# Análisis de incidentes informáticos en base a una distribución Bayesiana

Carlos A. Talay, Griselda Rojas, Carlos D. Amarilla, Dora S. Maglione, José L. Saenz

Unidad Académica Río Gallegos / Universidad Nacional de la Patagonia Austral

Lisandro de la Torre 1070, Río Gallegos, Santa Cruz (9400), Argentina

carlostalay@yahoo.com.ar, patagoniaustral@gmail.com, carlos.amarilla@gmail.com, dmaglione@disytel.com, jsaenz_lacaze@hotmail.com

Luis A. Marrone

L.I.N.T.I. – Universidad Nacional de La Plata

Calle 50 y 115, 1er. Piso, Edificio Bosque Oeste, La Plata (1900), Pcia. de Bs. As., Argentina

lmarrone@info.unlp.edu.ar

## Resumen

La evaluación de los incidentes informáticos es el primer paso para el diagnóstico y el posterior diseño de una estrategia efectiva que permita tomar medidas tendientes a reducir estos eventos. Este trabajo propone una metodología para realizar el tratamiento de los incidentes que se registran normalmente en centro de cómputos y que permita definir políticas tendientes a evitarlos.

**Palabras clave:** Incidentes informáticos, redes bayesiana, prevención de incidentes

## Contexto

El presente proyecto se encuentra dentro de la línea de investigación de S*eguridad Informática*, que desarrolla un grupo de investigadores de la UNPA-UARG conformado y en consolidación, financiado con fondos propios de la Unidad Académica Río Gallegos, destinados a proyectos de investigación acreditados. Este grupo de investigación cuenta, en esta ocasión, con la dirección del Ing. Luis Marrone, perteneciente a la UNLP.

## Introducción

El relevamiento y evaluación de incidentes informáticos dentro de un centro de cómputos es, conjuntamente con la optimización de los procedimientos, el camino que posibilita la mejora en el funcionamiento de una organización. El correcto tratamiento de los incidentes en un centro de cómputos posibilita evitar la eventual pérdida de tiempo e información, que indudablemente generará una mayor o

menor merma en la productividad dependiendo del grado del incidente.

La meta de toda organización es realizar el trabajo de la manera más simple y eficiente, con el menor riesgo posible. Si bien este criterio está aceptado, la implementación de una política eficiente en lo que respecta al manejo de incidentes no es, aún, una práctica generalizada. El registro y la documentación de los incidentes, y su posterior análisis permite la toma de acciones proactivas para evitar perdida de información, daños físicos sobre equipos o personas y la consiguiente pérdida de bienes y productividad en una organización en donde el centro de cómputos tienen un rol importante en la gestión de la información. Teniendo en cuenta esto y con vistas a poder generar un ambiente de trabajo seguro se debe comenzar por un correcto relevamiento de los incidentes que se producen, lo que implica poseer un registro histórico con los datos que caractericen estos incidentes y que permita hacer seguimiento de los mismos en forma continua. Luego, en base al análisis de estos datos se podrá actuar en consecuencia de acuerdo a las necesidades y políticas de una organización y de esta manera se realizarán correcciones sobre problemas de tipo puntual o crónico. En este sentido las estadísticas clásicas permiten hacer un análisis cuantitativo respecto a un parámetro, pero las redes bayesianas en cambio, permiten hacer un análisis de la interrelación existente entre la distribución de los incidentes y el efecto que estos causan sobre el desempeño de la organización.

## Las Redes Bayesianas

Las redes bayesianas, también conocidas como redes de creencias, constituyen una herramienta estadística que codifica relaciones probabilísticas entre un grupo de variables aleatorias de interés [1] [2]. Esta herramienta estadística es utilizada en los más variados campos de la ciencia [3] [4] [5] [6] [7] y ha mostrando ser una valiosa ayuda en el diseño de estrategias para la toma de decisiones y la elaboración de sistemas expertos. Cuando se utiliza en conjunción con técnicas estadísticas y datos concretos provenientes de conocimientos previos [8], el modelo-gráfico resultante proporciona varias ventajas al momento de realizar un análisis de las relaciones entre las causas y los efectos de las sucesivas etapas de un proceso aleatorio. Una de estas ventajas es que, como el modelo codifica dependencias entre todas las variables relacionadas mediante el vínculo de causa-efecto en las fases secuenciales del proceso, se puede obtener en forma dinámica la vinculación de las variables que afectan un evento y se pueden realizar acciones proactivas tomadas en función del análisis en las tendencias del comportamiento de esas variables, en un sentido de inferencia.

Otro aspecto interesante es que una red bayesiana se puede utilizar como sistema experto para aprender por sí mismo a perfeccionar el conocimiento de las relaciones causales, y por lo tanto, se puede utilizar para obtener una mejor comprensión sobre el dominio de un problema y predecir de las consecuencias de la intervención.

En tercer lugar, el modelo permite revelar la relación entre las causales y los efectos de los eventos que se registran y es dinámico en cuanto a que puede modificarse y eventualmente extenderse a través de la incorporación de nuevas ramas que vinculen otros efectos y causas no previstos inicialmente. Pudiéndose entonces, orientar en forma dinámica el tratamiento de determinados tipos de

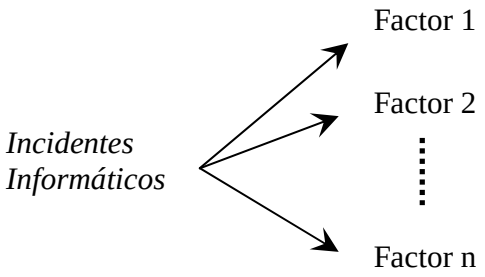
incidentes que generen efectos específicos que tengan un interés particular de estudio.

Bajo estas consideraciones vemos que la semántica probabilística, es una representación ideal para combinar el conocimiento previo (que a menudo viene dado en forma causal) y los datos disponibles, pudiéndose de esta manera llegar a la modelización de una serie de eventos y en consecuencia determinar cómo estos eventos afectan el comportamiento de un sistema.

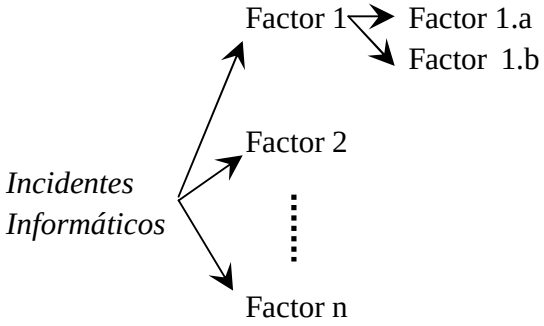
## Líneas de Investigación, Desarrollo e Innovación

La línea de investigación abordada en este proyecto es la de seguridad informática en su aspecto particular de clasificación y tratamiento de incidentes informáticos. Este proyecto mantiene y amplia la línea de investigación iniciada en un proyecto de investigación anterior, iniciado en la UNPA-UARG, con la identificación 29A198 y acreditado en esa misma unidad. Ese proyecto se basó en el estudio las técnicas para el manejo seguro de información dentro de un centro de cómputos, aspecto íntimamente vinculado con el estudio de las vulnerabilidades y los incidentes informáticos.

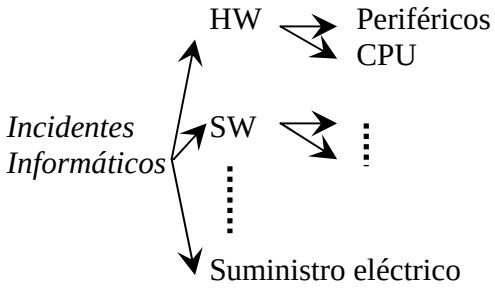
Para el desarrollo de este proyecto se buscó una manera innovadora de encuadrar los incidentes informáticos dentro de un esquema de estudio en donde, si bien los incidentes fueron registrados como hechos individuales, se agruparon y analizaron como eventos vinculados en forma condicional a una instancia superior que es la operatividad del centro de cómputos. Para obtener la información necesaria, se realizaron acuerdos con organizaciones públicas y privadas de tal manera de poder conformar bases de datos que proporcionen información básica como: Tipo de incidente, registro de inicio del incidente, registro de la resolución del incidente, forma en que afecta este incidente a la organización, etc. Con el aporte de esta información se alimenta la base de datos que representa la clasificación de incidentes, que tiene como raíz principal el evento *Incidentes Informáticos*, configurándose un grafo bayesiano que refleje los distintos casos que generen incidentes [9] y que responda al siguiente formato:




Inclusive se puede realizar un análisis de los factores que componen el primer nivel de este gráfico, generando un esquema arborescente del tipo:




Para entender cómo se aplica la red bayesiana a este caso en particular, se podría considerar completar este gráfico con parámetros propios de esta temática, como se indica en el siguiente gráfico:




De esta manera podemos llegar a realizar la modelación, en estructura de árbol, de la distribución de incidentes de que creamos conveniente para poder realizar el análisis de este evento.

En este contexto, con el modelo establecido y una base de datos que refleje los incidentes producidos durante un período de tiempo dado, se procede a alimentar este modelo. Como se comenta previamente, ante un comportamiento particular de un tipo de incidentes, se puede redefinir o expandir alguna de las ramas que componen el árbol definido, de tal manera que, por ejemplo la clasificación de incidentes determinada por "Periféricos" se pueda a su vez dividir en "periféricos de almacenamiento de información" y "otros periféricos". De esta manera se va ajustando la orientación del análisis de tal manera que se posibilita conformar un análisis adecuado a los fines del estudio.

## Resultados y Objetivos

Mediante este proyecto se logró plantear una metodología basada en un modelo matemático de red Bayesiana, que posibilita realizar un tratamiento adecuado de incidentes informáticos.

Al tomar como base este modelo, se puede realizar el manejo de los incidentes de manera tal que posibilite el diseño de una estrategia sistematizada que permita el diseño de una herramienta de software que posibilite asistir en la toma de decisiones tendiente a mejorar la operatividad de una organización.

En el transcurso de la toma de datos, se observó que una vez organizado el registro de incidentes, el flujo continuo de estos datos contribuye a la retroalimentación del sistema y esto permite actualizar permanentemente los resultados a medida que nueva información se incorpora, posibilitando determinar la tendencia en la aparición de incidentes.

En este contexto se analizó las bases de datos conformadas con los incidentes registrados de dos organizaciones. En los dos casos se pudo determinar una "tendencia" en el comportamiento que delineo un perfil de incidentes para cada organización. De esta manera pudo determinarse claramente que tipo de incidente tuvo preponderancia y por tanto analizar las medidas preventivas a tomar para lograr una baja ocurrencia y por tanto mejorar el rendimiento operativo de la organización [10].

En este momento, el proyecto que dio origen a este documento se encuentra finalizado. Mediante los resultados obtenidos se pudo probar que la técnica de clasificar los incidentes mediante el diseño de una red bayesiana dio un resultado positivo al momento de intentar identificar y modelar los factores que afectan, en forma sustancial, el rendimiento de un centro de cómputos de manera tanto cualitativa como cuantitativa. De esta forma, y en base al procesamiento de la información organizada de esta manera, se puede actuar en forma eficiente al momento de tomar acciones proactivas en la prevención de los incidentes informáticos.

Es así que, basados en la experiencia capitalizada en el desarrollo de este proyecto, se está analizando el diseño de una herramienta de software que materialice este conocimiento y aproveche los beneficios de las redes bayesianas en el tratamiento de incidentes. Esta herramienta de software se basará en la facilidad que brinda un modelo bayesiano en ser implementado bajo técnicas de inteligencia artificial, desarrollado mediante un sistema experto [11]. Con el desarrollo de esta herramienta, está contemplado gestionar las bases de datos que se han diseñado en base a las redes bayesianas, administrar y presentar estos datos de manera que el usuario cuente con la información necesaria para identificar tendencias en la aparición de incidentes, posibilitando enfocar recursos en áreas donde los incidentes causen los mayores inconvenientes.

## Formación de Recursos Humanos

Dentro de los objetivos del proyecto se encontraba consolidar la conformación de un equipo de trabajo radicado en la UNPA-UARG, que comenzó con el estudio de aspectos de seguridad informática y también afianzar los vínculos con equipos de investigación radicados en la UNLP que también han desarrollado conocimiento en este tema de investigación. En este contexto se convocó además a dos alumnos de la carrera de Licenciatura en Informática de la UNPA-UARG, para que puedan experimentar el trabajo en un equipo de investigación. Surgiendo también la posibilidad de iniciar el trabajo final de carrera una alumna de la carrera de Licenciatura, la que se encuentra actualmente en desarrollo.

## Referencias

[1] Jensen, F. (1996). An Introduction to Bayesian Networks. Springer.

[2] Langley, P., W. Iba, & K. Thompson (1992). An analysis of Bayesian classifiers. In Proceedings, Tenth National Conference on Artificial Intelligence (pp. 223–228). Menlo Park, CA: AAAI Press

[3] Boys, R. J., D. J. Wilkinson and T. B. L. Kirkwood (2008) Bayesian inference for a discretely observed stochastic kinetic model. Statistics and Computing, 18(2), 125-135.

[4] Neil, M., Fenton, N., Forey, S. & Harris, R. (2001), .Using Bayesian belief networks to predict the reliability of military vehicles. Computing & Control Engineering Journal, Feb. 2001, pp 11-20.

[5] David Heckerman (1997). Bayesian Networks for Data Mining, Journal of knowledge Discovery and Data Mining 1(1), pag. 79-119 Kluwer Acedemic Publishers

[6] Ferat Sahin, John S. Bay (2001). Structural Bayesian network learning in a biological decision theoretic intelligent agent and its application to a herding problem in the context of distributed multi agent system, 2001 IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pages: 1606 - 1611

[7] De la Fuente, E. I., García, J., y De la Fuente, L. (2002). Estadística bayesiana en la investigación psicológica. Metodología de las Ciencias del Comportamiento, 4, 185-200.

[8] D. Heckerman, D. Geiger, and D.M. Chickering, (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Machine Learning, vol. 20, pp. 197-243.

[9] Nir Friedman, Dan Geiger, Moises Goldszmidt (1997). Bayesian Network

Classifiers. Kluwer Academic Publishers. Manufactured in The Netherlands Machine Learning, 29, 131–163

[10] Talay, C. A., Rojas, G., Saenz, J. L., Maglione, D. S., Amarilla, C. y Marrone L. (2012). Tratamiento de incidentes informáticos mediante la utilización de redes bayesianas. XVIII CACIC - Bahía Blanca - Arg..

[11] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., y Spiegelhalter, D. J. (1999). Probabilistic networks and expert systems. Harrisonburg, VA: Springer.