INFORMES Y DOCUMENTOS

Informes

Esta sección de la International Review *presenta tres reseñas de informes recientes vinculados con el CICR que abordan la temática de este número: "Las tecnologías digitales y la guerra". Se incluyen los resúmenes de los tres informes y los enlaces a los textos completos.*

***

Informe del simposio: Riesgos digitales en conflictos armados, octubre de 2019

Este informe resume los puntos de acción y conclusiones principales del simposio sobre riesgos digitales en conflictos armados y en otras situaciones de violencia celebrado por el Comité Internacional de la Cruz Roja (CICR) en diciembre de 2018. En el encuentro, de dos días, se reunieron representantes de organizaciones humanitarias, académicos, empresas del ámbito de la tecnología y organismos gubernamentales, así como representantes de varios donantes. El debate se centró en el uso que hacen de las tecnologías digitales las partes en conflicto, las empresas privadas e incluso el sector humanitario (como parte de las respuestas humanitarias), y cómo ese uso puede poner en riesgo a las personas afectadas por crisis y volverlas más vulnerables, tanto en el ámbito virtual como real.

Disponible en https://shop.icrc.org/symposium-report-digital-risks-in-armed-conflicts-print-en (en inglés).

La cuestión de los metadatos humanitarios: "No causar daño" en la era digital, octubre de 2018

Las nuevas tecnologías son una fuente de grandes riesgos y oportunidades para la acción humanitaria. A fin de asegurar que su uso no cause daño alguno, las organizaciones humanitarias deben establecer e implementar estándares apropiados para la protección de datos, entre ellos, evaluaciones de riesgos que sean

confiables. Pero para ello se requiere una sólida comprensión de esas tecnologías, los riesgos que supone su uso y las estrategias que podrían evitarlos o mitigarlos. En este informe conjunto de Privacy International y el CICR, se intenta brindar la información necesaria para que quienes trabajan en el ámbito humanitario comprendan los riesgos que conlleva el uso de algunas nuevas tecnologías. Además, se analizan el principio de "no causar daño" y su aplicación en el mundo digital.

Disponible en www.icrc.org/en/download/file/85089/the_humanitarian_meta data_problem_-_icrc_and_privacy_international.pdf (en inglés).

Manual sobre protección de datos en la acción humanitaria, segunda edición, mayo de 2020

Este manual se publicó en el marco del proyecto sobre protección de datos en la acción humanitaria conducido por el Brussels Privacy Hub y el CICR. Está dirigido al personal de las organizaciones humanitarias que participa en el procesamiento de datos personales en las actividades humanitarias, en particular, los encargados de asesorar sobre las normas relativas a la protección de datos y su aplicación. El manual se basa en lineamientos, procedimientos de trabajo y prácticas existentes y establecidas en actividades humanitarias que se llevan a cabo en los contextos más volátiles y para beneficio de las víctimas más vulnerables de las emergencias humanitarias Procura ayudar a las organizaciones humanitarias a respetar las normas relativas a la protección de datos personales, generando sensibilización y proporcionándoles orientación específica sobre la interpretación de los principios relativos a la protección de datos en el contexto de la acción humanitaria, en particular cuando se emplean nuevas tecnologías.

Disponible en https://www.icrc.org/es/publication/manual-sobre-proteccion-de-datos-en-la-accion-humanitaria.

El potencial costo humano de las ciberoperaciones, mayo de 2019

Resumen

Las operaciones cibernéticas durante los conflictos armados: análisis de los retos para el derecho internacional humanitario

Las operaciones cibernéticas durante los conflictos armados son una realidad. Si bien solo algunos Estados han reconocido públicamente que llevan adelante ese tipo de operaciones, es sabido que las ciberoperaciones son parte de las operaciones militares actuales y que su uso seguramente aumente en el futuro.

Esta nueva realidad ha despertado un debate en torno a las normas del derecho internacional que se aplican a dichas operaciones. Al respecto, el CICR recuerda que, durante los conflictos armados, las ciberoperaciones están sujetas

a las normas del derecho internacional humanitario (DIH)[1]. Sin embargo, es evidente que el ciberespacio y estas nuevas operaciones militares dan lugar a varios cuestionamientos respecto de cómo se aplican a las ciberoperaciones algunas normas del DIH, originalmente concebidas para operaciones cinéticas.

El análisis de esos temas requiere la comprensión del uso previsto y del potencial militar de la tecnología cibernética. ¿Qué pretenden lograr los beligerantes mediante el uso de estas nuevas herramientas en áreas estratégicas, operacionales o tácticas durante un conflicto? ¿Cómo se comparan dichas tecnologías con otros medios de guerra existentes?

Además, para evaluar la protección que otorga el DIH a las personas civiles durante los conflictos armados y la posible necesidad de establecer regulaciones adicionales, los abogados y responsables de la formulación de políticas deben comprender el costo actual o potencial de la tecnología cibernética. De hecho, uno de los principales objetivos del DIH es proteger a la población civil de los efectos de las operaciones militares.

Finalidad y alcance de la reunión sobre el DIH y las ciberoperaciones

A fin de cumplir con su cometido de esclarecer el DIH y, de ser necesario, preparar su eventual desarrollo, el CICR supervisa la evolución de nuevas tecnologías que se utilizan como métodos y medios de guerra durante los conflictos armados. Este enfoque se basa en consideraciones interrelacionadas de los ámbitos jurídico, técnico, militar y humanitario.

A fin de realizar una evaluación realista de las capacidades cibernéticas y sus posibles consecuencias humanitarias a la luz de sus características técnicas, el CICR convocó a científicos y expertos en ciberseguridad de todas partes del mundo para que intercambiaran conocimientos acerca de las posibilidades técnicas, el uso previsto y los efectos potenciales de las operaciones cibernéticas. Durante los tres días del encuentro, reunió a representantes de empresas internacionales de tecnología de la información (TI), empresas especializadas en inteligencia sobre ciberamenazas, equipos de respuesta ante emergencias informáticas, una agencia de ciberseguridad nacional, especialistas en materia de ciberseguridad (de hospitales, redes de distribución eléctrica y otros servicios), conocedores del desarrollo y uso de las ciberoperaciones militares, abogados y académicos.

Los Estados y los militares son reticentes a divulgar sus capacidades cibernéticas, como los detalles de las ciberoperaciones realizadas en el contexto de los conflictos armados, y poco se sabe sobre los escasos ejemplos en los que se admitió su uso. Por eso, los expertos se enfocaron en algunas de las ciberoperaciones más complejas de las que tenemos conocimiento, independientemente de si se llevaron a cabo durante un conflicto armado o en tiempo de paz. El análisis de los aspectos técnicos de estos ataques y las vulnerabilidades específicas de sus objetivos

1 V., especialmente, CICR, *El derecho internacional humanitario y los desafíos de los conflictos armados contemporáneos*, Ginebra, 2015, pp. 39-44. Los límites que impone el DIH no legitiman el uso de la fuerza en el ciberespacio, que sigue rigiéndose por la Carta de las Naciones Unidas.

proporciona evidencias sólidas de lo que es técnicamente posible durante un conflicto armado. El eje principal del encuentro fue el riesgo de que las operaciones cibernéticas causen muertes, heridas o daños físicos y afecten el suministro de servicios esenciales para la población o la confiabilidad de los servicios de internet. Se examinaron las características específicas de las herramientas cibernéticas, la evolución de las ciberamenazas y el panorama de la ciberseguridad.

El CICR, que aborda este tema desde la perspectiva del derecho humanitario y la acción humanitaria, procura tener evidencias fiables para comprender, tanto como sea posible, los riesgos de los ciberataques[2] para la población civil. Gracias a esta reunión, el CICR pudo confirmar gran parte de sus propias investigaciones y complementar sus conclusiones con el valioso aporte de los expertos. Además, el evento fue sumamente provechoso porque permitió esbozar una imagen más completa y matizada de las ciberoperaciones y desmitificar algunos supuestos relacionados con la guerra cibernética.

Preocupaciones principales

El debate ayudó a poner en primer plano cuatro preocupaciones principales que tienen que ver con el potencial costo humano de las ciberoperaciones:

a. las vulnerabilidades específicas de algunos tipos de infraestructura;
b. el riesgo de que se produzca una reacción exagerada por un posible malentendido acerca de la finalidad de las ciberoperaciones hostiles;
c. la particular manera de proliferación de las herramientas cibernéticas;
d. los obstáculos que supone la dificultad de atribución de ciberataques para asegurar el cumplimiento del derecho internacional.

a. Vulnerabilidades específicas de algunos tipos de infraestructura: ciberataques que pueden afectar la prestación de asistencia de salud, sistemas de control industrial o la confiabilidad y disponibilidad de servicios clave de internet

Además de generar pérdidas económicas significativas, las ciberoperaciones pueden causar daños a la infraestructura de dos maneras, como mínimo. En primer lugar, pueden afectar la prestación de servicios esenciales a la población civil, como ha sucedido tras ciberataques a centrales eléctricas y al sector de la salud. En segundo y último lugar, pueden causar daños físicos, como lo demostraron el ataque Stuxnet a una central de enriquecimiento de uranio iraní en 2010 y el ataque a una fábrica de acero alemana en 2014.

2 Este informe usa los términos "ciberataques", "ciberoperaciones" y "operaciones cibernéticas" de manera técnica (uso cotidiano o coloquial) y no necesariamente responde a las definiciones establecidas en el DIH, a menos que se indique lo contrario.

Los ciberataques pueden afectar la prestación de servicios de salud

El sector de la salud está migrando a la digitalización y la interconectividad. Por ejemplo, los equipamientos médicos de los hospitales suelen estar conectados al sistema de tecnología de la información del establecimiento para permitir el archivado electrónico de forma automática. Los dispositivos biomédicos conectados, como los marcapasos y las bombas de insulina, permiten supervisar el estado de salud de los pacientes y el funcionamiento de los dispositivos en sí.

Hasta ahora, no ha habido una mejora en la ciberseguridad que acompañe esta creciente dependencia digital y la expansión en el "campo de acción para posibles ataques". Eso hace que infraestructuras como esas sean especialmente vulnerables y que los ciberataques en su contra puedan tener consecuencias graves en la salud y la vida de la población afectada.

Ciberataques a sistemas de control industrial, incluidos aquellos utilizados en infraestructura civil crítica

Los sistemas de control industrial están protegidos por mecanismos de seguridad complejos y suelen tener sistemas de redundancia integrados para garantizar su seguridad y su confiabilidad. Por ejemplo, las redes eléctricas cuentan con múltiples fuentes de energía para evitar que haya efectos masivos si se produce algún problema con una de sus partes. De todas formas, los ataques a nodos específicos pueden tener un impacto considerable, por ejemplo, si un sistema crítico (como un hospital) depende de un subsistema o nodo puntual, o si el ataque tiene consecuencias perjudiciales "en cascada").

Para llevar a cabo un ciberataque contra un sistema de control industrial, se necesita cierto nivel de conocimiento, complejidad y, por lo general, programas informáticos maliciosos, o *malware*, diseñados específicamente para ese fin. Hasta el momento, la frecuencia de tales ataques ha sido menor que la de otros tipos de ciberoperaciones. Sin embargo, son cada vez más frecuentes, y la gravedad de las amenazas aumenta más rápido de lo anticipado hace unos años. Existe el riesgo de que las herramientas desarrolladas por los actores con más recursos puedan reutilizarse con otros fines o sean adquiridas por otros actores que no tienen el conocimiento para desarrollarlas desde cero. Además, es posible que haya actores no detectados con la capacidad de atacar los sistemas de control industrial.

Los ciberataques pueden afectar la confiabilidad y disponibilidad de los servicios de internet

Los ciberataques que afectan el funcionamiento de servicios informáticos en la nube o servicios clave de internet –como el sistema de nombres de dominio (DNS, por sus siglas en inglés), indispensable para las comunicaciones en internet– pueden repercutir en todos los servicios que dependen de ellos. Sin embargo, los expertos consideran que, actualmente, es poco probable que los servicios clave de

internet se vean gravemente afectados, dada la gran redundancia implementada en el DNS y los altos estándares que suelen ofrecer los principales proveedores de los servicios en la nube. De todas formas, si hubiese una interrupción del servicio, podría ser masiva y tener repercusiones potencialmente graves, por ejemplo, si servicios vitales como las ambulancias dependen de la nube.

Finalmente, cabe destacar que ha habido ataques de denegación de servicio distribuido (DDoS, por sus siglas en inglés) dirigidos a servicios para la población provistos por el gobierno. Esos ataques se llevan a cabo a través de redes de robots informáticos, o "botnets", cada vez más extensas. La llegada de la "internet de las cosas" aumentará la cantidad de dispositivos conectados que pueden utilizarse para esos ataques. Además, un ataque de DDoS podría tener un impacto mayor al esperado por el autor del ataque, sobre todo si la información sobre la red objetivo no está completa.

b. Riesgo de que se produzca una reacción exagerada por un posible malentendido acerca de la finalidad de las ciberoperaciones hostiles

Las operaciones cibernéticas se pueden clasificar en dos categorías en función de su objetivo:

- actividades que abarcan la vigilancia, el reconocimiento y la exfiltración de datos e información –por ejemplo, con fines de espionaje–, mayormente conocidas como explotación de redes de computadoras (CNE, por sus siglas en inglés) u "operaciones de acceso";
- actividades que tienen como objetivo afectar un sistema o dispositivo específico –por ejemplo, alterar sus datos (eliminarlos, editarlos), modificar su disponibilidad (lograr la desactivación por un período breve o prolongado) o causar daños físicos (dañar el sistema)–, mayormente conocidas como ataques a redes de computadoras (CNA, por sus siglas en inglés) u "operaciones basadas en efectos".

La diferencia entre ambas es, principalmente, su finalidad. En lo que respecta al plano técnico, es posible que una CNE y una CNA usen los mismos pasos para acceder al objetivo y controlarlo. Las CNE pueden transformarse en CNA con relativa facilidad, en general, a través de cargas útiles específicas de distinta naturaleza. Aunque se puedan rastrear los pasos iniciales de los ataques, suele ser difícil determinar las intenciones del autor del ataque hasta que sea evidente el efecto final en el objetivo.

Cuando el objetivo no está seguro de la intención del ataque, podría pensar en las peores consecuencias que tendría si fuera un CNA y reaccionar de manera más exagerada que si hubiera sabido que se trataba de un CNE. El riesgo de escalada puede llevar a una reacción exagerada potencialmente perjudicial.

c. Proliferación de herramientas cibernéticas

Una tercera preocupación es la proliferación de herramientas cibernéticas. En ciertos aspectos, despierta inquietudes similares a las generadas por la proliferación de armas o de tecnología de doble uso, aunque siempre hay que tener presentes las particularidades de las herramientas cibernéticas.

Las herramientas y los métodos cibernéticos tienen un modo único de proliferación que es difícil de controlar. Por un lado, el ciberespacio es un dominio global: si un atacante puede burlar las medidas de ciberdefensa y ciberseguridad implementadas, podrá acceder a todos los nodos y datos almacenados en la red desde cualquier parte del mundo. Y por otro lado, las herramientas cibernéticas pueden rediseñarse o reutilizarse con otros fines. La combinación de estas dos características implica que, cuando las herramientas cibernéticas se usan, se roban, se filtran o están disponibles en general, otros actores que no son los desarrolladores originales podrían hallarlas, aplicar ingeniería inversa y reutilizarlas para sus propios fines.

Por último, el hecho de que los métodos y las herramientas cibernéticas puedan reutilizarse con otros fines es uno de los factores que dificultan la atribución técnica rápida y fiable de los ciberataques.

d. Atribución de los ataques

Si bien no fueron el eje principal del encuentro, la anonimidad de los ataques y la dificultad de atribuirlos a actores específicos también fueron temas de debate. Esos aspectos constituyen la cuarta área de preocupación.

El ciberespacio es un dominio complejo en el que operan múltiples agentes: jáqueres independientes, bandas criminales (en general, motivadas por un interés económico), Estados, grupos armados no estatales y otros actores no estatales. A veces, los actores cooperan: por ejemplo, los Estados pueden adquirir herramientas cibernéticas o instruir a otros que lleven a cabo ciberoperaciones en su nombre contra objetivos identificados.

Tanto el análisis forense digital como las capacidades de atribución de actividad cibernética maliciosa parecen estar mejorando. No obstante, la habilidad de los actores de cubrir sus rastros y hasta ocultar el origen de sus operaciones en internet, así como la capacidad de adquirir herramientas cibernéticas desarrolladas o implementadas por otros actores y reutilizarlas o rediseñarlas para otro fin siguen impidiendo la atribución de los ciberataques a actores específicos con rapidez y certeza. Esos factores obstaculizan la posibilidad de identificar a los actores que violan el DIH en el ciberespacio y atribuirles la responsabilidad que corresponda. Y eso es alarmante, porque responsabilizar a los autores de los ataques es una medida que ayuda a garantizar el cumplimiento del DIH. También es posible que se reduzca el umbral relativo al uso de los ciberataques y que esas operaciones constituyan una violación del derecho internacional, ya que los autores de los ataques pueden negar su responsabilidad.

Las operaciones cibernéticas durante los conflictos armados: repercusión en el derecho internacional humanitario

Se ha establecido que el derecho internacional se aplica a las ciberoperaciones. Más específicamente, el DIH y los principios de distinción, proporcionalidad, precaución, necesidad militar y humanidad restringen el uso de métodos y medios cibernéticos durante los conflictos armados. De todas formas, puede que sea necesario seguir debatiendo para aclarar cómo se aplica el DIH y si el marco que propone es suficiente y adecuado, o si, basándose en el derecho existente, se requeriría un mayor desarrollo.

Durante el encuentro, se ayudó a esclarecer qué cuestiones de interés humanitario deberían ser las principales áreas de atención. En síntesis, la minuciosa información disponible sobre las ciberoperaciones en tiempo de paz y el conocimiento más general sobre aquellas llevadas a cabo en tiempo de conflicto armado permiten esbozar el siguiente panorama sobre esas operaciones.

El principio de distinción en el ciberespacio

En primer lugar, hay que destacar que los ciberataques no necesariamente son indiscriminados. Como se detalla en el informe, las herramientas cibernéticas pueden estar diseñadas para autopropagarse (o no). Incluso si se autopropagan y se convierten en un problema de ciberseguridad para todos los dispositivos o sistemas afectados, pueden estar diseñados para dañar solo a un objetivo en particular. Si bien se han difundido casos de *malware* que se autopropagan y tienen efectos perjudiciales indiscriminados, muchas operaciones cibernéticas fueron de efectos bastante discriminados desde el plano técnico (lo que no implica que hayan sido lícitas).

Además, ciertos tipos de ciberataques, como los que tienen como finalidad causar daños a sistemas de control industrial, requieren el uso de herramientas cibernéticas personalizadas. En muchos casos, eso puede impedir que se realicen ciberataques a gran escala y de forma indiscriminada.

Y eso es importante desde la perspectiva del DIH, porque, contrario a la suposición común de que el principio de distinción no es pertinente en el ciberespacio a raíz de su interconectividad, no todas las herramientas cibernéticas ofensivas tienen inherentemente efectos indiscriminados. De hecho, se pueden personalizar de manera muy precisa para afectar solamente a objetivos específicos.

El potencial costo humano en primer plano

En segundo lugar, pero no por eso menos importante, está claro que las herramientas cibernéticas pueden causar daños sustanciales y tener efectos indiscriminados –en algunas ocasiones–, y que ciertos tipos de sistemas corren un mayor riesgo (seguramente, los más afectados sean los sistemas de salud). Además, las amenazas detectadas han avanzado más rápidamente de lo previsto, sobre todo en

lo que respecta a los ataques a sistemas industriales. Aún no se sabe mucho acerca de la rápida evolución de la tecnología, las capacidades y las herramientas desarrolladas por los actores con mejores recursos, y la medida en que el uso más extendido de las operaciones cibernéticas durante los conflictos armados podría apartarse de las tendencias observadas hasta ahora. En otras palabras, si bien el riesgo en materia de costo humano no parece demasiado alto por el momento –especialmente si se lo compara con la destrucción y el sufrimiento que siempre causan los conflictos–, la incertidumbre y el ritmo acelerado de cambio ameritan seguir de cerca la evolución de las ciberoperaciones.

Protección jurídica a través del DIH

Muchos de los ataques descritos en el informe se dirigieron contra infraestructura civil o la afectaron indiscriminadamente. En opinión del CICR, en caso de llevarse a cabo en tiempo de conflicto armado, esos ataques estarían prohibidos. Antes que nada, se prohibirían los ataques directos contra infraestructura civil y los ataques indiscriminados. Luego, incluso si la infraestructura o parte de ella fuera un objetivo militar (como una sección de una red eléctrica), el DIH exigiría limitar el ataque solo a esa sección y que no haya daños excesivos a los sectores civiles restantes. Además, el DIH establece que las partes en los conflictos deben tomar todas las precauciones posibles para impedir, o al menos minimizar, los daños incidentales a personas civiles y a bienes de carácter civil. Por último, aunque no constituyan un ataque según el DIH[3], tales operaciones podrían prohibirse por la protección específica conferida por el DIH a instalaciones médicas o bienes indispensables para la supervivencia de la población. Son protecciones poderosas que siguen teniendo total pertinencia dadas las características técnicas de las operaciones cibernéticas. Pero para que el DIH proporcione a los civiles protecciones jurídicas plenas contra los efectos de la guerra informática, los Estados deben comprometerse a aplicarlas y adoptar una interpretación de sus normas que sea efectiva para proteger a la población y la infraestructura civiles. En particular, se requeriría que los Estados reconocieran claramente que las operaciones cibernéticas que afectan el funcionamiento de la infraestructura civil están sujetas a las normas del DIH que rigen los ataques[4]. Esperamos que este informe ayude a demostrar la necesidad de tal interpretación a fin de garantizar la protección de las infraestructuras civiles.

3 En el DIH, el término "ataque" tiene un significado específico que no abarca todas las operaciones cibernéticas que se engloban en el uso coloquial de "ciberataques".

4 V. CICR, nota 1 mencionada arriba, pág. 41.

Posibles formas de reducir el potencial costo humano de las ciberoperaciones

Medidas de ciberseguridad

Más allá de las limitaciones impuestas por el DIH a quienes llevan a cabo las ciberoperaciones, es fundamental mejorar la ciberseguridad y la resiliencia de los actores potencialmente afectados. Si bien la ciberdefensa y la ciberseguridad mejoran todo el tiempo, los sistemas más antiguos, que tienen medidas de seguridad desactualizadas (o ninguna), son particularmente vulnerables a los ataques cibernéticos y seguirán siendo una fuente de preocupación en los años venideros. Tanto el sector público como el sector privado tienen un papel que desempeñar en lo que respecta a los estándares de la industria y la regulación jurídica.

En el sector de la salud, por ejemplo, se debería adaptar el entorno regulatorio en función del aumento de riesgo a través de los requisitos de estandarización, con la finalidad de garantizar la resiliencia frente a un ciberataque. La seguridad es un factor que se debe tener en cuenta a la hora de diseñar y fabricar dispositivos médicos, que, además, se deberían actualizar a lo largo de su vida útil (sin importar su duración). De manera similar, los estándares de la industria, ya sean impuestos o autoexigidos, son clave para los sistemas de control industrial. Eso incluye denunciar incidentes e intercambiar información con socios confiables.

En lo que respecta al DIH, las partes en conflictos armados deben tomar todas las precauciones posibles para proteger de los efectos de los ataques a las personas civiles y a los bienes de carácter civil que están bajo su control. Esta es una de las pocas obligaciones del DIH que los Estados deben implementar en tiempo de paz.

Divulgación de vulnerabilidades

La opción de preferencia para mejorar la seguridad en el ciberespacio debería ser divulgar las vulnerabilidades al desarrollador de *software* correspondiente, que se encargará de repararlas. En función de esa premisa, algunos Estados han puesto en marcha procesos de equidad para poner en la balanza los riesgos y los conflictos de interés, a fin de decidir si divulgarán o no las vulnerabilidades que identifiquen.

Medidas para evitar la proliferación

Los desarrolladores de armas cibernéticas deberían considerar la inclusión de obstáculos para que reutilizarlas sea difícil y costoso. Si bien, por sus características técnicas, es casi imposible garantizar que el *malware* no podrá ser reutilizado con otros fines, estrategias como encriptar la carga útil o incluir ciertos obstáculos en varios componentes del código podrían elevar el nivel de conocimientos necesarios para rediseñar las herramientas maliciosas. Actualmente, el DIH no impone una obligación expresa de obstaculizar la reutilización de herramientas cibernéticas

para otro fin, pero tomar esas medidas podría evitar que al menos algunos actores lo hagan y, por lo tanto, reduciría el riesgo de posteriores usos indebidos que conlleva la proliferación. Además, la forma de proliferación tan singular de las herramientas cibernéticas plantea dudas sobre la adecuación o suficiencia del derecho existente para abordar este fenómeno.

Señalización de ciertas instalaciones de infraestructura civil

Otra alternativa, que se basa en el derecho internacional existente, podría ser crear una "marca de agua digital" para identificar en el ciberespacio a actores o tipos de infraestructura que es necesario proteger (como los bienes a los que el DIH les confiere protección específica). Eso ayudaría a reconocerlos y evitar que sean los objetivos de ciberataques durante los conflictos armados. Se deben sopesar los posibles efectos positivos en cuanto a la prevención de daños no previstos a manos de actores que respetan el derecho y el riesgo de divulgar información sobre la infraestructura crítica a potenciales adversarios, incluidos criminales. El impacto positivo de esta alternativa podría depender, en parte, de que se facilite la atribución.

Mejoras en materia de atribución y responsabilización

Es importante destacar que mejorar las capacidades de atribución ayudaría a responsabilizar a quienes violan el derecho internacional en el ciberespacio, lo cual es una forma de fortalecer el cumplimiento del derecho. Además, en términos más generales, se fomentaría el comportamiento responsable en el ciberespacio.

De cara al futuro

Es probable que se sigan llevando a cabo operaciones cibernéticas durante los conflictos armados y que su uso siga manteniéndose en secreto. El análisis de sus consecuencias es una tarea a largo plazo de gran complejidad para la que se requieren conocimientos y cooperación multidisciplinarios de una gran variedad de partes interesadas.

Sobre la base de las conclusiones que dejó el encuentro, el CICR quisiera entablar un diálogo con los gobiernos, expertos y el sector de tecnología de la información. Esperamos con interés los comentarios sobre este informe para continuar supervisando la evolución de las ciberoperaciones, en particular, durante los conflictos armados. Se estudiará su potencial costo humano (y las formas de reducirlo) y se trabajará para alcanzar un consenso en la interpretación del DIH actual y, de ser necesario, para el desarrollo de normas complementarias que confieran protección efectiva a las personas civiles. Disponible en www.icrc.org/en/document/potential-human-cost-cyber-operations (en inglés).

Autonomía, inteligencia artificial y robótica: aspectos técnicos del control humano, agosto de 2019

Resumen

El CICR ha destacado la necesidad de mantener un control humano sobre los sistemas de armas y el uso de la fuerza, para asegurar el cumplimiento del derecho internacional y atender las preocupaciones éticas. Ese enfoque ha guiado el análisis del Comité en cuestiones jurídicas, éticas, técnicas y operacionales relacionadas con los sistemas de armas autónomos.

En junio de 2018, el CICR organizó una mesa redonda con expertos independientes en autonomía, inteligencia artificial (IA) y robótica para adquirir una comprensión más profunda de los aspectos técnicos del control humano y aprovechar su experiencia con los sistemas autónomos del sector civil. Este informe combina un resumen de los intercambios mantenidos en el encuentro con otras investigaciones. Además, se destacan las principales conclusiones del CICR, que no necesariamente reflejan las opiniones de los participantes. Las experiencias en el sector civil arrojan datos que pueden contribuir a la definición de estrategias para asegurar la implementación de un control humano comprometido, efectivo y apropiado en los sistemas de armas y el uso de la fuerza.

Los sistemas autónomos (robóticos) operan sin intervención humana y se basan en su interacción con el entorno. Estos sistemas plantean varios interrogantes. ¿Cómo podemos efectuar un control humano efectivo que supervise su funcionamiento? ¿Cómo pueden preverse las consecuencias de su uso? Mientras mayor sea la complejidad del entorno y de la tarea, mayor será la necesidad de tener un control humano directo y menor será la tolerancia de los sistemas autónomos, especialmente en tareas y entornos donde haya riesgo de daños a los bienes, lesiones o muerte (en otras palabras, actividades críticas para la seguridad).

Las personas ejercen cierto nivel de intervención sobre los sistemas autónomos –o funciones específicas– a través de la supervisión, lo que constituye el control humano conocido como "human-on-the-loop" y la capacidad de intervenir o desactivar el sistema. Esto requiere que el operador posea:

- conocimiento de la situación;
- tiempo suficiente para intervenir;
- un mecanismo que le permita intervenir (un enlace de comunicación o controles físicos) y recuperar el control o, si las circunstancias lo ameritan, desactivar el sistema.

Sin embargo, el control humano "human-on-the-loop" no es la panacea, ya que hay problemas de interacción entre el ser humano y la máquina, entre otros, el sesgo sobre la automatización, la falta de conocimiento de la situación como operador y la restricción moral.

La previsibilidad y la fiabilidad son el eje de los debates sobre la autonomía de los sistemas de armas, ya que son esenciales para cumplir con el DIH y evitar

consecuencias adversas para los civiles. También son fundamentales para el mando y el control militar.

Es importante distinguir entre fiabilidad, una medida que indica la frecuencia de falla de un sistema, y previsibilidad, una medida que indica el funcionamiento del sistema en una circunstancia en particular. La fiabilidad es una preocupación en todo tipo de sistema complejo, mientras que la previsibilidad es un problema de los sistemas autónomos. Además, cabe destacar la distinción entre previsibilidad en un sentido estricto (es decir, saber qué proceso guía el funcionamiento del sistema y cómo se realizan las tareas requeridas) y en un sentido más amplio (saber cuál será el resultado).

Es difícil determinar y verificar la previsibilidad y la fiabilidad de un sistema (robótico) autónomo. Ambos factores dependen no solo de su diseño técnico, sino también de la naturaleza del entorno, la interacción del sistema con ese entorno y la complejidad de la tarea. Sin embargo, fijar límites o imponer restricciones en el funcionamiento de un sistema autónomo –en especial, la tarea, el entorno, el plazo y el alcance geográfico de una operación– contribuye a que las consecuencias del uso del sistema sean más previsibles.

En términos generales, todos los sistemas autónomos son impredecibles hasta cierto punto, ya que responden a disparadores del entorno. Además, la creciente complejidad de los sistemas de control de *software* –en especial, los que se basan en la inteligencia artificial y en el aprendizaje automático– añade una capa de imprevisibilidad en términos más concretos, ya que el proceso que guía el funcionamiento del sistema es impredecible.

Dada la naturaleza de "caja negra" de gran cantidad de sistemas de aprendizaje informático, para el usuario es difícil –y, a menudo, imposible– comprender cómo el sistema obtiene un resultado determinado. Los algoritmos no solo son impredecibles, sino que también están sujetos a sesgos tanto intencionales como causales. Asimismo, no brindan explicaciones de sus resultados, lo cual hace considerablemente más difícil confiar en su uso y complica aún más la evaluación y verificación del desempeño de los sistemas autónomos. La vulnerabilidad de los sistemas de inteligencia artificial y aprendizaje informático a los engaños y a la suplantación de identidad, o "spoofing", amplifica los problemas centrales de previsibilidad y fiabilidad.

La visión artificial y el reconocimiento de imágenes son aplicaciones importantes del aprendizaje informático. Estas aplicaciones usan redes neuronales profundas (aprendizaje profundo o "deep learning"), que tienen un funcionamiento no predecible ni explicable y están sujetas a sesgos. Más fundamentalmente, las máquinas no ven como los seres humanos. No comprenden el significado ni el contexto, y eso implica que cometen errores que nunca cometería una persona.

Es significativo que los estándares de la industria para los sistemas robóticos autónomos críticos para la seguridad civil –como los robots industriales, los sistemas de piloto automático para aviones y los automóviles sin conductor– establezcan requisitos estrictos con respecto a la supervisión, intervención y capacidad de desactivación humanos (salvaguardias); previsibilidad, fiabilidad y

limitaciones operacionales. Los desarrolladores más importantes de inteligencia artificial y aprendizaje informático han destacado la necesidad de garantizar el control y el juicio humanos en aplicaciones sensibles –y abordar las cuestiones de seguridad y los sesgos–, sobre todo en los casos donde las aplicaciones pueden tener consecuencias graves para la vida de las personas.

La experiencia con los sistemas autónomos en el ámbito civil refuerza y profundiza algunas observaciones y preocupaciones del CICR con respecto a la autonomía de las funciones críticas de los sistemas de armas. Las consecuencias de usar sistemas de armas autónomos son impredecibles, debido a la incertidumbre del usuario sobre el objetivo específico y el momento y lugar exactos del ataque resultante. Esos problemas empeoran a medida que aumenta la complejidad del entorno o de la tarea y la libertad de acción en cuanto al tiempo y al espacio. Si bien la supervisión e intervención "human-on-the-loop" y la capacidad de desactivación son los requisitos mínimos absolutos para contrarrestar el nivel de riesgo, el diseño del sistema debe permitir una intervención humana significativa y oportuna (y ni siquiera esa es una situación ideal).

Todos los sistemas de armas autónomos siempre suponen un grado de imprevisibilidad surgido de su interacción con el entorno. Podría mitigarse esa imprevisibilidad, hasta cierto punto, imponiendo limitaciones operacionales a la tarea, el plazo y el alcance geográfico de la operación y al entorno. Sin embargo, el uso de *software* de control basado en la inteligencia artificial –y, especialmente, el aprendizaje informático, que incluye las aplicaciones de reconocimiento de imagen– conlleva un riesgo inherente de imprevisibilidad, sesgos y falta de explicabilidad. Esto aumenta las preocupaciones del CICR con respecto al uso de la inteligencia artificial y el aprendizaje informático para controlar las funciones críticas de los sistemas de armas y plantea interrogantes en relación con su uso en sistemas de apoyo a la toma de decisiones para establecer objetivos.

Esta reseña de aspectos técnicos pone de manifiesto la dificultad de ejercer un control humano sobre los sistemas autónomos (de armas) y demuestra que la inteligencia artificial y el aprendizaje informático podrían exacerbar el problema de manera exponencial. En última instancia, confirma la necesidad de que los Estados trabajen con urgencia para establecer límites con respecto a la autonomía de los sistemas de armas.

Además, refuerza la opinión del CICR según la cual los Estados deberían llegar a un acuerdo sobre el tipo y el grado de control humano necesarios para respetar el derecho internacional y atender preocupaciones éticas, y cuestiona si los sistemas de armas autónomos se pueden usar respetando las normas del DIH en todos los casos y no solo en los más elementales y simples.

Disponible en www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control (en inglés).

Establecer límites a la autonomía de los sistemas de armas: identificación de elementos prácticos del control humano, junio de 2020

CICR y SIPRI. Vincent Boulanin, Neil Davison, Netta Goussac y Moa Peldán Carlsson

Resumen[5]

Los retos que plantean los sistemas de armas autónomos han sido el eje central de debate intergubernamental en el marco de la Convención sobre ciertas armas convencionales (CCA) de las Naciones Unidas. A pesar de los desacuerdos permanentes sobre la necesidad de establecer nuevas normas, y de qué manera hacerlo, crece el consenso entre los Estados con respecto a que los sistemas de armas no pueden tener autonomía ilimitada: se debe mantener y ejercer responsabilidad humana sobre el uso de sistemas de armas y el uso de la fuerza en conflictos armados. Este informe aborda una cuestión compleja: cómo poner en práctica ese principio. Se analiza en profundidad el tipo y el grado de control humano que se deben ejercer en los sistemas de armas autónomos teniendo en cuenta los requisitos jurídicos, las preocupaciones éticas y las consideraciones operacionales. Los responsables de formular políticas encontrarán orientación útil para establecer el control humano como la base de los límites internacionalmente acordados sobre los sistemas de armas autónomos, desde normas y estándares hasta prácticas idóneas.

El presente informe es el resultado de un proyecto conjunto del CICR y el Instituto Internacional de Estocolmo para la Investigación de la Paz (SIPRI). El capítulo 1 presenta el contexto y el enfoque conceptual. El capítulo 2 analiza las consideraciones jurídicas, éticas y operacionales del control humano. El capítulo 3 ofrece orientaciones prácticas sobre el tipo y el alcance de las medidas de control y sus posibles combinaciones para respetar el derecho internacional humanitario y abordar cuestiones éticas sin dejar de lado las consideraciones relativas a las operaciones militares. El capítulo 4 detalla las principales conclusiones y recomendaciones para los responsables de formular políticas.

Un gran problema de los sistemas de armas autónomos es que se activan por el entorno, es decir que el usuario no conoce o no puede elegir el objetivo, el plazo o la ubicación de la aplicación de fuerza resultante. El funcionamiento de los sistemas de armas autónomos y la imprevisibilidad en cuanto a las consecuencias de su uso pueden generar un riesgo grave para las personas civiles y retos para el cumplimiento con el DIH, así como preocupaciones éticas fundamentales sobre el papel de los seres humanos en decisiones de vida o muerte y desafíos para el mando y control militar.

Por este motivo, los límites de la autonomía en los sistemas de armas es un tema clave para abordar esas cuestiones. Tras analizar los requisitos jurídicos, éticos y operacionales del control humano, queda claro que es necesario combinar los tres tipos de medidas de control mencionados a continuación.

5 Resumen © SIPRI 2020, citado con la debida autorización.

1. *Controles sobre los parámetros de uso de los sistemas de armas*. Incluyen las medidas que restringen el tipo de objetivo o tarea que puede manejar el sistema de armas autónomo, los límites temporales y espaciales de su funcionamiento y la limitación de los efectos del sistema. Además, establecen mecanismos de salvaguardia y desactivación.
2. *Controles sobre el entorno*. Incluyen las medidas que controlan o estructuran el entorno en el que se utiliza el sistema de armas autónomo (por ejemplo, permitir su uso solo si no hay personas civiles ni bienes civiles o excluir su uso mientras dure la operación).
3. *Controles sobre la interacción entre el ser humano y la máquina*. Incluyen las medidas que permiten al usuario supervisar el sistema de armas autónomo y, de ser necesario, intervenir en su operación.

Esas medidas de control pueden reducir o, al menos, compensar la imprevisibilidad inherente al uso de los sistemas de armas autónomos y mitigar los riesgos que conlleva su uso, sobre todo para los civiles. Desde un punto de vista jurídico, el usuario debe ejercer control suficiente para tener un grado razonable de certeza respecto a los efectos del uso del sistema de armas autónomo en un ataque y la capacidad de limitar su impacto en función de lo establecido por el DIH. Las consideraciones éticas pueden exigir otras restricciones, especialmente dadas las preocupaciones que despiertan los sistemas de armas autónomos diseñados o utilizados contra las personas.

El documento concluye con cinco recomendaciones. En primer lugar, los Estados deberían definir cómo se ponen en práctica las medidas necesarias para ejercer el control humano. Como esos tres tipos de medidas de control no están atados a ninguna tecnología en particular, constituyen una sólida base normativa para regular sistemas de armas autónomos actuales y futuros.

En segundo lugar, las medidas de control humano deberían ser consideraciones esenciales cuando se establecen límites acordados internacionalmente sobre la autonomía de los sistemas de armas, desde normas y estándares hasta prácticas idóneas. Esta labor debería guiarse por los requisitos jurídicos, éticos y operacionales para ejercer el control humano. El establecimiento de normas también debería centrarse en las obligaciones y responsabilidades de los seres humanos, no en soluciones tecnológicas, para que el enfoque siga siendo relevante, práctico y adaptable a futuras tecnologías.

En tercer lugar, los Estados deberían aclarar en qué casos las normas del DIH ya establecen limitaciones para el desarrollo y el uso de sistemas de armas autónomos y en qué casos podrían necesitarse nuevas normas, estándares y prácticas idóneas.

En cuarto lugar, todas las normas, los estándares y las prácticas idóneas nuevas deberían basarse en las prácticas vigentes y en los límites existentes sobre la autonomía establecidos en el DIH. Es probable que los lineamientos para las nuevas normas, estándares y prácticas idóneas se articulen más eficazmente definiendo limitaciones a los tipos específicos de sistemas de armas autónomos y sus formas y circunstancias de uso, y los requisitos de la supervisión y la intervención humanas.

Por último, los criterios relativos al control humano deberían ser una consideración esencial para el análisis, la investigación, el desarrollo y la adquisición de nuevos sistemas de armas.

Disponible en www.icrc.org/en/document/limits-autonomous-weapons (en inglés).