians alike. It is therefore necessary to reflect on the balance between surveillance for the protection of non-combatants and the protection of personal rights.

Conclusion

In the age of semi- and fully autonomous weapon systems, the traditional protective emblem is no longer sufficient to cover all cases of protected entities under the Geneva Conventions. It is therefore necessary to develop new strategies to ensure protection of these facilities and individuals. Here, we have discussed possibilities for new cross-frequency digital protective emblems. These would encompass both stationary and mobile protected entities, as well as autonomous weapon systems with different types of sensors. However, it would be necessary to install new software or equip the weapon systems with new emitters and sensors. Overall, it would take considerable effort to deploy these digital protective emblems, but considering the current developments in warfare, it is also highly necessary to update the protective emblem.