Seguridad MACSEC para tráfico TSN en aplicaciones críticas de Edge Computing

Astarloa, Armando; Garcia, Diego 1,* y Salas, Sergio; Martínez Joel 2.

1 Universidad del País Vasco/Euskal Herriko Unibertsitatea. Departamento de Tecnología Electrónica. Escuela de Ingeniería de Bilbao, 48013 Bilbao; armando.astarloa@ehu.eus

2 System-on-Chip engineering S.L., Ribera de Axpe 50, 48950, Erandio; sergio.salas@soc-e.com

Resumen

Esta contribución presenta una plataforma militar certificada para computación y comunicación en tiempo real con TSN en el Edge.
El análisis de video en tiempo real para la detección de objetos es acelerado mediante la inferencia de una red neuronal YOLO mediante una unidad procesadora DPU y módulos hardware especializados en el procesamiento de video.
La comunicación de los resultados críticos se comunica a los subsistemas de defensa como tráfico con requisitos de tiempo real estrictos mediante TSN. El video post-procesado se envía a los monitores como tráfico reservado con el fin de asegurar la calidad de servicio requerido para los operadores.
En esta comunicación se presentan los resultados de latencia obtenidos para la computación y la comunicación TSN realizada en claro y protegida con MACsec.

1. Introducción

En este trabajo se describe una plataforma militar certificada diseñada para la computación y comunicación en tiempo real en el Edge.

Los resultados críticos obtenidos mediante el análisis de video en tiempo real mediante la inferencia de redes neuronales se comunican de forma segura a los subsistemas de defensa que requieren una respuesta en tiempo real utilizando *Time-Sensitive Networking* (TSN). El video post-procesado se envía a los monitores reservando un ancho de banda específico, asegurando así la calidad de servicio necesaria para los operadores.

En la Sección 2 de esta comunicación se introduce TSN, MACsec, las redes neuronales utilizadas y el caso de uso desarrollado. La Sección 3 introduce el diseño general de la plataforma y del *System-on-Chip* (SoC). La Sección 4 presenta los resultados de latencia obtenidos, finalizando la comunicación con la Sección 5 que resume las contribuciones presentadas.

2. Estado del Arte

2.1. Time-Sensitive Networking

Los sistemas actuales del sector Aeroespacio&Defensa emplean Ethernet de forma generalizada. Bien Ethernet estándar, buses de campo, Ethernet combinado con mecanismos de sincronización avanzada, Ethernet de alta disponibilidad, y más recientemente, Ethernet Determinista (AFDX [1], TSN[2]).

TSN es una propuesta integral para una única solución basada en Ethernet. La base fundamental sobre la que se fundamenta TSN es el mecanismo denominado *Time-Aware Shaper* (TAS). TAS separa la comunicación en la red Ethernet en ciclos repetitivos de duración fija. Estos ciclos están divididos en ventanas temporales de acuerdo con la configuración TSN que haya sido acordada por los nodos que conforman dicha red. Es posible configurar y asignar a cada ventana temporal una o varias prioridades Ethernet de las ocho que están disponibles (IEEE 802.1Qbv).

TAS permite definir el número de ventanas temporales presentes en cada ciclo, su duración y el tipo de tráfico (basado en su prioridad) que es posible transmitir. Gracias a este modo de operación, el tráfico *Scheduled* tiene ventanas temporales dedicadas a garantizar el comportamiento determinístico de la red. La optimización del uso del ancho de banda se realiza mediante el uso del *Credit Based Shaper* (CBS), tal y como se especifica en IEEE 802.1Qav. Esta funcionalidad posibilita la priorización del tráfico de tipo *Reserved* respecto al *Best-effort* al reservar un determinado porcentaje del ancho de banda disponible. El tráfico *Best-effort* se acomoda en el resto de las ventanas temporales de cada ciclo de operación.

La sincronización temporal en el rango de nanosegundos entre todos los dispositivos que conforman la red TSN es provista mediante el uso del protocolo de sincronización IEEE 802.1 AS.

2.2. Seguridad MACSEC

El tráfico de tiempo real *Scheduled* e incluso el tráfico *Reserved* requieren mecanismos de autenticación y encriptación de baja latencia.

MACsec [3] proporciona autenticación y confidencialidad combinadas en la Capa 2 del modelo OSI. MACsec utiliza la suite criptográfica AES-GCM para garantizar la confidencialidad y la integridad de todo el tráfico de la red. El formato de trama y el conjunto de cifrado seleccionado

en el estándar facilitan implementaciones de hardware que ofrecen un procesamiento de baja latencia y alta capacidad de ancho de banda. Por lo tanto, la comunidad técnica de TSN ha identificado esta solución de seguridad como la alternativa más viable para la seguridad de las redes de TSN.

Figura 1: *Formato de trama MACsec.*

El formato de trama MACsec se muestra en la **Figura 1**. La trama está compuesta por una etiqueta de seguridad (SecTAG), los datos seguros y un valor de verificación de integridad (ICV). Los campos de SecTAG incluyen información de seguridad para determinar está utilizando solo confidencialidad o integridad, el número de asociación, un número de paquete y la identificación del canal seguro.

MACsec establece un único canal seguro unidireccional para cada dispositivo para transmitir tramas de MACsec a sus pares dentro de la asociación de conectividad. Una asociación de conectividad consiste en dos canales seguros, uno para el tráfico entrante y otro para el tráfico saliente. Todos los pares dentro de la asociación de conectividad utilizan el mismo *cipher* de cifrado. En el estándar se definen dos *ciphers* que proporcionan tanto encriptación como autenticación: GCM-AES128 y GCM-AES-256.

2.3. Descripción del caso de uso

Edge Intelligence se refiere a la migración de tareas de computación a dispositivos en el *Edge* que generan datos en bruto. La idea detrás de la combinación de *Edge Intelligence* y las tecnologías de Redes Neuronales Profundas (DNNs) es trasladar las tareas de inferencia de las redes neuronales realizadas en la nube o en servidores *Enterprise* dedicados al *Edge*. Este paradigma permite una mayor velocidad y menor latencia de computación ya que los dispositivos procesan datos en tiempo real utilizando DNNs habilitando el despliegue de aplicaciones de tiempo real.

La aplicación presentada en esta contribución aborda la detección y localización de objetos específicos mediante el análisis de un *stream* de video en tiempo real utilizando una red neuronal convolucional (CNN) y la comunicación de los resultados de forma determinista y segura.

Las CNNs son aplicables a la detección de diferentes objetos incluidos en esa imagen [4]. Las características de estas redes las hacen muy atractivas para su inferencia en el *Edge*. En concreto, en este proyecto propone el uso de la red YOLO [5], seleccionada por su velocidad en el proceso de inferencia, característica clave para poder afrontar aplicaciones de análisis de video en tiempo real.

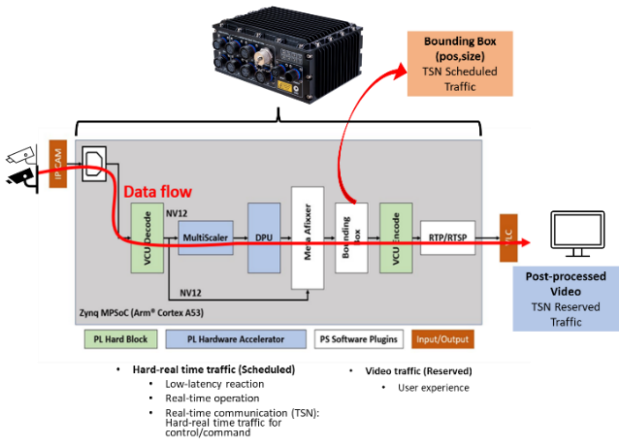

Figura 2*: Setup para la aceleración de análisis de video en Edge*

La red localiza los objetos detectados mediante la superposición de rectángulos (*Bounding box*) en el video post-procesado. Estas predicciones incluyen las coordenadas (x;y) que se refieren al centro de la caja delimitadora en relación con los extremos de la celda, la altura y el ancho en relación con toda la imagen, y la puntuación estadística sobre las probabilidades de contener un objeto. El resultado final incluye la ubicación de la caja delimitadora, la clase y la probabilidad de que el objeto pertenezca a esa clase sea correcta.

La **Figura 2** muestra el flujo de información en la aplicación desarrollada. El *stream* de video HD entra en la plataforma a través del puerto 1GbE. Un hardware específico para procesamiento de video, denominado VCU, acelera el proceso de decodificación H.264. El módulo hardware *Multiescaler* escala los *frames* del video para optimizar su uso en la red neuronal. El módulo DPU infiere la red neuronal YOLO, la cual ha sido previamente entrenada, cuantizada, optimizada y compilada para la aplicación. El software de aplicación post-procesa el video incluyendo los *bounding boxes* y los parámetros anteriormente presentados.

Estos parámetros se envían a los subsistemas de defensa como tráfico TSN *Scheduled*, puesto que se trata de tráfico con requisitos de tiempo real estricto. La codificación H.264 del video post-procesado se acelera de nuevo mediante el bloque hardware VCU. Este video, que incluirá la caja delimitadora incrustada, se transmitirá a pantallas remotas como *Reserved-traffic* para garantizar la calidad de servicio solicitada para la aplicación.

La aplicación de la seguridad MACsec al tráfico TSN se realiza forma independiente en cada puerto 1GbE mediante hardware dedicado con el fin de minimizar la potencial latencia añadida.

3. Plataforma segura de computación en el Edge con TSN seguro

La Figura 3 muestra los bloques principales de la plataforma involucrados en la comunicación TSN segura.

El SoC reconfigurable utilizado en el sistema es un AMD-Xilinx Ultrascale+ MPSoC. La sección PS representa la sección fija del circuito integrado, donde se sitúan las CPUs para el procesamiento software y un amplio número de periféricos hardware. La sección PL corresponde con la FPGA, donde se implementan los bloques hardware específicos de la aplicación. En este caso de uso, el switch TSN y los bloques hardware para la aceleración en la inferencia de la red neuronal y el procesamiento de video.

Los puertos de comunicación TSN comunican con circuitos integrados externos *Phyter* de Ethernet que proveen la interfaz física para conexión en cobre y el procesamiento MACsec integrado bidireccional.

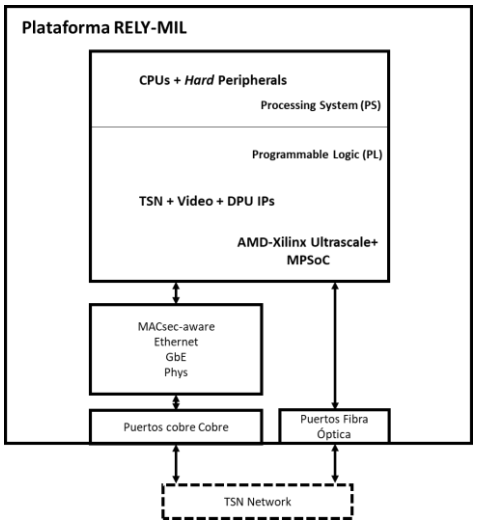

Figura 3: *Diagrama de bloques general.*

La Figura 4 detalla los bloques internos del SoC reconfigurable desarrollado. En la sección PS se hacen uso de dos controladores Ethernet integrados (MAC0 y MAC1), junto con el *cluster* de CPUs ARM Cortex A53. El software que ejecuta en estos procesadores incluye una distribución Linux como sistema operativo (Petalinux), el middleware de control de DPUs y TSN, y el software de aplicación encargado de controlar el flujo de video y su análisis.

Los módulos incluidos en la sección reconfigurable son el IP de switch TSN [6] previamente mencionado, el coprocesador de video VCU, el acelerador de la tarea de multi-escalado de *frames* de video y el coprocesador para inferencia de Redes Neuronales DPU (por sus siglas en Ingles, Deep-learning Processing Unit).

Arquitectura DPU	**Nº de cores**	**Latencia**
B4096	1	86 ms
B3136	1	120 ms
B2304	2	153 ms
B1600	2	215 ms

Tabla 1. *Latencias en el análisis de video en tiempo real.*

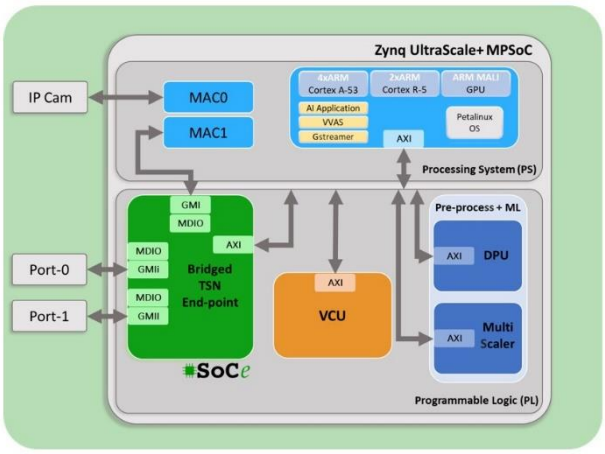

Figura 4: *Diagrama de bloques del diseño SoC.*

4. Resultados

La implementación del coprocesador DPU es configurable en tiempo de síntesis. El diseñador puede elegir distintos tamaños para ese modulo, ofreciendo distintos niveles de aceleración en función de éstos.

La Tabla 1 muestra las latencias obtenidas para el procesamiento del *stream* de video con resolución 1280x720 HD y un *framerate* de 25 FPS en el SoC con distintas configuraciones de arquitecturas DPU. Esta latencia representa el tiempo transcurrido desde la entrada del *frame* de video en el SoC hasta la generación del video post-procesado. Finalmente, para esta plataforma, tal y como se muestra en el diseño final mostrado en la **Figura 4**, la configuración usada es 1 core DPU B4096.

Desde el punto de vista de seguridad, todo el procesamiento y comunicación de datos en el interior del SoC se realiza en claro. Las tramas TSN, las cuales tienen el formato estándar Ethernet, se transmiten desde el SoC hasta el circuito integrado *Phyter* de Ethernet donde se protegen con MACsec antes de ser comunicadas a través del medio físico. En la recepción, la plataforma recibe las tramas MACsec que son autenticadas y desencriptadas antes de ser introducidas en el SoC.

La **Figura 7** muestra una captura del set-up implementado. El equipo RELYUM [7] comunica con las cámaras de video a través de un enlace GbE convencional. Los puertos TSN están conectados al equipo de pruebas IXIA.

La gráfica mostrada en la **Figura 5** representa los valores de latencia para el ensayo RFC2544 para los enlaces con interfaz física de cobre. En este ensayo se envían y reciben tramas de distinto tamaño entre el equipo de pruebas y pasando los mismos a través del equipo RELYUM.

Tal y como se puede comprobar, la latencia introducida en el camino completo aplicando un mecanismo de conmutación de paquetes *"Store&Forward"* se sitúa en el rango de 6-40 us según el tamaño de tramas. Esta latencia es aceptable incluso para las aplicaciones de control más

exigentes. Cabe destacar que la latencia introducida al proteger las comunicaciones con MACsec es similar ya que el circuito integrado *Phyter* de Ethernet emplea una *datapath* más rápido para el procesamiento hardware de las tramas cuando se aplica seguridad que cuando se realiza las comunicaciones en claro.

En el caso de utilizar puertos de comunicación TSN en claro con interfaz física de fibra óptica, las tramas de comunicación fluyen desde el SoC hasta el *transceiver* de forma directa. Para utilizar MACsec, los circuitos integrados *Phyter* de Ethernet se utilizan como coprocesadores off-chip de MACsec. Por tanto, las tramas de comunicación en claro se envían desde el SoC al *Phyter*. Este circuito devuelve las tramas al SoC protegidas, que son transmitidas al *transceiver* óptico. La recepción de las tramas MACsec siguen el mismo camino en sentido contrario.

La **Figura 6** muestra los resultados de la prueba RFC aplicado a los puertos de fibra de la plataforma. Como era esperable, el uso de los *Phyter* de Ethernet como co-procesadores off-chip de MACsec impacta en la latencia de las comunicaciones. Este impacto se reduce en el caso de que configure la operación *"Cut&Through"* para la conmutación de paquetes en el *Phyter* de Ethernet en el *loopback* implementado.

Para conocer la latencia total para el caso de uso presentado, deberán sumarse los resultados obtenidos para la computación y las comunicaciones.

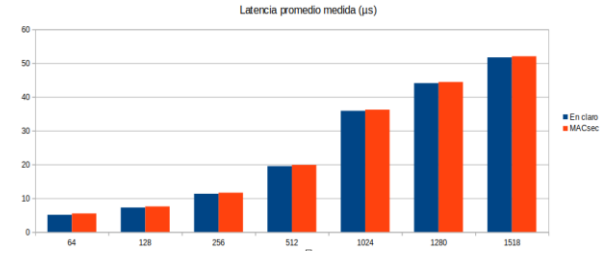

Figura 5: *Latencia en la comunicación MACsec (cobre).*

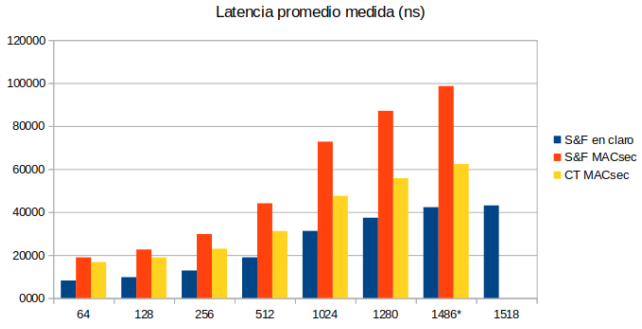

Figura 6: *Latencia en la comunicación MACsec (fibra óptica).*

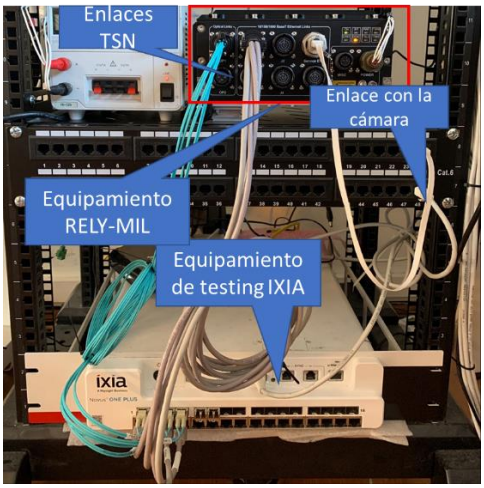

Figura 7*: Set-up de pruebas.*

5. Conclusiones y trabajo futuro

Esta contribución presenta una plataforma militar para computación y comunicación en tiempo real en el Edge.El análisis de video en tiempo real para la detección de objetos es acelerado mediante la inferencia de una red neuronal YOLO mediante una unidad procesadora DPU y módulos hardware especializados en el procesamiento de video. La comunicación de los resultados críticos se comunica a los subsistemas de defensa como tráfico con requisitos de tiempo real estrictos mediante TSN. El video post-procesado se envía a los monitores como tráfico reservado con el fin de asegurar la calidad de servicio requerido para los operadores.

Referencias

[1] ARINC Industry Activities, 2009 Arinc 664: Aircraft data network part 7 avionics FD switched ethernet network. https://standards.globalspec.com

[2] IEEE Time Sensitive Networking Task Group, 2018 IEEE 802.1 Standards. http://www.ieee802.org/1/pages/tsn.html.

[3] Institute of Electrical and Electronics Engineers (IEEE) , IEEE Std. 802.1AE-2006. IEEE standard for local and metropolitan area networks - media access control (MAC) security, 2006, https://standards.ieee.org/ieee/802.1AE.

[4] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," The Journal of Machine Learning Research, vol. 17, no. 1, pp. 1334–1373, 2016.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.

[6] System-on-Chip engineering S.L., 2018a Multiport TSN Switch IP. https://soc-e.com/mtsn-multiport-tsn-switch-ip-core/.

[7] Relyum by SoCe, "RELY-MIL-SWITCH-ROUTER High-availability. Military Switch Router," https://www.relyum.com, 2022.