Supervivencia en redes de sensores mediante técnicas multivariantes

Roberto Magán Carrión
Dpto. de Teoría de la Señal, Telemática y Comunicaciones
CITIC-UGR, Universidad de Granada
Email: rmagan@ugr.es

José Camacho Páez
Dpto. de Teoría de la Señal, Telemática y Comunicaciones
CITIC-UGR,Universidad de Granada
Email: josecamacho@ugr.es

Pedro García-Teodoro
Dpto. de Teoría de la Señal, Telemática y Comunicaciones
CITIC-UGR,Universidad de Granada
Email: pgteodor@ugr.es

***Abstract*—Actualmente las redes de sensores, y más en concreto las WSNs (*Wireless Sensor Networks*), están teniendo un gran auge debido a sus posibilidades de uso y al imparable desarrollo tecnológico. Por lo general existe una unidad de procesamiento central encargada de procesar los datos recibidos desde todos y cada uno de los sensores. Asegurar la integridad de dichos datos puede resultar crucial en ciertos ámbitos de utilización; por ejemplo, en la detección y extinción de incendios, donde normalmente están en juego vidas humanas. En este contexto, en el presente trabajo se aborda el problema de la modificación maliciosa de datos mediante la propuesta de un sistema de detección de anomalías e imputación de datos faltantes en base al análisis multivariante, explotando la densidad espacial y temporal de este tipo de redes con el objetivo de conseguir la supervivencia de la red.**

I. Introducción

En los últimos años, y gracias al avance en las tecnologías MEMS (*Micro-Electro-Mechanical Systems*), las redes de sensores inalámbricas o WSN (*Wireless Sensor Networks*) han tenido un gran auge [1]. Hoy es posible crear dispositivos más inteligentes, eficientes, baratos y con gran capacidad de comunicación, cuya configuración y uso tienen su límite casi exclusivamente en la imaginación.

Una red de sensores consiste en la agrupación estructurada o no [2] de cientos o miles de sensores con el objetivo de monitorizar, a través de mediciones específicas, un área o zona concreta. De manera general, existe una unidad de procesamiento central que recoge y procesa todos los datos que provienen de los sensores. Estos datos pueden ser enviados directamente a dicha unidad o encaminados a través de la red mediante la colaboración de todos los dispositivos (*multi-hop routing* [2]). También son usuales las agrupaciones (*clusters*) de sensores para proporcionar agregación de datos y que sea el nodo gestor del *cluster* el encargado de agrupar y enviar éstos. Se consigue así mayor eficiencia energética, un bien muy preciado en este tipo de redes.

Dos son los tipos de aplicaciones principales de las redes de sensores: la monitorización y el seguimiento (*monitoring and tracking*). Tanto en un caso como en el otro, las redes WSN poseen su ámbito de utilización dentro de numerosos campos como el militar, el médico y/o el industrial [1].

Aunque tradicionalmente se asume que estas redes son fijas, la introducción de movilidad aporta numerosas ventajas en cuanto a conectividad, coste, fiabilidad, eficiencia energética, etc. [3]. Se entronca así con las denominadas, bien conocidas y ampliamente usadas redes MANET (*Mobile Ad-hoc NETworks*).

Dado el amplio uso de las redes de sensores (móviles o no) en entornos hostiles (acciones militares, gestión de desastres naturales y/o medioambientales, etc.), se hace necesario la provisión de seguridad en el entorno. Existen así muchas e inherentes amenazas relativas a las redes WSN [4]. En concreto, y a modo de ejemplo, dentro del marco de uso de la detección y extinción de incendios, se pueden mencionar algunos ataques cuyo impacto resulta significativo.

Así, el ataque *data tampering, environmental tampering* o *tampering* [4] [5], que afecta a la integridad de los datos, puede tener consecuencias desastrosas si no se mitiga o detecta a tiempo. Como ejemplo, se podría dar el caso en que una persona (usualmente un pirómano) quisiera provocar fuego dentro del área que está siendo monitorizada. Dicha persona sólo tendría que alterar físicamente la medición proporcionada por uno o varios sensores con el propósito de desviar la atención de los servicios de extinción, de tal manera que el verdadero incendio (localizado en una zona distinta) evolucionara sin recibir la adecuada atención de los equipos de emergencias.

También, y relacionado con la propia comunicación entre sensores, más en concreto con el algoritmo de enrutamiento, aparece el ataque *drop* o *packet dropping* [4]. Imaginemos que toda la información de un grupo de sensores tiene que ser encaminada necesariamente (debido al algoritmo de enrutamiento) a través de un nodo comprometido, y el comportamiento de éste es descartar paquetes. Esto provocaría la no disponibilidad de los datos del área afectada. Volviendo al ejemplo forestal, ello podría significar la ocurrencia de un incendio sin que los equipos de extinción tuvieran consciencia de la situación, lo que retrasaría en definitiva la actuación adecuada sobre el incidente.

Se hace evidente en consecuencia, la necesidad de desplegar sistemas de seguridad robustos ante dichas amenazas, proveyendo de mecanismos adecuados para la corrección/obtención de los datos, contribuyendo así a la supervivencia del sistema, entendida ésta como *"la habilidad del sistema para conseguir sus objetivos en el tiempo y en presencia de ataques, fallos o accidentes"* [6].

En este contexto, se introduce aquí el uso de los métodos de análisis multivariante para la mejora de la seguridad en entornos WSN. Por una parte, PCA (*Principal Component Análisys*) [7], para la detección y monitorización de anomalías, y, por otra, TSR (*Trimmed Scores Regression*) [8] [9], para la imputación de datos faltantes.

Con estos métodos se persigue la detección supervisada de incidentes de seguridad, en concreto el ataque *data tampering*, así como la recuperación de datos alterados mediante la explotación de la correlación espacio-temporal entre sensores.

El resto del documento se organiza de la siguiente manera. En la Sección II se exponen algunos trabajos relacionados con la temática aquí planteada. La explicación teórica de los métodos de análisis multivariante empleados se expone en la Sección III. La simulación y los escenarios de uso se discuten en la Sección IV. En la Sección V se expone la capacidad de estos métodos en la detección de anomalías e incidentes de seguridad. En la Sección VI se aplican estos procedimientos para la imputación de datos como mecanismo de respuesta y, en definitiva, como propuesta de supervivencia. Por último, se finaliza el trabajo presentando en la Sección VII las principales conclusiones de los desarrollos realizados.

II. Trabajos relacionados

La aplicación de metodologías de análisis multivariante enfocadas a la seguridad en redes de comunicaciones es algo relativamente nuevo, si bien no ocurre lo mismo en el ámbito de los procesos industriales. Por ejemplo, en [10] se propone un sistema de detección de anomalías basado en PCA comparando distancias de Mahalanobis. Por su parte, en [11] se expone un sistema de detección de intrusiones también basado en componentes principales.

Este tipo de análisis encaja muy bien dentro del contexto de redes de sensores con alta densidad tanto espacial como temporal. Esto es así debido a la alta correlación existente en el sistema, característica que explota el análisis multivariante.

Algunos otros son los trabajos que aprovechan también esta singularidad de este tipo de redes. Por ejemplo, en [12] los autores proponen un método de detección de anomalías e imputación de datos faltantes empleando redes bayesianas que emplean la correlación espacial y temporal de estas redes.

En relación a la imputación de datos faltantes, es de mencionar que estas técnicas se han empleado habitualmente en el ámbito de la gestión de procesos industriales [13] [14], aunque también existen trabajos recientes en el contexto de las WSNs, como en [15], para la detección de anomalías e imputación de datos, o en [16], donde se presenta un método robusto ante la pérdida de datos.

III. Análisis multivariante para datos faltantes

La condición de multi-variabilidad está implícita en la propia naturaleza. Normalmente, para caracterizar un fenómeno natural físico es necesaria la intervención de varios factores (variables) estrechamente relacionados con él. Por ejemplo, la predicción del tiempo es muy dependiente del viento, la presión atmosférica y la temperatura, entre otros.

La descripción de datos y el modelado de su estructura, la discriminación y clasificación, así como la regresión y predicción [7], son escenarios de uso habituales para estas técnicas.

A. PCA: Principal Components Analysis

El objetivo principal de PCA es el modelado del espacio inicial de variables $\mathbf{X}$ en otro de menor dimensión en el que se concentra la información relevante (variabilidad) de las variables iniciales. Este modelo se puede entender como un cambio de variables, denominándose a las nuevas variables PCs (*Principal Components*). Sea $\mathbf{X}$ una matriz de datos, que contiene p variables medibles de un determinado fenómeno y n observaciones de dichas variables. Así, $\mathbf{X}$ tendrá dimensión $(n \,\mathrm{x}\, p)$. El modelo PCA sigue la expresión:

$$\mathbf{X} = \mathbf{T}_A \cdot \mathbf{P}_A^T + \mathbf{E}_A \tag{1}$$

donde A es el número de variables (PCs) del modelo, $\mathbf{T}_A(n \,\mathrm{x}\, A)$ es la matriz de puntuaciones (*scores*) que contiene las proyecciones de los puntos del espacio inicial sobre el nuevo espacio y a través de la cual podemos extraer información acerca de las observaciones o filas de nuestra matriz original $\mathbf{X}$. $\mathbf{P}_A(p \,\mathrm{x}\, A)$ es la matriz de cargas, cuyas columnas son combinaciones lineales de las variables iniciales. Esta matriz proporcionará información sobre la estructura de relación de las variables originales (diagrama de *loadings*). $\mathbf{E}_A(n \,\mathrm{x}\, A)$ es la matriz que contiene el error residual cometido en la aproximación y jugará un papel muy importante en la detección de anomalías correspondientes a los datos obtenidos de los sensores.

Monitorización multivariante: límites de control

Durante la monitorización usualmente se utilizan los estadísticos Q, que comprimen los residuos obtenidos en cada observación, y T^2 [17], que se obtiene de las puntuaciones. Con estos estadísticos, calculados a partir del modelado PCA en condiciones normales, se establecen unos límites de control para un determinado intervalo de confianza. Posteriormente, nuevas observaciones pueden ser contrastadas frente a los límites de Q y T^2, concluyéndose la aparición de anomalías cuando estos límites son rebasados. Adicionalmente, la contribución de las variables a cierta anomalía detectada puede ser investigada utilizando los gráficos de contribución [18]. Los estadísticos Q y T^2 se calculan para una determinada observación de acuerdo a las siguientes expresiones:

$$T_i^2 = \sum_{a=1}^{A} \left(\frac{\tau_{ai} - \mu_a}{\sigma_a} \right)^2 \tag{2}$$

$$Q_i = \sum_{j=1}^{J} (e_{ij})^2 \tag{3}$$

donde τ_{ai} representa la puntuación de la i-ésima observación en la a-ésima variable, μ_a y σ_a son la media y la desviación típica de las puntuaciones en dicha variable, respectivamente, y

e_{ij} representa el residuo correspondiente a la j-ésima variable de la i-ésima observación.

B. TSR: Trimmed Scores Regression

Varios son los métodos y propuestas que usan PCA en el contexto de datos faltantes, formándose dos grupos principales: los basados en regresión y los no basados en regresión [8]. Los primeros mejoran el rendimiento y tienen menor coste computacional [9]. Uno de estos métodos es TSR, que se utilizará aquí por las ventajas mencionadas.

TSR estima los valores de las puntuaciones o *scores* a partir de las puntuaciones incompletas, por ejemplo, rellenando las puntuaciones faltantes con ceros. Así, los valores faltantes de una observación incompleta x_{inc} son sustituidos por ceros. Podemos reordenar dicha observación como sigue:

$$x_{inc} = [x_1, \ldots, x_k, 0, \ldots, 0]^T \tag{4}$$

Las puntuaciones incompletas de x_{inc} con calculadas directamente usando el modelo PCA:

$$\tau_A^* = \mathbf{P}_A^T \cdot x_{inc} \tag{5}$$

Equivalentemente:

$$\tau_A^* = (\mathbf{P}_{A,k})^T \cdot x_{inc}^* \tag{6}$$

donde:

$$\mathbf{P}_{A,k}^* = \begin{bmatrix} p_{1,1} & \cdots & p_{A,1} \\ \vdots & \ddots & \vdots \\ p_{1,k} & \cdots & p_{A,k} \end{bmatrix} \tag{7}$$

$$x_{inc}^* = [x_1, \ldots, x_k]^T \tag{8}$$

siendo $p_{a,j}$ la carga correspondiente a la j-ésima variable en la a-ésima componente principal. Así, en (4) se completan las observaciones que faltan de cada una de las variables con el valor 0, a diferencia de (5), en donde únicamente se utilizan las variables disponibles en x_{inc}. Los datos de calibración $\mathbf{X}$ pueden ser usados para la estimación de los elementos faltantes en x_{inc} como sigue. Sea $\mathbf{X}^*$ una sub-matriz de $\mathbf{X}$ que contendrá únicamente las variables que están disponibles en x_{inc}. De esta manera, la matriz de puntuaciones incompletas correspondiente a los datos de calibración se puede obtener como sigue:

$$\mathbf{T}_A^* = \mathbf{X}^* \cdot \mathbf{P}_A^* \tag{9}$$

A partir de $\mathbf{T}_A$ y $\mathbf{T}_A^*$ se deriva la expresión:

$$\mathbf{T}_A = \mathbf{T}_A^* \cdot \mathbf{B} + \mathbf{F} \tag{10}$$

donde la matriz de coeficientes de regresión $\mathbf{B}$ se puede obtener aplicando mínimos cuadrados, ya que, típicamente, la inversión de $(\mathbf{T}_A^*)^T \cdot \mathbf{T}_A^*$ estará bien condicionada. En cualquier caso, si no lo está, podemos estimar $\mathbf{B}$ con métodos como PLS (*Partial Least Squares*) y PCR (*Principal Component Regression*) [7]. Después, $\mathbf{B}$ es usada para mejorar la estimación de las puntuaciones en (6):

$$\tau_A^{TSR} = (\mathbf{P}_A^* \cdot \mathbf{B})^T \cdot x_{inc}^* \tag{11}$$

Finalmente, las puntuaciones τ_A^{TSR} se pueden usar para estimar la observación inicial que contenía datos faltantes en la que se corrigen éstos:

$$\hat{x} = \mathbf{P}_A \cdot \tau_A^{TSR} \tag{12}$$

TSR basa su eficiencia en la asunción de correlación entre las variables iniciales, ya que los datos faltantes de una observación para varias variables se calculan a partir de valores ya existentes de otras variables. Un conjunto suficiente de muestras de las variables resulta también deseable, de tal manera que se consiga un mejor ajuste y calibración del modelo PCA del que se parte en inicio.

IV. Escenario de uso y simulación

En las redes de sensores es usual la existencia de una unidad de procesamiento central encargada de recoger y procesar la información de todos los sensores de la red. Aquí nos centraremos en el conjunto de datos recogidos por esta unidad.

Partimos de una red de 100 sensores homogéneamente distribuidos dentro de un área de 1 km^2 de superficie forestal. Cada uno de ellos efectuará mediciones de temperatura durante ciertos intervalos de tiempo y enviará dichos resultados a la unidad encargada del procesamiento. Esta distribución se escoge de manera análoga al sistema real proporcionado por la empresa Libelium [19].

Existen varios entornos de simulación exclusivos de WSNs [20], aunque usualmente sólo se centran en aspectos relacionados con la propia red (modelos de capa física, protocolos, propagación, etc.) dejando de lado la simulación del entorno propiamente dicho, por ejemplo, aspectos como el lugar donde se ubica la red y/o sus magnitudes físicas. Por este motivo, hemos desarrollado un simulador específico en el que se calcula la evolución de las temperaturas en un área concreta en presencia o no de fuego usando Matlab 2009b. Para ello nos basamos en la idea de [21], dónde se plantea un método de obtención de la temperatura que recibe cada sensor debido a la contribución de cada uno de los focos de fuego más cercanos a éste. Otro aspecto importante que se introduce en este trabajo es la aproximación de dichos focos mediante distribuciones gaussianas bi-variantes, idea de la que se partirá para emular tanto la temperatura que recibe un sensor en condiciones normales como en presencia de fuego. En la Figura 1 podemos observar el entorno de simulación considerado con los focos de temperatura en condiciones normales y de fuego, así como la distribución de los sensores.

El principal objetivo del simulador es realizar una prueba de concepto de la propuesta de supervivencia de este trabajo. Para calibrar el modelo inicial PCA nos basamos en una matriz $\mathbf{X}$ que contiene 100 observaciones (instantes de muestreo) de 100 variables (temperatura recogida en cada uno de los sensores)

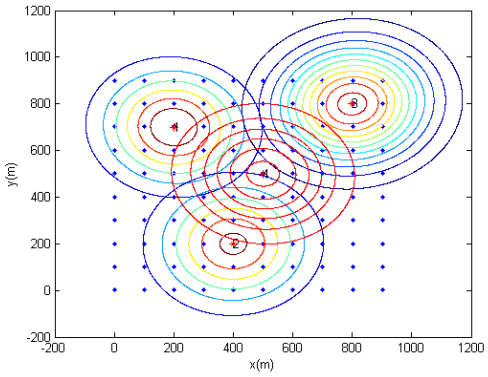

Fig. 1: Escenario de simulación. Se observa la distribución de los sensores, los focos de temperatura y un foco de fuego (en rojo y centrado).

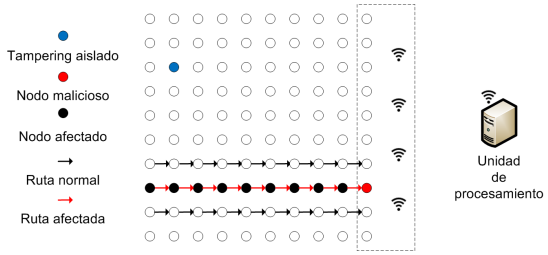

Fig. 2: *Tampering* aislado y en línea. Un nodo malicioso (en rojo) modifica los datos que provienen de los demás sensores.

en condiciones de temperatura normal, es decir, en ausencia de fuego.

Serán simulados varios casos posibles de ataques *data tampering*. En la Figura 2 se muestra un sensor afectado de manera aislada. También se aprecia el caso de que se vea afectada toda una línea de sensores debido al comportamiento malicioso del sensor más cercano a la unidad de procesamiento. Otro caso a estudiar es el *data tampering* en grupo como consecuencia del comportamiento anómalo del nodo CH (*Cluster Head*) durante la agregación de datos (Figura 3). Este último ejemplo coincide con métodos de encaminamiento que se usan en la actualidad, como el algoritmo de encaminamiento LEACH [2].

V. Análisis y detección de anomalías

Para efectuar la monitorización del sistema se han utilizado herramientas del PLS-toolbox de Matlab [22], específico para

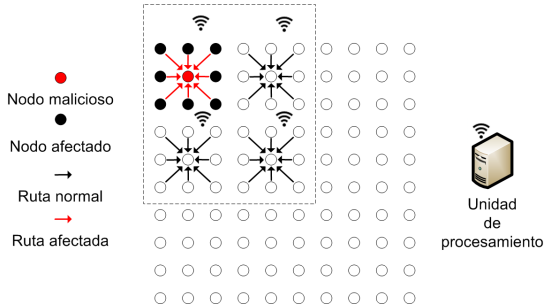

Fig. 3: *Tampering* en grupo. Un nodo malicioso (en rojo) modifica los datos que provienen de los demás sensores.

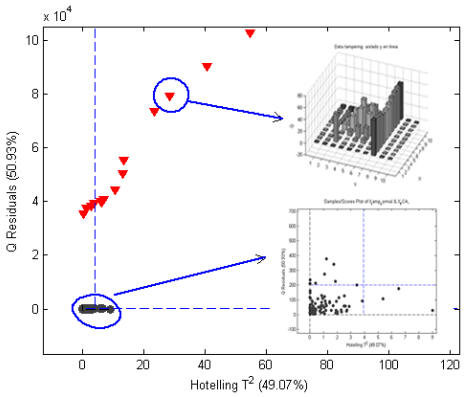

Fig. 4: Gráfico de monitorización: Modelo inicial (círculos en gris oscuro), con detalle de los límites de control (en azul y línea discontinua). Anomalía detectada (triángulos rojos). Detalle de la contribución a Q de una observación concreta (arriba a la derecha)

el análisis multivariante.

Confrontando el modelo PCA inicial con respecto a las observaciones que se están generando durante el proceso de monitorización, el sistema es capaz de indicar cuándo se está produciendo una anomalía. Esto es posible gracias a gráficos de monitorización como el que se presenta en la Figura 4, en donde se observa cómo evoluciona el sistema. Las nuevas observaciones (triángulos rojos) denotan la presencia de una anomalía.

Así, un observador que supervisa el sistema será alertado por éste cuando alguna de las muestras se desvíe lo suficiente como para rebasar los límites de control. Se han elegido estos límites de tal manera que el 95% de las muestras recogidas en la etapa de modelado y calibración quedasen por debajo de dichos límites.

En este punto no se sabe si se está produciendo un ataque o es el fuego el que está actuando. Para distinguir qué motivo originó dicho comportamiento, serán de ayuda los gráficos de contribución obtenidos a partir de los estadísticos Q, T^2, introducidos en el apartado de análisis multivariante.

Se usará la contribución a Q para la detección de ataques de *data tampering*, ya que proporcionará una visión de cuál es la discrepancia de una variable (sensor) con respecto al modelo inicial.

Una vez se ha detectado una anomalía, el sistema de monitorización muestra el correspondiente gráfico de contribución de la observación que produjo dicha anomalía. En este gráfico, el patrón típico de un fuego se distingue del patrón que sigue un ataque malicioso, como se ilustra a continuación.

En las Figuras 6 y 7 se observa que la aportación de aquellos sensores al estadístico Q en los que se está produciendo el ataque se desvía con respecto al perfil o patrón de contribución en presencia de fuego únicamente (Figura 5). La persona encargada de supervisar el sistema debe ser capaz de distinguir entre ambos patrones. Nótese que para que una persona malintencionada pueda lograr obtener un patrón parecido al fuego, es necesario comprometer un número muy elevado de sensores.

De la Figura 6 se puede concluir que estamos ante la

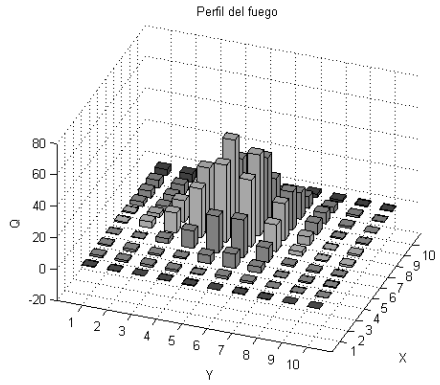

Fig. 5: Perfil generado por una observación correspondiente únicamente a la presencia de fuego.

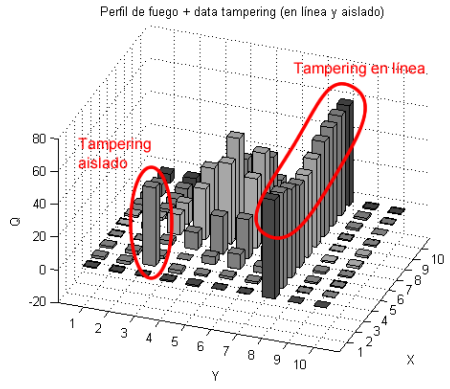

Fig. 6: Perfil generado por la misma observación que en la Figura 5, con la presencia de fuego y *data tampering* en una línea de sensores y en un sensor aislado.

presencia de un ataque *data tampering* que afecta a una línea completa de sensores (debido al uso del encaminamiento en la Figura 2) y a un sensor aislado. De forma análoga, en la Figura 7, vemos cómo un ataque de *data tampering* afecta a un grupo de sensores de tamaño 3x3, al comprometer un *Cluster Head* (Figura 3). En ambos casos, el patrón de temperaturas dista mucho del patrón gaussiano típico de un fuego.

VI. Imputación de datos y supervivencia de la red

Más allá de la detección de ataques, están las actuaciones que se deberían llevar a cabo para contribuir a la supervivencia del sistema, en el sentido de proporcionar un mecanismo de respuesta que mitigase el efecto del ataque detectado. La propuesta de este artículo es tratar como valores faltantes los datos detectados como fraudulentos y estimar su valor original con técnicas de imputación.

En esta sección se evaluará la técnica de imputación de datos TSR para dos casos concretos: *data tampering* sobre una línea de sensores y un sensor aislado (Figura 6), así como sobre un grupo de sensores agrupados en forma de celda de 3x3 (Figura 7).

A. Data tampering: línea de sensores y un sensor aislado

Tras la aplicación del método de imputación de datos sobre los ataques que se producen en la Figura 6, se obtiene el

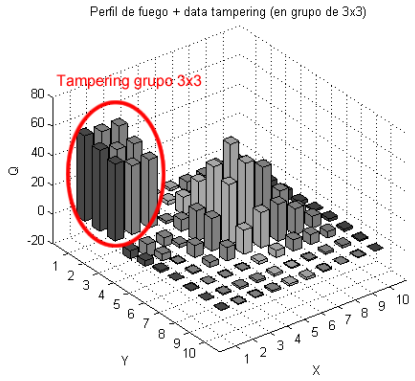

Fig. 7: Perfil generado por la misma observación que en la Figura 5, con la presencia de fuego y *data tampering* en grupo de sensores de tamaño 3x3.

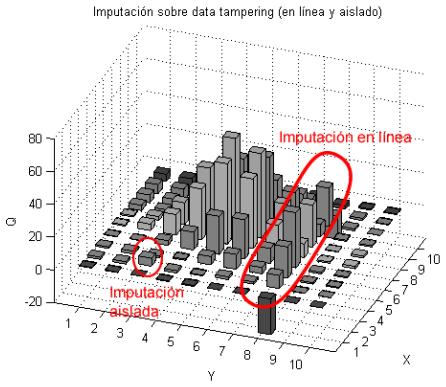

Fig. 8: Resultado de la imputación para *data tampering* aislado y en línea

resultado de la Figura 8. Se observa cómo el método es capaz de estimar el valor actual del sensor atacado de forma aislada (comparar con la Figura 6). Su reducción es notable, como esperábamos que fuese, ya que los sensores vecinos poseen valores muy cercanos a cero.

Notamos también la adecuación de los valores que corresponden a la línea de sensores. Se percibe una adaptación de éstos a la estructura que sigue el patrón de fuego, sobre todo en la parte central de la línea, zona afectada por éste. La raíz cuadrada del error cuadrático medio en la estimación es de 48,6419.

B. Data tampering: grupo de sensores

En este caso tendremos una situación similar a la anterior, difiriendo únicamente en que la agrupación de sensores se produce en la esquina superior izquierda de la red de sensores, tal y como se muestra en la Figura 7.

En la Figura 9 se observa cómo se reduce notablemente la contribución de los sensores. Esto se puede ver comparando con la Figura 7. En este caso, la raíz cuadrada del error cuadrático medio es 60,3301, algo mayor al del apartado anterior. Esto es debido a que el método es capaz de recuperar mejor los datos cuando se tiene más información de correlación disponible, situación que ocurre cuando es una línea de sensores la afectada. En el presente caso no existe tanta

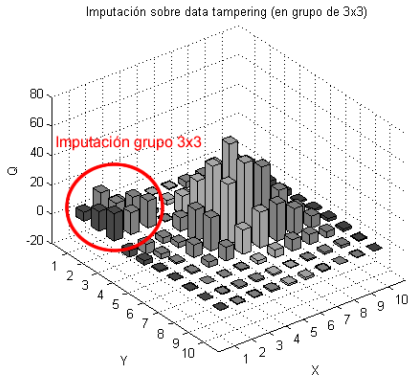

Fig. 9: Resultado de la imputación para *data tampering* agrupado con tamaño 3x3.

información de correlación disponible útil para la estimación. De tal manera que, por ejemplo, el sensor situado en mitad del grupo no dispondrá de datos válidos de correlación de los vecinos inmediatos porque éstos también están afectados por el ataque.

VII. Conclusiones

A partir de los resultados obtenidos, se ha comprobado la capacidad de detectar y dar respuesta a ataques de *data tampering* usando esquemas basados en el análisis multivariante, abriendo así una nueva línea de actuación en la mejora de la supervivencia en redes de sensores.

A pesar de que sólo se ha tratado un ataque concreto, se intuye que éstos métodos podrían ser válidos para otros ataques, en concreto *dropping* y *delay*. En particular para este último, habida cuenta de la correlación temporal de los datos [23].

Se ha comprobado la dependencia de estos métodos con el algoritmo de encaminamiento usado para el transporte de los datos. Proponemos así como línea futura de investigación la creación o mejora de métodos de encaminamiento de cara a la preservación de la información de correlación de la red. Éste puede ser un problema complejo en especial en el contexto de redes de sensores móviles.

Otra línea futura interesante es la mejora de los resultados de imputación utilizando otras organizaciones de los datos, como los propuestos en el análisis multivariante de imágenes [24]. Algunos resultados preliminares reflejan el interés en esta posibilidad.

También se plantea como trabajo futuro la realización de un simulador que se adapte más fielmente a escenarios naturales de aplicación, dentro del contexto de la detección y extinción de incendios.

Finalmente, estamos trabajando en la disposición de mecanismos automáticos de detección que permitan una mayor capacidad de respuesta en base a la generación de alarmas ante la ocurrencia de incidentes de seguridad, sin la intervención manual del hombre (o como complemento de ésta).

Agradecimientos

Este trabajo ha sido parcialmente financiado por el MICINN a través del proyecto TEC2011-22579.

Referencias

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6– 28, Dec. 2004.

[3] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," *ACM Trans. Sen. Netw.*, vol. 8, no. 1, pp. 7:1–7:31, 2011.

[4] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 2, pp. 52–73, 2009.

[5] A. D. Wood and J. A. Stankovic, "Denial of service in sensor networks," *IEEE Computer*, vol. 35, no. 10, pp. 54–62, Oct. 2002.

[6] M. Lima, A. dos Santos, and G. Pujolle, "A survey of survivability in mobile ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 11, pp. 66–77, 2009.

[7] K. H. Esbensen, *Multivariate Data Analysis - in practice*, Esberg, Aalborg University, Ed. CAMO, 2009.

[8] F. Arteaga and A. Ferrer, "Dealing with missing data in MSPC: several methods, different interpretations, some examples," *Journal of Chemometrics*, vol. 16, no. 8-10, pp. 408–418, Aug. 2002.

[9] ——, "Framework for regression-based missing data imputation methods in on-line MSPC," *Journal of Chemometrics*, vol. 19, no. 8, pp. 439–447, Aug. 2005.

[10] N. Chitradevi, V. Palanisamy, K. Baskaran, and U. B. Nisha, "Outlier aware data aggregation in distributed wireless sensor network using robust principal component analysis," in *2010 International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, Jul. 2010, pp. 1–9.

[11] M. A. Livani and M. Abadi, "A PCA-based distributed approach for intrusion detection in wireless sensor networks," in *2011 International Symposium on Computer Networks and Distributed Systems (CNDS)*. IEEE, Feb. 2011, pp. 55–60.

[12] E. W. Dereszynski and T. G. Dietterich, "Spatiotemporal models for Data-Anomaly detection in dynamic environmental monitoring campaigns," *ACM Trans. Sen. Netw.*, vol. 8, no. 1, pp. 3:1–3:36, 2011.

[13] J. Flores-Cerrillo and J. F. MacGregor, "Latent variable MPC for trajectory tracking in batch processes," *Journal of Process Control*, vol. 15, no. 6, pp. 651–663, Sep. 2005.

[14] J. Camacho, J. Picó, and A. Ferrer, "Bilinear modelling of batch processes. part II: a comparison of PLS soft-sensors," *Journal of Chemometrics*, vol. 22, no. 10, pp. 533–547, Oct. 2008.

[15] Y. Li and L. E. Parker, "A spatial-temporal imputation technique for classification with missing data in a wireless sensor network," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008. IROS 2008*. IEEE, Sep. 2008, pp. 3272–3279.

[16] J. C. Lim and C. J. Bleakley, "Robust data collection and lifetime improvement in wireless sensor networks through data imputation," in *2010 Fifth International Conference on Systems and Networks Communications (ICSNC)*. IEEE, Aug. 2010, pp. 64–69.

[17] H. Hotelling, *Multivariate Quality Control. Techniques of Statistical Analysis*, New York, Ed. MacGraw-Hill, 1947.

[18] T. Kourti and J. MacGregor, "Multivariate spc methods for process and product monitoring," *Journal of Quality Technology*, vol. 28, 1996.

[19] Libelium. Last accessed: Feb. 2012. [Online]. Available: http://www.libelium.com/wireless_sensor_networks_to_detec_forest_fires/

[20] D. H. A. Kellner, K. Behrends, "Simulation environments for wireless sensor networks," IFI-TB, Tech. Rep., 2010.

[21] G. X. E. S. Manolakos, D. V. Manatakis, "Temperature field modeling and simulatioin of wireless sensor network behaviour during a spreading wildfire," in *16th European Signal Processing Conference (EUSIPCO 2008)*. EURASIP, 2008.

[22] PLSToolbox 3.5 for use with Matlab. Last accessed: Feb. 2012. [Online]. Available: http://www.eigenvector.com/software/pls_toolbox.htm

[23] J. Camacho, J. Picó, and A. Ferrer, "Bilinear modelling of batch processes. part i: theoretical discussion," *Journal of Chemometrics*, vol. 22, no. 5, pp. 299–308, May 2008.

[24] M. H. Bharati, J. Liu, and J. F. MacGregor, "Image texture analysis: methods and comparisons," *Chemometrics and Intelligent Laboratory Systems*, vol. 72, no. 1, pp. 57–71, Jun. 2004.