Representación matemática de la incertidumbre en modelos genéricos de gestión de la confianza

Julio José Teca Nemesio

Instituto Tecnológico de Informática

jutene@fiv.upv.es

Technical Report ITI-ITE-08/19

Representación matemática de la incertidumbre en modelos genéricos de gestión de la confianza

Julio José Teca Nemesio

Instituto Tecnológico de Informática

Technical Report ITI-ITE-08/19

e-mail: jutene@fiv.upv.es

27 de noviembre de 2008

Índice

Bibliografía

1

Introducción

Muchos de los ataques y vulnerabilidades de los sistemas resultan de la capacidad de los atacantes de sortear los mecanismos de seguridad preventivos. Por ello, los procesos de detección y reacción resultan fundamentales.

Un protocolo de gestión de la confianza es un protocolo mediante el cuál los nodos de un sistema distribuido pueden obtener evidencias que les permitan establecer vínculos de confianza con otros nodos. Dichas relaciones de confianza pueden utilizarse como base para desarrollar mecanismos de detección y reacción. Algunos ejemplos de aplicación son los protocolos seguros de encaminamiento, los protocolos de autenticación de claves o los protocolos para garantizar criterios de calidad de servicio.

En este trabajo presentamos un modelo genérico de gestión de la confianza para sistemas distribuidos asíncronos que tolera la presencia de nodos defectuosos y nodos maliciosos, incluso cuando estos exhiban un comportamiento bizantino. Además, los nodos maliciosos no pueden afectar sin ayuda a la reputación o las relaciones de confianza de otros nodos.

En nuestro modelo, los valores de creencia se verán afectados por la interacción entre agentes, por el paso del tiempo y también por la rectificación de los recomendantes.

Por simplicidad, basamos la autenticación de los mensajes utilizados en el protocolo en criptografía asimétrica. Pensamos que esto no reduce la generalidad de nuestra propuesta, especialmente si consideramos que incluso algunos mecanismos de seguridad para la autenticación y el encaminamiento seguro en redes ad hoc, que presentan cambios impredecibles de topología y nodos con importantes limitaciones de potencia, se basan en este tipo de criptografía [32, 13, 17, 31, 19, 3, 27, 16].

Dado que en nuestro modelo la confianza está orientada a características, depende del contexto y es multifacetada, la utilización de un protocolo de pertenencia a grupos [4, 9, 10, 18, 21, 25, 20, 14] no resulta de utilidad puesto que no puede asumirse que, en cierto instante de tiempo y contexto, todo nodo confíe en los restantes nodos del sistema ni que tenga criterio de asesoramiento. Nuestro protocolo realiza estimaciones sobre los nodos activos del sistema en base a las referencias que obtiene de los mensajes intercambiados en los procesos de asesoramiento y reevaluación. Dichas nociones se utilizan para estimar los umbrales de tolerancia.

En nuestro modelo, la representación de la confianza se basa en la teoría de las funciones de creencia de Shafer-Dempster [29]. Esta teoría proporciona un enfoque no Bayesiano a la utilización de la probabilidad matemática para cuantificar juicios subjetivos. Existen numerosas revisiones de esta teoría, no obstante, las funciones de creencia de Shafer-Dempster siguen constituyendo, junto con los intervalos de confianza de Neyman-Pearson y los tests de representatividad de Fisher, las principales herramientas de razonamiento subjetivo mediante probabilidades.

En el siguiente apartado estableceremos la importancia de la incertidumbre en las relaciones de confianza. En la literatura se han propuesto diversos modelos para representar la confianza y agregar las evidencias, en la sección “Aproximaciones previas al problema de la confianza” describimos brevemente y discutimos algunas de las más representativas. En el apartado “Gestión de la confianza” enfatizamos la diferencia entre fiabilidad y confianza. Posteriormente, en la sección “Enfoques no frecuentistas de la teoría de probabilidades” enunciamos las limitaciones de las aproximaciones Bayesianas al problema de la representación de

la confianza. En esta sección proponemos nuestro modelo de representación, evaluación, reevaluación y revocación de la confianza. Finalmente, destacamos las principales aportaciones de este trabajo y proporcionamos una breve descripción de posibles trabajos futuros de investigación.

2

Concepto de confianza

La cooperación es una necesidad humana básica. Incluso para competir, de un modo no destructivo, es necesario confiar hasta cierto punto en que los rivales cumplirán ciertas reglas.
Esto se aplica igualmente a política y a procesos económicos. La teoría de juegos nos ha proporcionado una mejor comprensión del hecho de que la ausencia de cooperación es posible aun cuando ésta podría beneficiar a las entidades implicadas. Es un error fundamental asumir que porque cierto comportamiento cooperativo pudiese beneficiar a cada individuo de un grupo, las entidades de dicho grupo vayan a adoptar ese comportamiento.
Aun cuando dos entidades tengan motivos adecuados para cooperar, necesitarán conocer las motivaciones de la otra y confiar entre sí. La simple creencia de una de las entidades en que la otra podría perjudicarla puede empujarla a atacarla como mecanismo de auto-defensa. Asimismo, la creencia de la primera entidad en la falta absoluta de cooperación de la segunda o su creencia en que la otra entidad no confiará en ella podría tener un efecto similar.
Por tanto, la incertidumbre está íntimamente relacionada con el concepto de confianza. Las entidades desconocen los intereses ocultos de otras entidades. No obstante, es importante destacar que la ausencia de credibilidad no implica la ausencia de motivos para cooperar y tampoco implica incredulidad.
Existen otras fuentes de incertidumbre como los errores o las limitaciones en los mecanismos de observación y combinación de evidencias, y los errores de comunicación.

3

Aproximaciones previas al problema de la confianza

En este apartado describimos brevemente un conjunto representativo de aproximaciones al problema de la gestión de la confianza y discutimos su interés en base a dos aspectos: la representación de la confianza, enfatizando en la capacidad para modelar la incertidumbre, y el mecanismo de combinación que permite agregar las evidencias.

3.1. El modelo BBK

El modelo BBK [2] es un protocolo orientado a la autenticación de claves que representa las valoraciones de confianza mediante un valor real perteneciente al intervalo [0,1] y que utiliza dos contadores enteros p y n para almacenar el número de experiencias positivas y negativas, respectivamente.
Este protocolo identifica dos tipos de relación: relaciones de confianza directa y relaciones de confianza de recomendación. Las valoraciones de confianza se realizan en base a una aproximación probabilista convencional (probabilidad condicional Bayesiana).
La valoración de las relaciones se realiza en base a la expresión:

$$\begin{aligned} P(r > \alpha | ok = p) &= \frac{P(r > \alpha, ok = p)}{P(ok = p)} &= \frac{\int_\alpha^1 x^p \mathrm{d}x}{\int_0^1 x^p \mathrm{d}x} \\ &= \frac{(p+1)^{-1}(1-\alpha^{p+1})}{(p+1)^{-1}} &= 1 - \alpha^{p+1} \end{aligned}$$

donde la variable aleatoria ok representa el número de experiencias positivas y r es la fiabilidad de la entidad.
Las relaciones de confianza directa se valoran mediante la fórmula:

$$v_z(p) = 1 - \alpha^p$$

Se utiliza p en lugar de $p + 1$ para garantizar que a las entidades desconocidas se les asigna un grado de confianza nulo.
Las relaciones de confianza de recomendación se valoran mediante la expresión:

$$v_r(p, n) = \begin{cases} 1 - \alpha^{p-n} & \text{si } p > n \\ 0 & \text{en otro caso} \end{cases}$$

Los recomendantes se juzgan a partir de las experiencias que la entidad tiene con los recomendados.

En el modelo BBK las relaciones de confianza son transitivas. Dadas las entidades y relaciones de la figura 1:

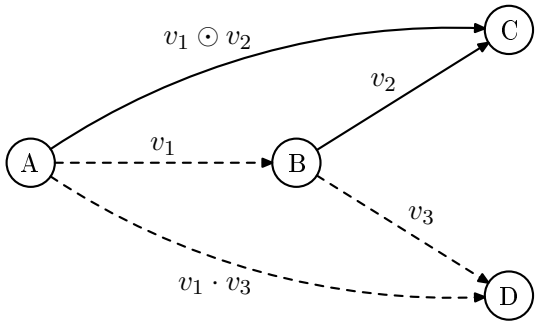

Figura 1: Derivación de confianza

donde el trazo discontinuo representa las relaciones de confianza de recomendación y el trazo continuo denota las relaciones de confianza directa, es posible derivar la relación de confianza directa entre las entidades A y C mediante la expresión:

$$v_1 \odot v_2 = 1 - (1 - v_2)^{v_1}$$

asimismo se puede derivar la relación de confianza de recomendación entre A y D utilizando el siguiente producto:

$$v_1 \cdot v_3$$

Dado que pueden existir diferentes caminos de cierto nodo A a determinado nodo B, es posible que se hayan derivado diferentes estimaciones de su afinidad a través de los distintos caminos. En BBK el mecanismo de combinación de valoraciones es la media aritmética. En el caso de las relaciones de recomendación, dadas n relaciones de confianza de recomendación entre las mismas entidades y con respecto a la misma clase de confianza V_i, la expresión utilizada es:

$$V_{com} = \frac{1}{n} \sum_{i=1}^{n} V_i$$

La expresión usada para combinar las valoraciones asociadas a las relaciones de confianza se basa en el concepto de recomendación. Las experiencias de cierta entidad con la entidad que va a ser recomendada son propagadas por los caminos de recomendación, experimentando reducciones en base a las correspondientes valoraciones de las relaciones de confianza de recomendación. Dado que es posible que las mismas experiencias se propaguen a cierta entidad por diferentes caminos, debe realizarse su media para obtener un valor único. El valor de confianza combinado puede calcularse mediante la siguiente fórmula:

$$V_{com} = 1 - \prod_{i=1}^{m} \alpha^{\frac{1}{n_i} \left(\sum_{j=1}^{n_i} \overline{V} i,j \right) \cdot p_i} = 1 - \alpha^{\sum_{i=1}^{m} \frac{1}{n_i} \left(\sum_{j=1}^{n_i} \overline{V} i,j \right) \cdot p_i}$$

3.1.1. Discusión

En este algoritmo, la representación de las relaciones de confianza directa y de recomendación sigue un enfoque frecuentista Bayesiano. La principal limitación de esta aproximación radica en la incapacidad de representar la incertidumbre, que resulta fundamental para la cuantificación de juicios subjetivos. Por otro lado, las probabilidades que se utilizan en el modelo no tienen sentido a no ser que todos los parámetros de confianza se interpreten como las probabilidades de eventos bien definidos del mismo experimento aleatorio.

El algoritmo BBK proporciona mecanismos de derivación tanto de las relaciones de confianza directa como de las relaciones de confianza de recomendación. No consideramos apropiado asumir que las relaciones de confianza son transitivas. En el mundo real, la relaciones de confianza no son incondicionalmente transitivas. Las entidades deberían tener poder de decisión sobre sus relaciones de confianza con otras entidades. La transitividad ,en el contexto de la confianza, no puede ser automatizada.

El mecanismo de combinación de evidencias utilizado se basa en la media aritmética. La media es un mecanismo de agregación excesivamente sensible a la inclusión de nuevas evidencias, de modo que puede crecer o decrecer conforme aparecen nuevas relaciones de confianza. Si el conjunto de relaciones de confianza entre dos entidades con respecto a la misma clase de creencia aumenta, el comportamiento de la combinación de valores no puede predecirse, es decir, que la media es no monótona. De este modo, es imprescindible conocer si todas las relaciones de confianza han sido derivadas ya que de lo contrario deberían considerarse las densidades de probabilidad implicadas.

3.2. El modelo Schillo

En el modelo Schillo [28], las entidades intercambian un vector n-dimensional de valores binarios que representan los resultados de un conjunto de interacciones. Cada valor se corresponde con un evento, que puede ser el resultado de una interacción directa entre el nodo que proporciona la información y cierta entidad de interés o puede ser el resultado de cierto evento observado por la entidad informadora. Dado el carácter binario de la representación, las experiencias u observaciones solamente pueden ser calificadas como positivas (1) o negativas (0).

En este modelo, el cálculo de la confianza se basa en una aproximación probabilista y el mecanismo utilizado para combinar los valores de confianza consiste en obtener una secuencia de eventos sobre los que se aplica un procedimiento probabilístico para predecir el comportamiento futuro de la entidad de interés.

Este mecanismo de combinación de evidencias se basa en dos premisas: Primera, las entidades que actúan como testigos no pueden mentir pero si ocultar información positiva con el objeto de difamar a otras entidades. Segunda, la información intercambiada por las entidades debe ser un conjunto de observaciones y no un resumen de las mismas.

Bajo estas asunciones se considera que una entidad, para evitar ponerse en evidencia, no tiene motivaciones para mentir dado que otros agentes podrían haber observado el mismo hecho y determinar que no ha dicho la verdad.

El proceso de combinación consiste, por tanto, en la estimación de la cantidad de información que ha sido ocultada por los agentes, utilizando distribuciones binomiales, y en el cálculo de la probabilidad de que el testigo mienta. Esto último se hace tomando como referencia experiencias directas en las que haya intervenido el testigo. A continuación, a partir de la estimación de la cantidad de información que ha ocultado cada testigo se reconstruyen los vectores proporcionados por los testigos bajo la asunción de que la distribución probabilística que describe la información ocultada coincide con la distribución de probabilidades asociada a la información conocida. Finalmente, los vectores son combinados y resultado se utiliza para calcular la probabilidad de que la interacción con la entidad de interés será positivo:

$$P(i = p) = \frac{\text{Número de experiencias positivas}}{\text{Número de eventos}}$$

3.2.1. Discusión

El principal defecto de esta aproximación es que no especifica cómo encontrar buenos testigos. No se proporciona un proceso de remisión a través del cuál los agentes puedan ayudarse mutuamente a encontrar testigos. Bajo este marco, sería posible que varios testigos se aliasen para mentir utilizándose entre sí como coartada. La absolución por exculpación mutua es una de las formas más elementales de cooperación. De igual modo, dichos nodos también podrían generar información espúrea.

El mecanismo de combinación de evidencias propuesto es únicamente válido asumiendo que todos los agentes tendrán la misma opinión sobre cierto evento específico y por tanto no será necesario resolver el problema de la correlación entre evidencias.

Otro de los inconvenientes que presenta es que la información intercambiada entre los agentes consiste en evaluaciones sobre eventos individuales. Esto tiene dos consecuencias: Primera, la información puede no evaluar el comportamiento general de la entidad. Segunda, se requiere intercambiar un mayor número de mensajes que al utilizar opiniones de carácter global. Este tipo de estrategia podría utilizarse únicamente en casos muy concretos.

3.3. El modelo Abdul-Rahman-Hailes

El modelo Abdul-Rahman-Hailes [1] utiliza etiquetas lingüísticas para representar diferentes grados de confianza. A diferencia de otros modelos [7, 23] las etiquetas lingüísticas no están relacionadas con la teoría de conjuntos difusos.

En este modelo, los agentes almacenan el grado de confianza que tienen en cierta entidad. Dicho grado de confianza es un miembro del conjunto ordenado $E = \{vg, g, b, vb\}$, donde vg se corresponde con la etiqueta lingüística "very good", g representa "good", b equivale a "bad" y vb a "very bad". Este grado de confianza expresa la opinión global del agente sobre la entidad de interés.

El mecanismo de combinación se realiza en cuatro pasos. En el primero, se descarta la información que proviene de agentes desconocidos, es decir, aquellos agentes para los que no existe un precedente que permita contrastar la información proporcionada y, a partir de dicha comparación, establecer una distancia semántica utilizable en recomendaciones futuras. Una distancia semántica es una medida que permite a un agente adaptar, de acuerdo a su propia percepción, la información suministrada por otros agentes. Esta medida calcula la desviación entre la información recibida de otros agentes y la experiencia propia.

En el segundo paso, a cada agente conocido se le asigna un peso de acuerdo a su confiabilidad. El valor de confiabilidad se calcula en base a las distancias entre las recomendaciones de los agentes y las experiencias reales derivadas de confiar en esas recomendaciones. La figura 2 establece la relación entre confiabilidad y peso.

En tercer lugar, a partir de la distancia semántica, se ajusta la información proporcionada por cada agente. Finalmente, para cada grado de confianza recomendado $t \in \{vg, g, v, cb\}$, se suman los pesos de los agentes que dieron soporte a ese grado. El vector de valores obtenido representa el nivel de apoyo a cada grado. El grado de confianza combinado será aquél que tenga el mayor soporte. En caso de que existiesen varios grados con el valor máximo, la agregación final se hace en base a la siguiente regla:

- Si los valores máximos se corresponden con los grados de "good" y "very good", dándose menor apoyo a los grados de "bad" y "very bad", entonces se devolverá la etiqueta lingüística mg, "mostly good".
- Si los valores máximos se corresponden con los grados de "bad" y "very bad", dándose menor apoyo a los grados de "good" y "very good", entonces se devolverá la etiqueta semántica mb, "mostly bad".
- En otro caso, se devolverá la etiqueta semántica gb, "equal amount of good and bad".

Grado de confiabilidad	Peso
0	9
1	5
2	3
3	1
Desconocido	0

Figura 2: Asignación de pesos en función del grado de confiabilidad de los agentes

3.3.1. Discusión

El principal problema que presenta la aproximación de Abdul-Rahman y Hailes se encuentra en la inicialización del sistema (*bootstrapping*). En el mundo real, el ciclo natural de la confianza se inicia con la

selección de compañeros, a los que se asignará cierto grado inicial de confianza. Una vez el sistema está en funcionamiento, comienza a observar y recoger evidencias del comportamiento de los nodos. Entonces, se actualizan los niveles de confianza estimados.
El algoritmo propuesto por Abdul-Rahman y Hailes no respeta el ciclo de confianza, dado que asume la preexistencia de evidencias. Inicialmente no pueden existir agentes conocidos, puesto que todavía no se ha realizado interacción alguna. Para alcanzar el estado propuesto en este algoritmo, se requeriría de la existencia previa de un sistema de recomendación.
A diferencia del modelo Schillo, en este modelo las opiniones son de carácter global. No obstante, no se toman en consideración diversos contextos. Además, no es posible representar el grado de confianza que los agentes tienen en los valores que intercambian.
Únicamente se considera el concepto de incertidumbre en caso de empate entre las valoraciones combinadas y, en cualquier caso, la representación de la incertidumbre en base a etiquetas lingüísticas es muy limitada.

3.4. El modelo ReGreT

En el modelo ReGreT [26] la información de confianza se almacena en un vector bidimensional de valores reales $\langle \text{Trust}_{w\to t}(\varphi), \text{TrustRL}_{w\to t}(\varphi)\rangle$. La primera componente $\text{Trust}_{w\to t}(\varphi) \in [-1,1]$ es el valor de confianza en cierta entidad t en relación a determinado aspecto de su comportamiento φ desde el punto de vista del agente w. La segunda componente $\text{TrustRL}_{w\to t}(\varphi) \in [0,1]$ es el grado de fiabilidad que el agente que actúa de testigo le confiere al valor de confianza. En este caso, la información considera globalmente la entidad de interés, pero con respecto a un aspecto concreto del comportamiento de la misma.
ReGreT utiliza un mecanismo de combinación de evidencias bastante complejo que permite establecer la credibilidad de los agentes en base a relaciones sociales. El valor combinado de credibilidad, denotado como R, se calcula mediante la suma ponderada de los valores de confianza (Trust) recibidos. El peso asociado a cada valor de confianza es la credibilidad normalizada del agente que la envió.
Para establecer el grado de fiabilidad de la opinión de cierto agente se utiliza el valor mínimo entre la credibilidad del agente y la fiabilidad que el propio agente proporciona. Cuando los agentes son confiables puede usarse el valor de fiabilidad que proporcionan. En caso contrario, se usará la credibilidad asociada al agente ($\mathrm{witnessCr}$) como medida de la fiabilidad de la información:

$$\begin{aligned} \mathrm{R}_{\alpha\to t}(\varphi) &= \sum_{w_i\in W} \omega^{w_i\cdot t}\cdot \mathrm{Trust}_{w_i\to t}(\varphi) \\ \mathrm{RL}_{\alpha\to t}(\varphi) &= \sum_{w_i\in W} \omega^{w_i\cdot t}\cdot \min(\mathrm{witnessCr}(a,w_i,t), \mathrm{TrustRL}_{w_i\to t}(\varphi)) \end{aligned}$$

donde a es el evaluador que está realizando la combinación y:

$$\omega^{w_i\cdot t} = \frac{\mathrm{witnessCr}(a,w_i,t)}{\sum_{w_j\in W}\mathrm{witnessCr}(a,w_j,t)}$$

3.4.1. Discusión

Aunque ReGreT permite representar el grado de fiabilidad asociado a los valores transmitidos por los agentes, utiliza un simple valor real para representar la evaluación de la confianza. Este hecho limita la expresividad del modelo y lo hace incapaz de representar el concepto de incertidumbre.

El grado de credibilidad que los agentes asocian a la información que suministran permite al algoritmo distinguir si un resultado neutral de la combinación de evidencias proviene de la agregación de valores de confianza totalmente contradictorios o si por el contrario se trata de un verdadero estado neutral, ya que el resultado de agregar información contradictoria tendrá un bajo grado de fiabilidad asociado. Sin embargo, es imposible distinguir entre el resultado de combinar valores contrarios extremos de confianza de una valoración que tenga un bajo grado de fiabilidad debido a la falta de evidencias suficientes.

3.5. El modelo Afras

El modelo Afras [7] está basado en la teoría de los conjuntos difusos. Los conjuntos difusos son una extensión del concepto clásico de conjunto, donde la pertenencia de los elementos al conjunto se expresa en términos binarios de acuerdo a una condición bivalente (un elemento pertenece o no pertenece al conjunto). Por el contrario, en la teoría de los conjuntos difusos es posible expresar de un modo gradual la pertenencia de un elemento a un conjunto. Esto se hace en base a una función de pertenencia que devuelve un valor en el intervalo [0,1].
En el modelo Afras, la reputación que los testigos asignan a cierta entidad de interés se almacena en un conjunto difuso en el espacio [0,100]. El grado de fiabilidad que el testigo confiere a dicha reputación está implícita en la forma que adopta el propio conjunto difuso. De esta manera, un conjunto difuso amplio significa menor confianza y viceversa.
En esta aproximación, las nuevas evidencias se agregan al valor de reputación actual mediante una media aritmética ponderada. Si $R_{t-1}^{E\to A}$ representa la reputación de cierta entidad A en el instante de tiempo $t-1$ desde el punto de vista del evaluador E, y $R_t^{T\to A}$ representa la reputación que el testigo T ha comunicado sobre A, entonces la nueva evidencia puede ser agregada mediante la expresión:

$$R_t^{E\to A} = w \cdot R_t^{T\to A} + (1-w) \cdot R_{t-1}^{E\to A}$$

donde el peso $w = \mathrm{cg}(R_t^{E\to T}) \in \left\lfloor 0, \frac{1}{2} \right\rfloor$, siendo cg el centro de gravedad del conjunto difuso y $R_t^{E\to T}$ la reputación del testigo T desde el punto de vista del evaluador E en el instante de tiempo t.

3.5.1. Discusión

Aunque la utilización de conjuntos difusos proporciona gran potencia expresiva a esta aproximación, también impide la diferenciación explícita entre incertidumbre e imprevisibilidad al combinar la evaluación de la reputación con el grado de incertidumbre asociada a la misma en la propia representación difusa.
El mecanismo de combinación de evidencias únicamente requiere de la información más reciente y la información combinada fruto de agregaciones anteriores. Aunque esta representación es compacta no permite la reevaluación de los valores de reputación previos. Esto es un grave inconveniente, dado que la confianza es un proceso dinámico.

3.6. El modelo Ramchurn

El modelo Ramchurn [23], al igual que Afras, está basado en la teoría de conjuntos difusos y también utiliza un mecanismo de combinación de evidencias basado en una media aritmética ponderada. Sin embargo, la forma en la que se utilizan los conjuntos difusos para representar la información que intercambian los agentes es totalmente diferente.

En este modelo se utiliza un esquema contractual para representar relaciones de compromiso entre los agentes. El cumplimiento de un contrato se expresa en términos de variaciones absolutas de utilidad, denotadas ΔU, entre el contrato firmado y el contrato implementado. Los agentes comparten un conjunto de etiquetas lingüísticas, donde cada etiqueta $L \in \{Bad, Average, Good\}$ se modela como un conjunto difuso en el dominio de las desviaciones de utilidad $\Delta U = [-1, 1]$ que se especifica con la función de pertenencia $\mu_L(u) : [-1, 1] \rightarrow [0, 1]$.
La información que intercambian los agentes es un conjunto de tres niveles de confianza, uno por cada conjunto difuso. Un nivel de confianza se define como el nivel de pertenencia del comportamiento de un agente a, con respecto a cierta cuestión x, a un término lingüístico L y se denota $C(e, a, x, L)$, donde e es un evaluador.
El fragmento del conjunto difuso definido por $C(e, a, x, L)$ representa un rango de valores en el eje de abscisas que indica el intervalo de desviaciones de utilidad esperado en tiempo de ejecución en relación con la cuestión x por el agente a.
La información intercambiada por los agentes consiste en la opinión que estos tienen sobre cierta entidad de interés, expresada en términos de variaciones de utilidad.
El mecanismo de combinación de evidencias asume que los agentes pertenecen a grupos sociales y otorga mayor credibilidad a aquellas informaciones que provienen de agentes pertenecientes a un grupo social relevante. Esto se tiene en cuenta en la expresión que se utiliza para agregar los niveles de confianza recibidos para cada etiqueta lingüística (conjunto difuso):

$$\text{Rep}(e, a, x, L) = \sum_{G_i \in G} w_i \cdot min_{e \in G_i} \text{C}(e, a, x, L)$$

donde $w_i > w_j$ si la importancia social de G_i es mayor que la importancia social de G_j.

3.6.1. Discusión

La utilización de conjuntos difusos aporta expresividad al modelo de Ramchurn. No obstante, el rango de utilidad esperada refleja a su vez el grado de incertidumbre por lo que no es posible distinguir entre una evaluación que presente una gran incertidumbre de otra que represente una elevada variabilidad en el grado esperado de utilidad. Por tanto, no es posible interpretar si el agente que proporciona la información es incapaz de interpretar la entidad de interés o si dicha entidad se comporta de forma caótica.

3.7. Conclusión

Los actuales modelos de confianza han sido criticados por no hacer una clara relación entre confianza y reputación, así como por considerar estos conceptos de un modo independiente del tiempo y el contexto [22]. Grandison y Sloman concluyen que algunos sistemas [6, 5, 12, 30] se limitan a sí mismos a una subsección del problema de la gestión de la confianza, sin considerar la actualización de la confianza en base a la información disponible [11].
Entre los sistemas analizados no existe ninguno de carácter genérico y únicamente los modelos BBK, ReGreT y Ramchurn consideran, en cierta medida, que la confianza es dependiente del contexto.
Entre los modelos estudiados, las aproximaciones que presentan mayor expresividad son aquéllas basadas en los conjuntos difusos pero no son capaces de representar el concepto de incertidumbre, distinguiéndola claramente de comportamientos imprevisibles o muy variables.

Finalmente, en algunos de los modelos la representación de la confianza y la reputación no permite la reevaluación, dado que el mecanismo de combinación destruye las evidencias durante el proceso de agregación.

4

Gestión de la confianza

El problema de la confianza en sistemas multi-agente está relativamente bien definido. Existe diversidad de protocolos de representación y evaluación de la confianza. No obstante, en muchas aproximaciones a la gestión de la confianza se asimilan de forma implícita los conceptos de confianza y fiabilidad. En estos modelos, se toman como entrada medidas subjetivas de carácter general, relacionadas con la fiabilidad, que una vez tratadas estadísticamente se emplean como índices de la confianza.

En [15] se considera la posibilidad de que las entidades puedan manifestar comportamientos pasionales, estableciéndose dos categorías de entidades: entidades racionales (*rational entities*) y entidades emocionales (*passionate entities*).

Las entidades emocionales son aquellas que tienen "voluntad propia" y que disponen de mecanismos que les proporcionan un comportamiento cuasi-humano. En la actualidad, los algoritmos, los protocolos, el software o el hardware difícilmente pueden caracterizarse como emocionales. Únicamente las personas, organizaciones humanas o una combinación de las mismas pueden presentar rasgos emocionales. Nunca los sistemas por sí solos.

Las entidades racionales son aquellas que no tienen libre albedrío y no pueden actuar de manera benevolente o maliciosa de modo intencionado. Por tanto, una entidad racional no puede aprovechar la confianza de otra entidad para atacarla.

De acuerdo con este planteamiento los ataques siempre tienen origen en entidades emocionales y la confianza no se deposita en el comportamiento de las entidades racionales sino en su resistencia a los intentos de manipulación maliciosa por parte de entidades emocionales.

Dado que la confianza se basa en el conocimiento, si sólo consideramos entidades de tipo racional, utilizar una aproximación frecuentista de la teoría de probabilidades tiene sentido. Las entidades racionales no pueden actuar de manera ilícita pero pueden fallar. En este sentido, fiabilidad y confianza coinciden.

Un enfoque frecuentista de la teoría de probabilidades debería emplearse únicamente si pueden ignorarse las relaciones entre entidades racionales y emocionales, es decir, entre entidades de confianza y atacantes.

5

Enfoques no frecuentistas de la teoría de probabilidades

Dada la existencia de entidades emocionales, que pueden exhibir comportamientos bizantinos, la evaluación de la confianza no puede reducirse al análisis estadístico de las interacciones entre entidades durante cierto periodo de tiempo, es decir, a la calidad de servicio en un sentido amplio.
Un algoritmo de gestión de la confianza sin supervisión podría llegar a obtener resultados inconsistentes, ya que evidencias obtenidas de diferentes fuentes podrían ser genuinamente contradictorias. Por tanto, para afrontar la incertidumbre epistemológica y aleatoria, y evaluar la plausibilidad de los resultados obtenidos es necesario modelar la incertidumbre de un modo consistente y utilizar razonamiento incierto.
En la actualidad, la mayor parte de trabajos de investigación que se enfrentan al concepto de incertidumbre, desde el punto de vista de la fiabilidad, utilizan aproximaciones frecuentistas de la teoría de probabilidades o estrategias basadas en la lógica difusa.
La principal desventaja del enfoque Bayesiano es que no permite distinguir entre la ausencia de creencia y la incredulidad. Los axiomas de la teoría de probabilidades se basan en la propiedad de la completitud, que afirma que "cierta probabilidad puede asignarse a toda proposición bien definida". Dado un espacio de muestras arbitrario Θ, tal que $A \in \Theta$, de acuerdo a la segunda regla de Bayes:

$$\mathrm{P}(A \cup A^c) = 1 \tag{1}$$

aplicando la ley de aditividad de Bayes sobre la ecuación 1:

$$\mathrm{P}(A \cup A^c) = \mathrm{P}(A) + \mathrm{P}(A^c) = 1 \tag{2}$$

Esto implica que $\mathrm{P}(A)$ no puede incrementarse a no ser que $\mathrm{P}(A^c)$ se decremente.
En la teoría Bayesiana, la asimilación de nuevas evidencias se realiza en base a la regla de condicionamiento de Bayes. Esta regla no resulta adecuada por dos razones: primero, tiene un efecto asimétrico al obligar a condicionar el conocimiento previo en base a la proposición que representa la nueva evidencia y segundo, se asume que como resultado de considerar esta nueva evidencia deberá establecerse una única proposición como cierta.
A su vez, como se argumenta en [29], la teoría Bayesiana no permite representar el concepto de incertidumbre o ignorancia de un modo sencillo, de forma que pueden originarse inconsistencias.
Por otro lado, la teoría de cuantificación monótona, que incluye la lógica difusa, al no imponer el principio de razón mínima, permite emitir afirmaciones acerca de la verosimilitud simultánea de varios eventos pero sin tener que realizar asunciones sobre los eventos para los que no disponemos de suficiente información. Por tanto, esta teoría es más apropiada para tratar con la incertidumbre. No obstante, en lógica difusa, no se exige la definición clara de las proposiciones, de modo que los conceptos y sus relaciones son vagos. Esta interpretación no es relevante en el problema de la gestión de la confianza y tampoco proporciona un álgebra de combinación de evidencias apropiada.
La teoría de la evidencia de Shafer-Dempster [29], también perteneciente a la teoría de cuantificación no aditiva, ofrece un formalismo que permite respaldar o refutar de un modo explícito, asignando medidas de credibilidad a conjuntos de proposiciones, es decir, expresando de forma manifiesta grados de ignorancia.
La regla de Dempster, que constituye el núcleo de la teoría de funciones de creencia de Shafer, proporciona un álgebra que permite combinar evidencias de forma que evidencias concordantes se refuerzan mutuamente y evidencias contradictorias se debilitan entre sí.
Las funciones de creencia definidas por Shafer son apropiadas en aquellas situaciones en las que la utilización directa de un enfoque frecuentista de la teoría de probabilidades no es posible.

5.1. Definiciones básicas

Definición 1 (Marco de discernimiento) Un marco de discernimiento Θ_X es un conjunto exhaustivo de proposiciones o hipótesis mutuamente excluyentes en relación a cierta cuestión de interés X. Asumiremos que el marco de discernimiento es finito.

Definición 2 (Función de masa) Una función de masa $[\Psi]_m$ en X asigna a cada subconjunto A de Θ_X un valor en $[0,1]$, es decir $[\Psi]_m : 2^{\Theta_X} \rightarrow [0,1]$. Debe satisfacerse:

$$\sum_{A \subseteq \Theta_X} [\Psi(A)]_m = 1 \tag{3}$$

De forma intuitiva, $[\Psi(A)]_m$ es la porción de creencia asignada a la evidencia A que no ha sido asignada a ningún subconjunto propio de A. En ocasiones, se impone una segunda condición $[\Psi(\emptyset)]_m = 0$. Si la función de masa cumple esta segunda condición se denomina normalizada y se representa $[\Psi]_M$.

Los conjuntos $A \subseteq \Theta_X$ tales que $[\Psi(A)]_m \neq 0$ se denominan conjuntos focales.

Definición 3 (Función de creencia) Una función de creencia $[\Psi]_b$ en X asigna a cada subconjunto A de Θ_X un valor en $[0,1]$, es decir $[\Psi]_b : 2^{\Theta_X} \rightarrow [0,1]$. Puede obtenerse $[\Psi]_b$ en base a una función de masa:

$$[\Psi(A)]_b = \sum_{B \subseteq A} [\Psi(B)]_m \tag{4}$$

Si $[\Psi(\emptyset)]_b = 0$ entonces se dice que está normalizada y se representa $[\Psi]_B$.

En la teoría de la evidencia de Shafer-Dempster cierto grado de creencia puede ser asignado a una proposición, pero no necesariamente deberá ser asignado a dicha proposición o a su negación. Esto implica que:

$$[\Psi(A)]_B + [\Psi(A^c)]_B \leq 1 \tag{5}$$

de modo que:

$$[\Psi(A)]_B + [\Psi(A^c)]_B + \mu = 1 \tag{6}$$

donde μ representa el grado de incertidumbre.

5.2. Normalización de funciones

La normalización de las funciones de masa y creencia se realiza mediante la siguientes transformaciones:

$$[\Psi(A)]_M = \begin{cases} 0 & \text{si } A = \emptyset \\ \dfrac{[\Psi(A)]_m}{1 - [\Psi(\emptyset)]_m} & \text{en otro caso} \end{cases} \tag{7}$$

$$[\Psi(A)]_B = \frac{[\Psi(A)]_b - [\Psi(\emptyset)]_b}{1 - [\Psi(\emptyset)]_b} \tag{8}$$

Se asume que las evidencias no son totalmente contradictorias, es decir, que $[\Psi(\emptyset)]_m \neq 1$ y $[\Psi(\emptyset)]_b \neq 1$.

5.3. Combinación de evidencias

Dadas dos funciones de creencia $[\Psi_1]_B$ y $[\Psi_2]_B$ definidas sobre el marco de discernimiento Θ_X, con las asignaciones básicas de probabilidad $[\Psi_1]_M$ y $[\Psi_2]_M$, respectivamente. Si $A_1, A_2, \ldots, A_k$ y $B_1, B_2, \ldots, B_l$ son los elementos focales correspondientes a $[\Psi_1]_M$ y $[\Psi_2]_M$.
La función de masa $[\Psi]_M : 2^{\Theta_X} \rightarrow [0, 1]$:

$$[\Psi(A)]_M = \begin{cases} 0 & \text{si } A = \emptyset \\ \dfrac{\sum\limits_{A_i \cap B_j = A} [\Psi_1(A_i)]_M [\Psi_2(B_j)]_M}{1 - \Phi} & \text{en otro caso} \end{cases} \tag{9}$$

es una asignación básica de probabilidad para todo $A \subset \Theta_X$, donde Φ representa la contradicción entre las evidencias:

$$\Phi = \sum_{A_i \cap B_j = \emptyset} [\Psi_1(A_i)]_M [\Psi_2(B_j)]_M , \Phi < 1 \tag{10}$$

La función de creencia dada por $[\Psi]_M$ se denomina suma ortogonal de $[\Psi_1]_M$ y $[\Psi_2]_M$ y se denota $[\Psi_1]_M \oplus [\Psi_2]_M$.
Por tanto, dadas dos funciones de creencia definidas sobre la misma base de discernimiento pero basadas sobre diferentes cuerpos de evidencia, la regla de combinación de Dempster permite calcular una nueva función de creencia que considere ambas evidencias.
Es posible combinar un número arbitrario de funciones de creencia aplicando recursivamente la regla de combinación Dempster:

$$(([\Psi_1]_B \oplus [\Psi_2]_B) \oplus [\Psi_3]_B) \ldots) \oplus [\Psi_n]_B \tag{11}$$

En [29] se demuestra que el resultado de aplicar la expresión 11 es independiente del orden en que se combinen las funciones de creencia.
Un mecanismo de combinación de evidencias debería poseer las siguientes propiedades:

- Sensibilidad: la inclusión de nuevas evidencias no debería introducir grandes cambios o cambios bruscos.
- Monotonía: el resultado de la combinación debería ser predecible.
- Conmutatividad: el orden en que se combinen las evidencias no debería influir en el resultado de la combinación.

La regla de combinación de Dempster satisface todas las propiedades.

5.4. Ontología genérica

El espacio de proposiciones podría contener un número arbitrario de hipótesis, no obstante, con el fin de plantear un marco de discernimiento de propósito general, utilizaremos un conjunto mínimo de hipótesis genéricas. Consideraremos el espacio de proposiciones formado únicamente por dos hipótesis: Y y N. La primera representa el soporte a cierta cuestión de interés, la segunda la oposición a dicha cuestión.
Dado que $\Theta_X = \{Y, N\}$, si asumimos que no existe contradicción en nuestra ontología ($[\Psi(\emptyset)]_m = 0$), la función de creencia aplicada al marco de discernimiento presentará la siguiente forma:

$$[\Psi(\Theta_X)]_B = \sum_{A \subset \Theta_X} [\Psi(A)]_m = [\Psi(Y)]_m + [\Psi(N)]_m + [\Psi(\Theta_X)]_m = 1 \tag{12}$$

Donde los valores iniciales de las funciones de creencia correspondientes a las hipótesis, que coinciden con sus funciones de masa, podrán establecerse mediante cualquier proceso de estimación: inferencia, deducción, computabilidad, teorías probabilistas, restricciones o una combinación de las mismas. Para evitar la pérdida de generalidad no nos decantaremos por ningún formalismo concreto.
La ignorancia absoluta puede representarse mediante la denominada *función de creencia vacua*:

$$[\Psi(A)]_B = \begin{cases} 1 & \text{si } A = \Theta_X \\ 0 & \text{en otro caso} \end{cases} \tag{13}$$

La certeza absoluta en que la proposición P es cierta se expresa mediante la siguiente función de creencia:

$$[\Psi(x)]_B = \begin{cases} 1 & \text{si } x = \{P\} \\ 0 & \text{en otro caso} \end{cases} \tag{14}$$

5.5. Relaciones de confianza y de asesoramiento

En el contexto de la gestión de la confianza distinguimos dos tipos de relaciones entre las entidades: relaciones de confianza y relaciones de asesoramiento.

5.5.1. Relación de confianza

La confianza se define como la firme creencia en la competencia de una entidad para actuar de forma fiable y segura en un contexto específico. A la inversa, la desconfianza se define como la firme creencia en la incompetencia de una entidad para actuar de forma fiable y segura en un contexto específico.
La ausencia de confianza no equivale a desconfianza, debe considerarse un estado neutral, caracterizado por la falta de conocimiento en el que una entidad no posee suficientes evidencias para emitir un criterio sobre otra.
En nuestra arquitectura, la confianza tiene las siguientes características:

1. Una relación de confianza se establece entre, exactamente, dos entidades.
2. Una relación de confianza es reflexiva.
3. Una relación de confianza no es simétrica. La confianza es unidireccional. Si existe confianza mutua entre dos entidades, lo representamos mediante dos relaciones de confianza independientes. De este modo ambas relaciones pueden gestionarse de forma autónoma.
4. Una relación de confianza es condicionalmente transitiva. Muchos protocolos asumen que la confianza es transitiva. Pero que A confíe en B y que B confíe en C no implica que A confíe en C. La decisión de confiar en una entidad o desconfiar de ella, o de hacerlo en una determinada cadena de confianza es decisión de la entidad. Ningún protocolo puede automatizar las relaciones de confianza.
5. Una relación de confianza está contextualizada. Depende del escenario, del contexto.
6. La confianza es dinámica. Le afectan dos factores: el tiempo y la interacción entre entidades.

7. La confianza es multifacetada. Está basada en múltiples características, en facetas que se evalúan por separado y que, posteriormente se combinan, mediante un procedimiento algebraico para obtener uno o más índices de confianza.
8. La confianza es explícita. Una entidad tiene asociados gran cantidad de aspectos. La confianza nunca se refiere a una entidad como si de una unidad se tratase, sino que está orientada a funciones.
9. La confianza es relativa. Cada entidad del sistema tiene criterio propio.

Una relación de confianza entre dos entidades A y B, en un contexto C y en el instante de tiempo t se representa:

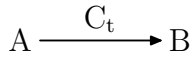

Figura 3: Relación de confianza

o también como $A \, \mathrm{Rc}_{\mathrm{t}}^{\mathrm{c}} \, B$.

5.5.2. Relación de asesoramiento

Las relaciones de recomendación son de fundamental importancia en nuestra sociedad dado que es imposible tener un conocimiento directo de todas las entidades en las que se debe confiar.
Por otro lado, una relación de confianza se establece entre dos entidades y no es incondicionalmente transitiva. Un modelo basado únicamente en relaciones de confianza no evolucionaría con la suficiente fluidez ya que el único dinamismo provendría del paso del tiempo y de la interacción con las entidades en las que la entidad confía de forma explícita.
Con el objeto de facilitar el proceso de inicialización del sistema (*bootstrapping*) y aportar mayor dinamismo al modelo, consideraremos un segundo tipo de relación que denominamos relación de asesoramiento.
Las relaciones de asesoramiento son un tipo específico de relación de confianza en la que una entidad consulta a otra sobre sus relaciones de confianza o asesoramiento.
Una relación de asesoramiento entre dos entidades A y B, en un contexto C, en el instante de tiempo t se representa:

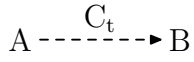

Figura 4: Relación de asesoramiento

o también como $A \, \mathrm{Ra}_{\mathrm{t}}^{\mathrm{c}} \, B$.
Al igual que las relaciones de confianza explícita, las relaciones de asesoramiento son no simétricas. Ambas, relaciones de confianza y relaciones de asesoramiento, son independientes. Consecuentemente, una entidad puede confiar explícitamente en otra pero no confiar en sus recomendaciones y viceversa.
En base al asesoramiento recibido, una entidad podrá decidir si desea establecer nuevas relaciones de confianza o si prefiere considerar las recomendaciones como tales.
Las recomendaciones se utilizan para obtener criterios de evaluación sobre otras entidades mediante las evidencias obtenidas a través de las relaciones de asesoramiento. No debe confundirse este tipo de relación con una relación de confianza incondicionalmente transitiva. En nuestro modelo, una entidad confiará o desconfiará de otra, o en cierta cadena de confianza, por iniciativa propia.
En las recomendaciones, el asesor proporciona a la entidad asesorada los valores de creencia asociados a las relaciones de confianza explícita y asesoramiento de otras entidades. Dichos valores de creencia podrán ser reevaluados por la entidad en función del grado de credibilidad que la entidad le conceda a su asesor.

En nuestro modelo cada entidad almacena localmente la información sobre sus relaciones de confianza y asesoramiento. No existe una estructura de datos global que concentre la información sobre estas relaciones.

Nuestra propuesta de gestión descentralizada de la confianza, en la que las relaciones de confianza son extremadamente volátiles, resulta especialmente apropiada en infraestructuras de clave pública, dada la ineficiencia asociada a la revocación de certificados.

La confianza es contextual, multifacetada y explícita. Por tanto, los valores de creencia deben expresarse sobre aspectos específicos de las entidades. Bajo un modelo de control de acceso basado en roles, dichos roles podrían utilizarse como características o aspectos susceptibles de evaluación. De esta forma una entidad podría confiar en otra como servidor de nombres pero no como autoridad de atributos. Dicha entidad confiaría en la faceta (rol) de servidor de nombres de la otra, pero no en su faceta de autoridad de atributos.

Dado que la ontología que estamos considerando está formada únicamente por dos proposiciones, expresaremos las relaciones de confianza y asesoramiento mediante vectores tridimensionales:

$$[\Psi]_M = ([\Psi(\{Y\})]_M, [\Psi(\{N\})]_M, [\Psi(\{Y,N\})]_M) \tag{15}$$

donde la primera componente representa el grado de creencia, la segunda el grado de incredulidad y la tercera el grado de incertidumbre.

El contexto C en base al cuál se evaluará y reevaluará la confianza vendrá dado por el nivel de aplicación. De este modo diferentes subcomponentes, o hilos de ejecución, de una entidad podrían tener diferentes grados de creencia para las mismas facetas de otras entidades.

La figura 5 muestra la base de datos de confianza de una entidad para una aplicación concreta (contexto C) en un instante de tiempo t:

Relaciones de confianza	Identificador	Faceta	Creencia	Id. recomendante	Timestamp
	$\mathrm{ID_A}$	$\mathrm{Rol_A}$	$[\Psi_\alpha]_M$	ID_α	t_α
		…	…	…	…
		$\mathrm{Rol_C}$	$[\Psi_\beta]_M$	ID_β	t_β
	…	…	…	…	…
	$\mathrm{ID_B}$	$\mathrm{Rol_D}$	$[\Psi_\omega]_M$	ID_ω	t_ω
		…	…	…	…
		$\mathrm{Rol_F}$	$[\Psi_\gamma]_M$	ID_γ	t_γ
Relaciones de asesoramiento	Identificador		Creencia	Id. recomendante	Timestamp
	$\mathrm{ID_C}$		$[\Psi_\theta]_M$	ID_θ	t_θ
	$\mathrm{ID_D}$		$[\Psi_\epsilon]_M$	ID_ϵ	t_ϵ
	…		…	…	…
	$\mathrm{ID_E}$		$[\Psi_\kappa]_M$	ID_κ	t_κ
Recomendaciones	Identificador	Faceta	Creencia	Id. recomendante	Timestamp
	$\mathrm{ID_F}$	$\mathrm{Rol_G}$	$[\Psi_\delta]_M$	ID_δ	t_δ
		…	…	…	…
		$\mathrm{Rol_I}$	$[\Psi_\mu]_M$	ID_μ	t_μ
	…	…	…	…	…
	$\mathrm{ID_G}$	$\mathrm{Rol_J}$	$[\Psi_\chi]_M$	ID_χ	t_χ
		…	…	…	…
		$\mathrm{Rol_L}$	$[\Psi_\tau]_M$	ID_τ	t_τ

Figura 5: Base de datos de confianza de una entidad para una aplicación concreta en el instante t

Se asume que las entidades poseen un identidad persistente. La persistencia del identificador es necesaria para que la entidad sea reconocible durante su tiempo de vida y no pueda ocultar su actividad previa.

Bajo una infrastructura de clave pública, ésta podría utilizarse como identificador puesto que toda entidad posee una y es única en el sistema. No obstante, las claves públicas pueden ser revocadas o caducar y ser

reemplazadas por otras nuevas.
Como alternativa proponemos que se utilice como identificador el resumen digital de la imagen inicial del proceso en memoria combinada con alguna otra fuente de entropía.

5.6. Aplicación de la regla de combinación de Dempster

Es posible que se obtengan diferentes cuerpos de evidencia relativas a cierto marco de discernimiento. En una situación como la descrita en la figura 6 la entidad A recibe información sobre la entidad D a través de dos fuentes diferentes:

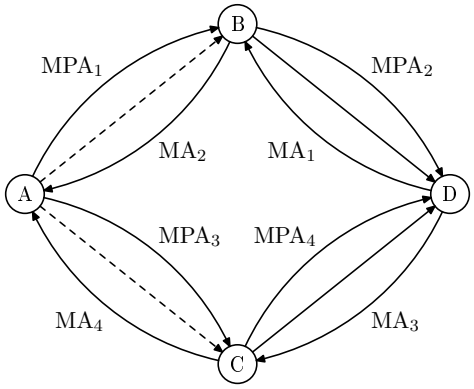

Figura 6: La entidad A recibe múltiples valores de creencia respecto a la entidad D

Existen tres posibles situaciones: (1) las evidencias se corroboran mutuamente, (2) las evidencias se contradicen pero son compatibles y (3) las evidencias son totalmente contradictorias.
Si las evidencias se corroboran entre sí:

$$\begin{array}{lllllllll} [\Psi_{B,D}(\{Y\})]_m & = & 0{,}6 & [\Psi_{B,D}(\{N\})]_m & = & 0{,}2 & [\Psi_{B,D}(\{Y,N\})]_m & = & 0{,}2 \\ [\Psi_{C,D}(\{Y\})]_m & = & 0{,}7 & [\Psi_{C,D}(\{N\})]_m & = & 0{,}2 & [\Psi_{C,D}(\{Y,N\})]_m & = & 0{,}1 \end{array}$$

la combinación de las evidencias mediante la regla de Dempster daría como resultado:

$$[\Psi(\{Y\})]_M = \frac{0{,}62}{0{,}74} \quad [\Psi(\{N\})]_M = \frac{0{,}1}{0{,}74} \quad [\Psi(\{Y,N\})]_M = \frac{0{,}02}{0{,}74}$$

Como era de esperar, el resultado obtenido también da soporte a la proposición Y.

Si las evidencias se contradicen pero son compatibles:

$$\begin{array}{lllllllll} [\Psi_{B,D}(\{Y\})]_m & = & 0{,}6 & [\Psi_{B,D}(\{N\})]_m & = & 0{,}2 & [\Psi_{B,D}(\{Y,N\})]_m & = & 0{,}2 \\ [\Psi_{C,D}(\{Y\})]_m & = & 0{,}2 & [\Psi_{C,D}(\{N\})]_m & = & 0{,}7 & [\Psi_{C,D}(\{Y,N\})]_m & = & 0{,}1 \end{array}$$

los valores de creencia resultantes de la combinación de las evidencias es la siguiente:

$$[\Psi(\{Y\})]_M = \frac{0{,}22}{0{,}54} \quad [\Psi(\{N\})]_M = \frac{0{,}3}{0{,}54} \quad [\Psi(\{Y,N\})]_M = \frac{0{,}02}{0{,}54}$$

El resultado obtenido es el esperado y la combinación favorece ligeramente a la proposición N.

Por el contrario, si las evidencias son totalmente contradictorias:

$$\begin{array}{lllllllll} [\Psi_{B,D}(\{Y\})]_m & = & 1 & [\Psi_{B,D}(\{N\})]_m & = & 0 & [\Psi_{B,D}(\{Y,N\})]_m & = & 0 \\ [\Psi_{C,D}(\{Y\})]_m & = & 0 & [\Psi_{C,D}(\{N\})]_m & = & 1 & [\Psi_{C,D}(\{Y,N\})]_m & = & 0 \end{array}$$

la combinación daría como resultado:

$$[\Psi(\{Y\})]_M = 0 \quad [\Psi(\{N\})]_M = 0 \quad [\Psi(\{Y,N\})]_M = 0$$

por tanto, $[\Psi(\emptyset)]_M = 1$ y nuestra ontología es inconsistente. En este caso se dice que no existe la suma ortogonal de las funciones de creencia.

5.7. Proyección de los valores de creencia

En [8] se demuestra que las funciones de distribución β pueden representar las probabilidades a posteriori de eventos binarios. La fórmula 16 muestra como dicha distribución, que está indexada por los parámetros σ y ω, puede expresarse en base a la función Γ:

$$\beta\left(\delta|\sigma,\omega\right) = \frac{\Gamma(\sigma+\omega+2k)}{\Gamma(\sigma+k)\Gamma(\omega+k)}\delta^{\sigma-1+k}(1-\delta)^{\omega-1+k} \tag{16}$$

donde $0 \leq \delta \leq 1$, $\sigma \geq 1-k$, $\omega \geq 1-k$ y $k \geq 0$.

La esperanza de la distribución β puede calcularse mediante la expresión:

$$\overline{\delta} = \frac{\sigma+k}{\sigma+\omega+2k} \tag{17}$$

Aplicando la fórmula 17, utilizando la partición $\{\{Y\},\{N\}\},\{\{Y,N\}\}$ de 2^{Θ_X} y tomando $[\Psi(\{Y,N\})]_M$ como parámetro k, podemos proyectar los valores de creencia sobre un espacio de probabilidades unidimensional, $\mathrm{E} : [\Psi]_M \rightarrow [0,1]$:

$$\mathrm{E}([\Psi]_M) = \frac{[\Psi(\{Y\})]_M + [\Psi(\{Y,N\})]_M}{[\Psi(\{Y\})]_M + [\Psi(\{N\})]_M + 2\,[\Psi(\{Y,N\})]_M} \tag{18}$$

Dado que al proyectar estamos eliminando información (la componente de incertidumbre), la relación entre los valores de creencia y sus proyecciones es unívoca pero no biunívoca, es decir, que pueden producirse colisiones.

$$
\begin{array}{lcll}
[\Psi_1]_M & = & \{1,0,0\} & \mathrm{E}([\Psi_1]_{\mathrm{M}}) = \quad 1 \\
[\Psi_2]_M & = & \{0,1,0\} & \mathrm{E}([\Psi_2]_{\mathrm{M}}) = \quad 0 \\
[\Psi_3]_M & = & \{0,0,1\} & \mathrm{E}([\Psi_3]_{\mathrm{M}}) = \quad \frac{1}{2} \\
[\Psi_4]_M & = & \left\{\frac{1}{2},0,\frac{1}{2}\right\} & \mathrm{E}([\Psi_4]_{\mathrm{M}}) = \quad \frac{2}{3} \\
[\Psi_5]_M & = & \left\{\frac{1}{2},\frac{1}{2},0\right\} & \mathrm{E}([\Psi_5]_{\mathrm{M}}) = \quad \frac{1}{2} \\
[\Psi_6]_M & = & \left\{0,\frac{1}{2},\frac{1}{2}\right\} & \mathrm{E}([\Psi_6]_{\mathrm{M}}) = \quad \frac{1}{3}
\end{array}
\tag{19}
$$

Figura 7: Proyección de los valores de creencia en un espacio unidimensional

En el ejemplo de la figura 7 puede observarse que $\mathrm{E}([\Psi_3]_{\mathrm{M}}) = \mathrm{E}([\Psi_5]_{\mathrm{M}})$.
No obstante, dadas dos funciones de masa $[\Psi_i]_M$, $[\Psi_j]_M$ definidas sobre el mismo marco de discernimiento, podemos establecer la relación de comparación $>$ como:

$$
[\Psi_i]_M > [\Psi_j]_M \equiv \begin{cases} \mathrm{E}([\Psi_{\mathrm{i}}]_{\mathrm{M}}) > \mathrm{E}([\Psi_{\mathrm{j}}]_{\mathrm{M}}) \\ \vee \\ \mathrm{E}([\Psi_{\mathrm{i}}]_{\mathrm{M}}) = \mathrm{E}([\Psi_{\mathrm{j}}]_{\mathrm{M}}) \quad \wedge \quad [\Psi_i(\{Y,N\})]_M < [\Psi_j(\{Y,N\})]_M \end{cases}
\tag{20}
$$

Por tanto, $[\Psi_5]_M > [\Psi_3]_M$.

Es importante destacar que es posible generalizar las distribución β para que considere un número arbitrario de eventos. Por tanto, sería posible proyectar valores de creencia asociados a un marco de discernimiento de cardinal arbitrario.

5.8. Reevaluación de evidencias

Para reflejar la volatilidad de las relaciones de confianza y asesoramiento, es necesario que las entidades puedan rectificar sus valores de creencia y difundirlas a los nodos a los que proporcionaron evidencias. Esto se llevará a la práctica mediante procesos de decisión sucesivos. Por ello, aunque las evidencias podrían agregarse mediante la regla de combinación de Dempster, reduciendo el tamaño de la base de datos, es necesario conservarlas.
En nuestro modelo, el tiempo también afectará a los valores de creencia. En este sentido, las evidencias vinculadas a las relaciones de recomendación y de confianza, a las que se asocia una marca de tiempo, tienen un periodo de validez finito.

5.9. Proceso de decisión tolerante a ataques

Las recomendaciones deben ser interpretadas como los valores de creencia que una entidad proporciona a otra, pero no necesariamente como los valores de creencia que realmente sostiene. Estos podrían ser diferentes en caso de que el nodo recomendante fuese malicioso o defectuoso.

Aunque la regla de combinación de Dempster posee propiedades que la convierten en una mecanismo de agregación adecuado, asume que todas las evidencias son legítimas. Si consideramos la expresión 21 es posible apreciar el efecto de agregar tres evidencias que se corroboran entre sí:

$$\left.\begin{array}{rcl} [\Psi_1]_M & = & \{0.7,\ 0.1,\ 0.2\} \\ [\Psi_2]_M & = & \{0.8,\ 0,\ 0.2\} \\ [\Psi_3]_M & = & \{0.6,\ 0.1,\ 0.3\} \end{array}\right\} \rightarrow [\Psi]_M = \{0.971,\ 0.015,\ 0.015\} \quad (21)$$

En la expresión 22 una sola evidencia, extremadamente contradictoria, altera significativamente los valores de creencia. Este hecho hace sensible a la regla de Dempster a la presencia de nodos defectuosos o maliciosos.

$$\left.\begin{array}{rcl} [\Psi_1]_M & = & \{0.7,\ 0.1,\ 0.2\} \\ [\Psi_2]_M & = & \{0.8,\ 0,\ 0.2\} \\ [\Psi_3]_M & = & \{0.6,\ 0.1,\ 0.3\} \\ [\Psi_4]_M & = & \{0.0,\ 0.985,\ 0.015\} \end{array}\right\} \rightarrow [\Psi]_M = \{0.333,\ 0.662,\ 0.005\} \quad (22)$$

Si κ es el factor de normalización de la regla de combinación de Dempster, entonces es posible calcular el grado de contradicción entre dos evidencias definidas sobre el marco de discernimiento Θ_X mediante la fórmula:

$$\log(\kappa) = -\log\left(1 - \sum_{A_i \cap B_j = \emptyset} [\Psi_1(A_i)]_M [\Psi_2(B_j)]_M\right) = -\log(1 - \Phi) \quad (23)$$

donde $[\Psi_1]_M$,$[\Psi_2]_M$ son las funciones de masa que están siendo combinadas y A_i,B_j son, respectivamente, elementos focales de dichas funciones de masa. El valor $\log(\kappa)$ pertenece al intervalo $[0, \infty)$.
La figura 8 muestra el grado de contradicción entre las evidencias del ejemplo anterior:

$[\Psi_i]_M$	$[\Psi_i]_M$	$\log(\kappa)$
{0.7, 0.1, 0.2}	{0.0, 0.985, 0.015}	0.508
{0.8, 0, 0.2}	{0.0, 0.985, 0.015}	0.674
{0.6, 0.1, 0.3}	{0.0, 0.985, 0.015}	0.388

Figura 8: Cálculo del grado de contradicción entre evidencias

En [24] Reiter introduce la importancia de la redundancia en la toma de decisiones. Utilizando este concepto, y considerando el grado de contradicción entre evidencias, es posible definir un algoritmo que particione el espacio de hipótesis cuando existan evidencias contradictorias en la base de hechos, para reducir el efecto erosivo.
Durante el proceso de decisión, cada nodo podrá aportar una única opinión en relación a otro agente, en cierto contexto. Dicho proceso finalizará cuando existan suficientes evidencias que den soporte a una de las hipótesis del marco de discernimiento. Posteriormente, algunos nodos podrían reevaluar los valores de creencia que ofrecieron, iniciándose otro proceso de votación.

En nuestro modelo, el grado de sospecha se incrementa paulatinamente y acaba por aislar totalmente a los nodos maliciosos impidiéndoles participar en el proceso de recomendación. Una vez un nodo es introducido en el conjunto de sospechosos quedará marcado como nodo malicioso permanentemente. De esta forma, los nodos maliciosos no podrán recuperar credibilidad exhibiendo, de un modo ocasional, comportamiento colaborativo.
La condición de parada es una función del número de nodos activos en el sistema. Dicho valor se estima a partir de los mensajes intercambiados por los agentes en el proceso de evaluación de la confianza. Conforme el número de intermediarios, vinculados mediante relaciones de recomendación, se incremente será necesario aportar más evidencias para alcanzar el umbral de decisión. Los nodos se eliminarán del conjunto de nodos activos cuando expire el tiempo de vida asociado a la referencia, cuando el nodo sea calificado como sospechoso o cuando se revoque el certificado asociado al agente.

Funciones auxiliares:

$\langle [\Psi]_M, id, t \rangle$ median$_Y(\epsilon)$:
    Sort ϵ;
    if $|\epsilon| \mod 2 \neq 0$ **then**
        $i := \left\lceil \frac{|\epsilon|}{2} \right\rceil$;
    else
        $i := \frac{|\epsilon|}{2}$;
    end
    return ϵ_i;
end

$\langle [\Psi]_M, id, t \rangle$ median$_N(\epsilon)$:
    Sort ϵ;
    if $|\epsilon| \mod 2 \neq 0$ **then**
        $i := \left\lceil \frac{|\epsilon|}{2} \right\rceil$;
    else
        $i := \left\lceil \frac{|\epsilon|+1}{2} \right\rceil$;
    end
    return ϵ_i;
end

void reset():
    for id **in** activos **do**
        vote$_{id} := false$;
    end
    $\epsilon := \emptyset$;
    $\epsilon_Y := \emptyset$;
    $\epsilon_N := \emptyset$;
    rep$_\epsilon := [\Psi_\emptyset]_M$; // $\{0, 0, 0\} \equiv [\Psi_\emptyset]_M$
    rep$_{\epsilon_Y} := [\Psi_\emptyset]_M$;
    rep$_{\epsilon_N} := [\Psi_\emptyset]_M$;

```
        contradiccion:=false;
end
```

```
[Ψ]_M  aggregate_evidences():
        E = ε_1.[Ψ]_M;
        for ε_i in ε − {ε_1} do
                E = E ⊕ ε_i.[Ψ]_M;
        end
        return E;
end
```

Inicialización:

```
activos := ∅;
sospechosos := ∅;
ε := ∅;
ε_Y := ∅;
ε_N := ∅;
rep_ε := [Ψ_∅]_M ; // {0, 0, 0} ≡ [Ψ_∅]_M
rep_{ε_Y} := [Ψ_∅]_M;
rep_{ε_N} := [Ψ_∅]_M;
contradiccion:=false;
```

```
while(true) do

void add_evidence():
        Precondición:
                m ∈ rcvq ∧ vote_{m.id} = false ∧ m.id ∉ sospechosos ∧
                m.[Ψ]_M ≠ {1,0,0} ∧ m.[Ψ]_M ≠ {0,1,0} ∧ m.[Ψ]_M ≠ {0,0,1} ∧
                m.[Ψ]_M ≠ {0,0,0}
        Efecto:
                m:=rcv();
                vote_{m.id} := true;
                for id in m.PATH−{id_self} do
                        activos := activos ∪ {id};
                end
                activos := activos − sospechosos;
                if contradiccion = false then
                        α := −log(1 − Φ_{m.[Ψ]_M, rep_ε.[Ψ]_M});
                        if α < Ω_∅ then
                                ε := ε ∪ ⟨m.[Ψ]_M, m.id, m.t⟩;
                                rep_ε := median_Y(ε);
                                if |ε| ≥ Π then
                                        result:=aggregate_evidences();
                                        reset();
                                end
                        else
                                contradiccion:=true;
                                ε := ε ∪ ⟨m.[Ψ]_M, m.id, m.t⟩;
                                max_ε := {ε_i ∈ ε | ε_i.[Ψ]_M > ε_j.[Ψ]_M, ∀ε_j ∈ ε, i ≠ j};
                                min_ε := {ε_i ∈ ε | ε_j.[Ψ]_M > ε_i.[Ψ]_M, ∀ε_j ∈ ε, i ≠ j};
                                ε_Y := {max_ε};
```

```
30                  ϵ_N := {min_ϵ};
31                  for ϵ_i in ϵ − {max_ϵ, min_ϵ} do
32                      α_Y := −log(1 − Φ_{ϵ_i.[Ψ]_M, max_ϵ.[Ψ]_M});
33                      α_N := −log(1 − Φ_{ϵ_i.[Ψ]_M, min_ϵ.[Ψ]_M});
34                      if α_Y < α_N then
35                          ϵ_Y := ϵ_Y ∪ ϵ_i;
36                      else
37                          if α_N < α_Y then
38                              ϵ_N := ϵ_N ∪ ϵ_i;
39                          else
40                              ϵ := ϵ − {ϵ_i};
41                          end
42                      end
43                  end
44                  rep_{ϵ_N} := median_N(ϵ_N);
45                  rep_{ϵ_Y} := median_Y(ϵ_Y);
46                  n := max(Π, |activos|);
47                  if |ϵ_Y| ≥ ⌈(2n + 1)/3⌉ then
48                      ϵ := ϵ_Y;
49                      for ϵ_i in ϵ_N do
50                          Sospecha_{ϵ_i.id} := Sospecha_{ϵ_i.id} + Δ;
51                          if Sospecha_{ϵ_i.id} ≥ δ then
52                              sospechosos := sospechosos ∪ ϵ_i.id
53                          end
54                      end
55                      result:=aggregate_evidences();
56                      reset();
57                  else
58                      if |ϵ_N| ≥ ⌈(2n + 1)/3⌉ then
59                          ϵ := ϵ_N;
60                          for ϵ_i in ϵ_Y do
61                              Sospecha_{ϵ_i.id} := Sospecha_{ϵ_i.id} + Δ;
62                              if Sospecha_{ϵ_i.id} ≥ δ then
63                                  sospechosos := sospechosos ∪ ϵ_i.id
64                              end
65                          end
66                          result:=aggregate_evidences();
67                          reset();
68                      end
69                  end
70              end
71          else
72              α_Y := −log(1 − Φ_{m.[Ψ]_M, rep_{ϵ_Y}.[Ψ]_M});
73              α_N := −log(1 − Φ_{m.[Ψ]_M, rep_{ϵ_N}.[Ψ]_M});
74              if α_Y < α_N then
75                  ϵ_Y := ϵ_Y ∪ ⟨m.[Ψ]_M, m.id, m.t⟩;
76                  rep_{ϵ_Y} := median_Y(ϵ_Y);
77              else
78                  if α_N < α_Y then
79                      ϵ_N := ϵ_N ∪ ⟨m.[Ψ]_M, m.id, m.t⟩;
80                      rep_{ϵ_N} := median_N(ϵ_N);
81                  end
82              end
83              n := max(Π, |activos|);
```

```
84                  if |ϵ_Y| ≥ ⌈(2n+1)/3⌉ then
85                      ϵ := ϵ_Y;
86                      for ϵ_i in ϵ_N do
87                          Sospecha_{ϵ_i.id} := Sospecha_{ϵ_i.id} + Δ;
88                          if Sospecha_{ϵ_i.id} ≥ δ then
89                              sospechosos := sospechosos ∪ ϵ_i.id
90                          end
91                      end
92                      result:=aggregate_evidences();
93                      reset();
94                  else
95                      if |ϵ_N| ≥ ⌈(2n+1)/3⌉ then
96                          ϵ := ϵ_N;
97                          for ϵ_i in ϵ_Y do
98                              Sospecha_{ϵ_i.id} := Sospecha_{ϵ_i.id} + Δ;
99                              if Sospecha_{ϵ_i.id} ≥ δ then
100                                 sospechosos := sospechosos ∪ ϵ_i.id
101                             end
102                         end
103                         result:=aggregate_evidences();
104                         reset();
105                     end
106                 end
107             end
108         end
109 end
```

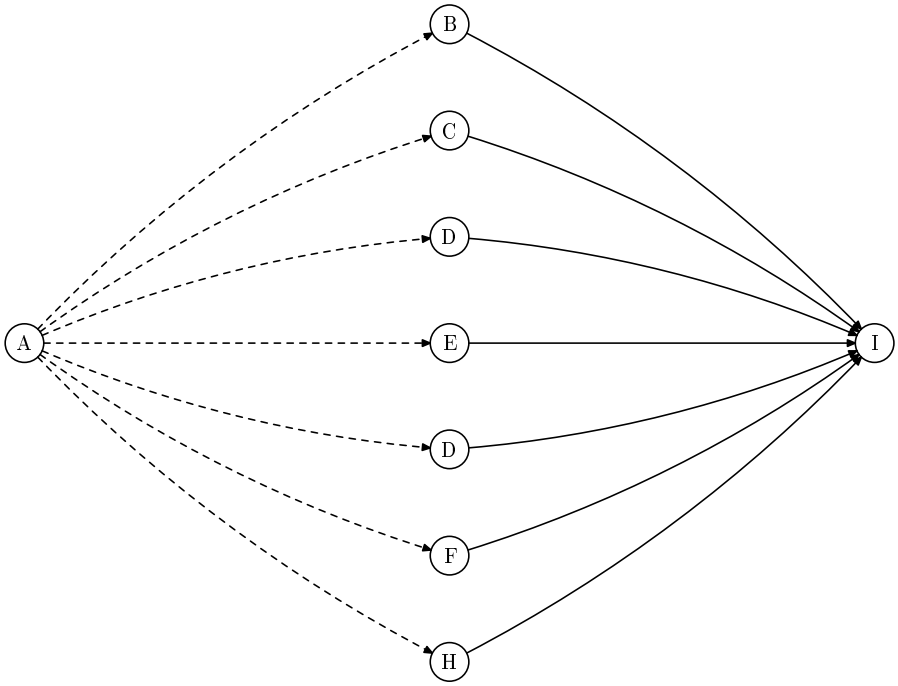

Figura 9: Secuencia de mensajes de petición de asesoramiento

Dado el ejemplo de la figura 9, en el que el nodo A ha pedido asesoramiento sobre cierto nodo cuyo perfil coincide con el del agente I, suponiendo que la secuencia de evidencias que llegasen a A fuese la siguiente:

$\langle \{0.7,\ 0.1,\ 0.2\}, id_B, t_1 \rangle$
$\langle \{0.8,\ 0,\ 0.2\}, id_C, t_2 \rangle$
$\langle \{0.2,\ 0.7,\ 0.1\}, id_D, t_3 \rangle$
$\langle \{0.6,\ 0.2,\ 0.2\}, id_E, t_4 \rangle$
$\langle \{0.1,\ 0.6,\ 0.3\}, id_F, t_5 \rangle$
$\langle \{0.65,\ 0.1,\ 0.25\}, id_G, t_6 \rangle$
$\langle \{0.5,\ 0.2,\ 0.3\}, id_H, t_7 \rangle$

y tomando $\Omega_\emptyset = 0{,}2$ y $\Pi = 5$, la traza del algoritmo de decisión sería:

$\langle \{0.7,\ 0.1,\ 0.2\}, id_B, t_1 \rangle \rightarrow A,\ \ E([\Psi_B]_M) = 0.75$
activos $= \{B\}, \epsilon = \{B_\epsilon\}, rep_\epsilon = B_\epsilon$

$\langle \{0.8,\ 0,\ 0.2\}, id_C, t_2 \rangle \rightarrow A,\ \ E([\Psi_C]_M) = 0.83$
activos $= \{B, C\}, \alpha = 0.04$
$\epsilon = \{B_\epsilon, C_\epsilon\}, rep_\epsilon = B_\epsilon$

$\langle \{0.2,\ 0.7,\ 0.1\}, id_D, t_3 \rangle \rightarrow A,\ \ E([\Psi_D]_M) = 0.27$
activos $= \{B, C, D\}, \alpha = 0.31, max_\epsilon = \{C_\epsilon\}, min_\epsilon = \{D_\epsilon\}, \alpha_Y = 0.04, \alpha_N = 0.31$
$\epsilon_Y = \{B_\epsilon, C_\epsilon\}, \epsilon_N = \{D_\epsilon\}$
$rep_{\epsilon_Y} = B_\epsilon$
$rep_{\epsilon_N} = D_\epsilon$

$\langle \{0.6,\ 0.2,\ 0.2\}, id_E, t_4 \rangle \rightarrow A,\ \ E([\Psi_E]_M) = 0.67$
activos $= \{B, C, D, E\}, \alpha_Y = 0.1, \alpha_N = 0.27$
$\epsilon_Y = \{B_\epsilon, C_\epsilon, E_\epsilon\}, \epsilon_N = \{D_\epsilon\}$
$rep_{\epsilon_Y} = B_\epsilon$
$rep_{\epsilon_N} = D_\epsilon$

$\langle \{0.1,\ 0.6,\ 0.3\}, id_F, t_5 \rangle \rightarrow A,\ \ E([\Psi_F]_M) = 0.31$
activos $= \{B, C, D, E, F\}, \alpha_Y = 0.24, \alpha_N = 0.09$
$\epsilon_Y = \{B_\epsilon, C_\epsilon, E_\epsilon\}, \epsilon_N = \{D_\epsilon, F_\epsilon\}$
$rep_{\epsilon_Y} = B_\epsilon$
$rep_{\epsilon_N} = F_\epsilon$

$\langle \{0.65,\ 0.1,\ 0.25\}, id_G, t_6 \rangle \rightarrow A,\ \ E([\Psi_G]_M) = 0.72$
activos $= \{B, C, D, E, F, G\}, \alpha_Y = 0.06, \alpha_N = 0.22$
$\epsilon_Y = \{B_\epsilon, C_\epsilon, E_\epsilon, G_\epsilon\}, \epsilon_N = \{D_\epsilon, F_\epsilon\}$
$rep_{\epsilon_Y} = G_\epsilon$
$rep_{\epsilon_N} = F_\epsilon$

$\langle \{0.5,\ 0.2,\ 0.3\}, id_H, t_7 \rangle \rightarrow A,\ \ E([\Psi_H]_M) = 0.62$
activos $= \{B, C, D, E, F, G, H\}, \alpha_Y = 0.09, \alpha_N = 0.17$
$\epsilon_Y = \{B_\epsilon, C_\epsilon, E_\epsilon, G_\epsilon, H_\epsilon\}, \epsilon_N = \{D_\epsilon, F_\epsilon\}$
$rep_{\epsilon_Y} = G_\epsilon$
$rep_{\epsilon_N} = F_\epsilon$

$result := \{0.992,\ 0.007,\ 0.001\}$

El algoritmo se detendría en este punto, una vez el cardinal de la partición que da soporte a la confiabilidad del nodo I ha alcanzado dos terceras partes del número de nodos activos. A raíz de esta votación, el grado de sospecha sobre los nodos D y F se incrementaría en un valor Δ.

6

Protocolo de intercambio, revocación y reevaluación de criterios de confianza

El protocolo que presentamos en este apartado sería compatible con representaciones de la confianza alternativas a las funciones de creencia, así como con bases de datos de confianza con un formato diferente al propuesto en el apartado anterior.
Aunque ciertos nodos intermedios no fuesen accesibles en determinado instante debido a sobrecarga o fallo, los valores de creencia podrían ser transferidos al nodo asesorado. Éste, en función de la longitud de la cadena de confianza resultante y de sus componentes, podría estimar la validez de dichos valores. En este sentido, el algoritmo proporciona cierta tolerancia a fallos.

6.1. Estructura de los mensajes

El protocolo de gestión de la confianza se basa en tres tipos de mensajes: mensajes de petición de asesoramiento, mensajes de asesoramiento y mensajes de refresco.
Una entidad puede solicitar recomendaciones enviando mensajes de petición de asesoramiento (*MPA*) a aquellas entidades con las que mantiene una relación de asesoramiento. Como respuesta podrá recibir uno o más mensajes de asesoramiento (*MA*). Las recomendaciones podrán reevaluarse o revocarse en base a los mensajes de refresco (*MDR*).
La autenticación de los mensajes impide que los nodos maliciosos o defectuosos puedan afectar, sin ayuda, a la reputación o las relaciones de confianza de otros nodos.

6.1.1. Mensaje de petición de asesoramiento (*MPA*)

Un mensaje *MPA* tiene los siguiente campos:

1. $\mathrm{ID_S}$: Es el identificador de la entidad que solicita asesoramiento.
2. $\mathrm{REQUEST_{ID}}$: Es el identificador del mensaje de petición de asesoramiento. Este campo debe tener suficiente variabilidad para identificar unívocamente el mensaje.
3. REQUEST HISTORY: Es la secuencia de identificadores de mensajes de petición de asesoramiento originada en la búsqueda. Es necesario para permitir reevaluar o revocar el asesoramiento.
4. $\mathrm{VALIDITY}$: Indica el periodo de validez del mensaje.
5. QUERY: Es el conjunto de características que definen el perfil de entidad en la que el asesorado está interesado. Esta consulta consistirá en una serie de características relacionadas mediante operadores lógicos de conjunción, disyunción y negación. Si dichas características fuesen roles, la consulta podría ser una fórmula similar a:

   $$(\mathrm{Rol_A} \wedge \mathrm{Rol_B} \wedge \neg \mathrm{Rol_C}) \vee \mathrm{Rol_D}$$

   En este caso, la entidad buscaría asesoramiento sobre entidades que o bien fuesen miembro del $\mathrm{Rol_D}$ o que fuesen simultáneamente miembro de los roles $\mathrm{Rol_A}$ y $\mathrm{Rol_B}$ pero no del $\mathrm{Rol_C}$.
6. PATH: Es la secuencia de identificadores de las entidades asesoras generada en la búsqueda. Se utiliza para prevenir la formación de ciclos. Antes de enviar un mensaje de petición de asesoramiento a un nodo puede comprobarse si dicho nodo ya aparece en la secuencia del PATH, en cuyo caso se aborta el envío.
7. $\mathrm{AUTHENTICATOR}$: Se utiliza para validar la autenticidad de la consulta y, posteriormente, la validez de las respuestas.

Mensaje de asesoramiento (*MA*)
Un mensaje *MA* tiene los siguiente campos:

1. $\mathrm{ID_S}$: Es el identificador de la entidad que solicitó el asesoramiento. Se obtiene del mensaje *MPA* asociado.
2. $\mathrm{REQUEST_{ID}}$: Es el identificador del mensaje de petición de asesoramiento asociado a este mensaje de asesoramiento.
3. REQUEST HISTORY: Es la secuencia de identificadores de mensajes de petición de asesoramiento originada en la búsqueda. Es necesario para permitir reevaluar o revocar el asesoramiento. Se obtiene del mensaje *MPA* asociado a este mensaje de asesoramiento.
4. PATH: Es la secuencia completa de identificadores de las entidades asesoras que han intervenido en el camino de asesoramiento. Este campo toma valor cuando la búsqueda alcanza una entidad que tiene una relación de confianza con otra que posee las características que buscamos. Después, en los restantes mensajes *MA*, sólo se difunde.
5. AUTHENTICATOR: Es una cadena de confianza. Cada asesor de la secuencia firma el camino de asesoramiento. Incluye la petición de asesoramiento original y la recomendación.
6. AUTHINFO: Es la información asociada a la entidad o entidades recomendadas. Incluirá un conjunto de valores de creencia, el periodo de validez de la recomendación, los identificadores correspondientes y sus certificados de clave pública. Este campo toma valor cuando la búsqueda alcanza una entidad con una relación de confianza con una entidad que presenta el perfil de interés. A partir de este punto, sólo se difunde en los restantes mensajes *MA*.

Mensaje de refresco (*MDR*)
Para que las recomendaciones puedan reevaluarse o revocarse en base a mensajes *MDR*, todos los nodos deben almacenar un histórico de los mensajes *MA* que han enviado. Estos mensajes proporcionan la información necesaria para hacer llegar la reevaluación o revocación al nodo que originalmente realizó la petición de asesoramiento, pasando por todos los asesores que tomaron partido.
Un mensaje *MDR* tiene los siguientes campos:

1. $\mathrm{ID_{AS}}$: Es el identificador de la entidad asesora que ha decidido modificar su recomendación.
2. REQUEST HISTORY: Es la secuencia de identificadores de mensajes de petición de asesoramiento originada en la búsqueda. Es necesario para permitir reevaluar o revocar el asesoramiento. Se obtiene del mensaje *MA*.
3. PATH: Es la secuencia completa de los identificadores de las entidades asesoras que han intervenido en el camino de asesoramiento. Se obtiene del mensaje *MA*.
4. AUTHENTICATOR: Es una cadena de confianza. Cada asesor de la secuencia firma el camino de asesoramiento. Se obtiene a partir del mensaje *MA*.
5. AUTHINFO: Es la información de la entidad o entidades que se recomendaron, reevaluada.

6.2. Traza del algoritmo de evaluación, reevaluación y revocación de la confianza

Supongamos que A busca entidades de confianza que estén autorizadas a utilizar el rol $\mathrm{Rol_A}$ o el rol $\mathrm{Rol_B}$. A mantiene una relación de asesoramiento con B (representada con flechas de trazo discontinuo). A su vez, B mantiene relaciones de asesoramiento con C y E, y estos mantienen relaciones de confianza explícita (representada por flechas en color azul) con las entidades D y F, respectivamente. Si D es miembro del rol $\mathrm{Rol_A}$ y F es miembro del rol $\mathrm{Rol_B}$, podemos considerar la siguiente secuencia de mensajes:

6.2.1. Secuencia de mensajes de petición de asesoramiento

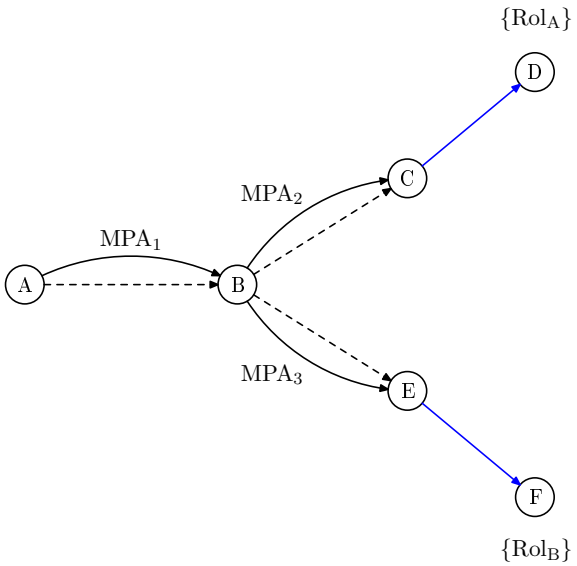

Figura 10: Secuencia de mensajes de petición de asesoramiento

MPA_1
$\mathrm{ID_S} = \mathrm{ID_A}$
$\mathrm{REQUEST_{ID}} = \mathrm{RQ0001}$
$\mathrm{REQUEST\,HISTORY} = \emptyset$
$\mathrm{VALIDITY} = \mathrm{validity_1}$
$\mathrm{QUERY} = \mathrm{Rol_A} \vee \mathrm{Rol_B}$
$\mathrm{PATH} = \{\mathrm{ID_A}, \mathrm{ID_B}\}$
$\mathrm{AUTHENTICATOR} = \mathrm{K_A^-}(\mathrm{ID_A}\|\|\mathrm{QUERY}\|\|\mathrm{TS_1})$

MPA_2
$\mathrm{ID_S} = \mathrm{ID_B}$
$\mathrm{REQUEST_{ID}} = \mathrm{RQ0002}$
$\mathrm{REQUEST\,HISTORY} = \mathrm{RQ0001}$
$\mathrm{VALIDITY} = \mathrm{validity_2}$
$\mathrm{QUERY} = \mathrm{Rol_A} \vee \mathrm{Rol_B}$
$\mathrm{PATH} = \{\mathrm{ID_A}, \mathrm{ID_B}, \mathrm{ID_C}\}$
$\mathrm{AUTHENTICATOR} = \mathrm{K_B^-}(\mathrm{K_A^-}(\mathrm{ID_A}\|\|\mathrm{QUERY}\|\|\mathrm{TS_1}))$

MPA_3

$ID_S = ID_B$

$REQUEST_{ID} = RQ0003$

$REQUEST\,HISTORY = RQ0001$

$VALIDITY = validity_3$

$QUERY = Rol_A \vee Rol_B$

$PATH = \{ID_A, ID_B, ID_E\}$

$AUTHENTICATOR = K_B^-(K_A^-(ID_A||QUERY||TS_1))$

6.2.2. Secuencia de mensajes de asesoramiento

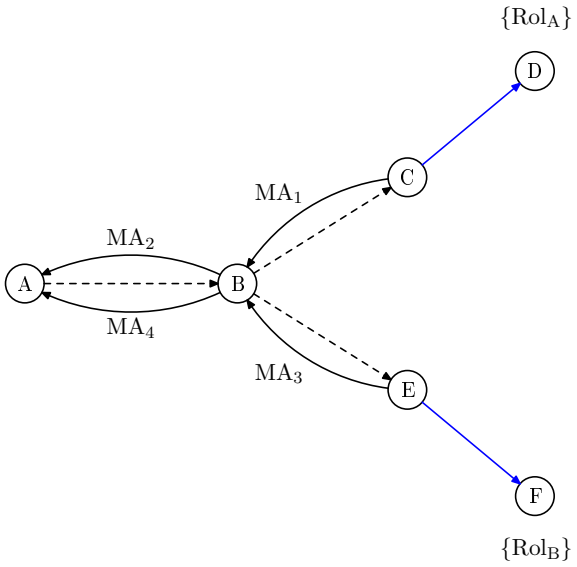

Figura 11: Secuencia de mensajes de asesoramiento

MA_1

$ID_S = ID_B$

$REQUEST_{ID} = RQ0002$

$REQUEST\,HISTORY = RQ0001$

$PATH = \{ID_A, ID_B, ID_C\}$

$AUTHENTICATOR =$

$K_C^-(ID_A||ID_B||ID_C||K_A^-(ID_A||QUERY||TS_1)||C\,Rc_t^c\,D)$

$AUTHINFO =$

$ID_D, Cert_D, validity_x, C\,Rc_t^c\,D$

MA_2

$ID_S = ID_A$

$REQUEST_{ID} = RQ0001$

$REQUEST\,HISTORY = \emptyset$

$PATH = \{ID_A, ID_B, ID_C\}$

$AUTHENTICATOR =$

$K_B^-(K_C^-(ID_A||ID_B||ID_C||K_A^-(ID_A||QUERY||TS_1)||C\,Rc_t^c\,D))$

$AUTHINFO =$

$ID_D, Cert_D, validity_x, C\,Rc_t^c\,D$

MA_3

$ID_S = ID_B$

$REQUEST_{ID} = RQ0003$

$REQUEST\,HISTORY = RQ0001$

$PATH = \{ID_A, ID_B, ID_E\}$

$AUTHENTICATOR =$

$K_E^-(ID_A||ID_B||ID_E||K_A^-(ID_A||QUERY||TS_1)||E\,Rc_t^c\,F)$

$AUTHINFO =$

$ID_F, Cert_F, validity_y, E\,Rc_t^c\,F$

MA_4

$ID_S = ID_A$

$REQUEST_{ID} = RQ0001$

$REQUEST\,HISTORY = \emptyset$

$PATH = \{ID_A, ID_B, ID_E\}$

$AUTHENTICATOR =$

$K_B^-(K_E^-(ID_A||ID_B||ID_E||K_A^-(ID_A||QUERY||TS_1)||E\,Rc_t^c\,F))$

$AUTHINFO =$

$ID_F, Cert_F, validity_y, E\,Rc_t^c\,F$

6.2.3. Secuencia de mensajes de refresco

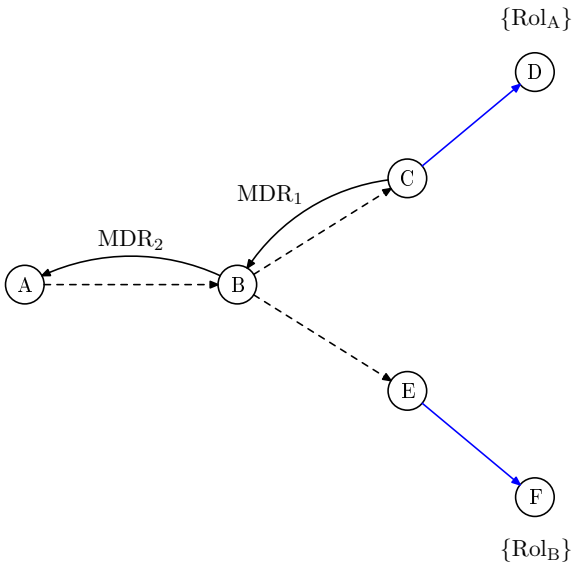

Figura 12: Secuencia de mensajes de refresco

C decide modificar todas las recomendaciones que ha hecho sobre D. Para ello busca entre los mensajes *MA* que almacena y selecciona aquéllos en los que el campo *AUTHINFO* haga referencia a D. Localiza MA_1 y comprueba que no ha caducado la validez de su recomendación (campo $\mathrm{validity_x}$ del campo *AUTHINFO* del mensaje MA_1). A partir de los campos de MA_1 genera el mensaje de refresco MDR_1, en el que ha modificado el campo AUTHINFO original.

MDR_1
$\mathrm{ID_{AS}} = \mathrm{ID_C}$
$\mathrm{REQUEST\ HISTORY} = \mathrm{RQ0001}$
$\mathrm{PATH} = \{\mathrm{ID_A}, \mathrm{ID_B}, \mathrm{ID_C}\}$
$\mathrm{AUTHENTICATOR} =$
$\mathrm{K_C^-(ID_A||ID_B||ID_C||K_A^-(ID_A||QUERY||TS_1)||C\ Rc_{t'}^c\ D)}$
$\mathrm{AUTHINFO} =$
$\mathrm{ID_D, Cert_D, validity_x', C\ Rc_{t'}^c\ D}$

A partir del campo $\mathrm{ID_S}$ del mensaje MA_1, C sabe que debe enviárselo a B. B busca entre los mensajes *MA* que almacena, aquél cuyo campo $\mathrm{REQUEST_{ID}}$ coincida con el último elemento del campo $\mathrm{REQUEST\ HISTORY}$ del mensaje MDR_1 y cuyo campo PATH coincida con el del mensaje MDR_1. Encuentra MA_2 y genera el mensaje:

MDR_2

$\mathrm{ID_{AS}} = \mathrm{ID_C}$

$\mathrm{REQUEST\ HISTORY} = \emptyset$

$\mathrm{PATH} = \{\mathrm{ID_A}, \mathrm{ID_B}, \mathrm{ID_C}\}$

$\mathrm{AUTHENTICATOR} =$

$\mathrm{K_B^-(K_C^-(ID_A||ID_B||ID_C||K_A^-(ID_A||QUERY||TS_1)||C\ Rc_{t'}^c\ D))}$

$\mathrm{AUTHINFO} =$

$\mathrm{ID_D, Cert_D, validity'_x, C\ Rc_{t'}^c\ D}$

A recibe el mensaje y dado que el campo REQUEST HISTORY está vacío y que su identificador encabeza el campo PATH, sabe que no debe reenviar el mensaje porque va dirigido a él. Puede comprobar el autenticador para verificar que la secuencia del PATH coincide con la del autenticador.

7

Conclusiones y trabajo futuro

Entre las aproximaciones previas a la gestión de la confianza analizadas en este documento no existe ninguna de carácter genérico, sólo un reducido número de ellas considera la naturaleza contextual de la confianza, no todas utilizan mecanismos de combinación de evidencias que permitan reevaluar los valores de creencia y ninguna puede representar de un modo claro y preciso la incertidumbre asociada a las relaciones de confianza.

El algoritmo de toma de decisiones que proponemos está altamente parametrizado, por lo que podría adoptar enfoques progresistas o conservadores en función de las necesidades de los agentes. Aunque el formalismo que hemos utilizado para representar la confianza y agregar las evidencias permitiría establecer un número arbitrario de hipótesis, en este trabajo, por simplicidad, consideramos el marco de discernimiento más sencillo. No obstante, el mecanismo de proyección basado en la distribución β y la relación de orden que proponemos podrían adaptarse a un número arbitrario de eventos.

En resumen, creemos que el modelo que proponemos es suficientemente genérico para ser incorporado como un servicio distribuido de gestión de la confianza de propósito general, que podría ser utilizado por los agentes para establecer redes de confianza.

La incertidumbre se incorpora a dos niveles en el modelo. Primero, en la representación no Bayesiana de la confianza, utilizando funciones de creencia. Segundo, en el algoritmo de evaluación de las evidencias que utiliza el concepto de redundancia y la estimación de los agentes activos en el sistema para la toma de decisiones, sin hacer presunciones basadas en la marginalidad de los valores. La evaluación de la confianza en nuestro modelo refleja su carácter dinámico, permitiendo la reevaluación de los valores de creencia no sólo por interacción directa o por factores temporales, sino también a través de la rectificación por parte de los agentes asesores.

Finalmente, pensamos que sería de interés formular un protocolo matemático que incorporase ciertas reglas de juego que permitiesen establecer los valores iniciales de creencia de los nodos y, a su vez, minimizar el riesgo asociado a las interacciones iniciales entre los agentes del sistema. Dicho algoritmo podría estar basado en la teoría de juegos y la teoría del caos.

Bibliografía

[1] Alfarez Abdul-Rahman y Stephen Hailes. Supporting trust in virtual communities. En *HICSS '00: Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 6*, pág. 6007, Washington, DC, USA, 2000. IEEE Computer Society.

[2] Thomas Beth, Malte Borcherding y Birgit Klein. Valuation of trust in open networks. págs. 3–18. Springer-Verlag, 1994.

[3] Jim Binkley y William Trost. Authenticated ad hoc routing at the link layer for mobile systems. En Springer Netherlands, editor, *Wireless Networks*, volumen 7, págs. 139–145. 2001.

[4] Kenneth P. Birman y Thomas A. Joseph. Reliable communication in the presence of failures. *ACM Trans. Comput. Syst.*, 5(1):47–76, 1987.

[5] Matt Blaze, Joan Feigenbaum y Angelos D. Keromytis. Keynote: Trust management for public-key. En *Infrastructures (Position Paper). Lecture Notes in Computer Science 1550*, págs. 59–63, 1999.

[6] Matt Blaze, Joan Feigenbaum y Jack Lacy. Decentralized trust management. En *In Proceedings of the 1996 IEEE Symposium on Security and Privacy*, págs. 164–173. IEEE Computer Society Press, 1996.

[7] J. Carbo, J. Molina y J. Davila. Trust management through fuzzy reputation. *International Journal in Cooperative Information Systems*, págs. 135–155, 2003.

[8] George Casella y Roger L. Berger. *Statistical Inference*. Duxbury press, Thomson Learning, 2002.

[9] B.A. Coan y G. Thomas. Agreeing on a leader in real-time [fault tolerant computer system]. *Real-Time Systems Symposium, 1990. Proceedings., 11th*, págs. 166–172, 1990.

[10] F. Cristian. Reaching agreement on processor group membership in synchronous distributed systems. *Distributed Computing*, 4(4):175–188, 1991.

[11] Tyrone Grandison. Specifying and analysing trust for internet applications. En *In Towards The Knowledge Society: eCommerce, eBusiness, and eGovernment, The Second IFIP Conference on E-Commerce, E-Business, E-Government (I3E 2002), IFIP Conference Proceedings*, págs. 145–157. Kluwer, 2002.

[12] Yang hua Chu, Joan Feigenbaum, Brian Lamacchia, Paul Resnick y Martin Strauss. Abstract referee: Trust management for web applications, 1997.

[13] Jean-Pierre Hubaux, Levente Buttyán y Srdan Capkun. The quest for security in mobile ad hoc networks. En *MobiHoc '01: Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, págs. 146–155, New York, NY, USA, 2001. ACM.

[14] F. Jahanian, A. Fakhouri y R. Rajkumar. Processor group membership protocols: Specification, design and implementation. En *Proc. 22th Symp. Reliable Distributed Systems*, págs. 2–11, 1993.

[15] Audun Jøsang. The right type of trust for distributed systems. Informe técnico, The Norwegian University of Science and Technology, 1996.

[16] Jiejun Kong, Haiyun Luo, Kaixin Xu, Daniel Lihui Gu, Mario Gerla y Songwu Lu. Adaptive security for multi-layer ad-hoc networks. En *Special Issue of Wireless Communications and Mobile Computing*, págs. 533–547. Wiley Interscience Press, 2002.

[17] Jiejun Kong, Petros Zerfos, Haiyun Luo, Songwu Lu y Lixia Zhang. Providing robust and ubiquitous security support for mobile ad-hoc networks. En *ICNP*, págs. 251–260, 2001.

[18] H. Kopetz, G. Grünsteidl y J. Reisinger. *Fault-tolerant membership service in a synchronous distributed real-time service*. Dependable Computing For Critical Applications. Springer-verlag edición, 1991.

[19] Haiyun Luo, P. Zerfos, Jiejun Kong, Songwu Lu y Lixia Zhang. Self-securing ad hoc wireless networks. *Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on*, págs. 567–574, 2002.

[20] S. Mishra, L.L. Peterson y R.D. Schlicting. A membership protocol based on partial order. *Dependable Computing for Critical Applications*, 2:309–331, 1992.

[21] L.E. Moser, P.M. Melliar-Smith y V. Agrawala. Membership algorithms for asynchronous distributed systems. En *Proc. 21th International Conf. Distributed Computing Systems*, págs. 480–488, 1991.

[22] L. Mui, M. Mohtashemi y A. Halberstadt. A computational model of trust and reputation. *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, págs. 2431–2439, Jan. 2002.

[23] S. Ramchurn, C. Sierra, L. Godo y N.R. Jennings. Devising a trust model for multi-agent interactions using confidence and reputation. *International Journal of Applied Artificial Intelligence*, 2004.

[24] Michael K. Reiter. A secure group membership protocol. *IEEE Trans. Softw. Eng.*, 22(1):31–42, 1996.

[25] A.M. Ricciardi y K.P. Birman. Using process groups to implement failure detection in asynchronous environments. En *Proc. 10th ACM Symp. Principles of Distributed Computing*, 1991.

[26] J. Sabater. *Trust and reputation for agent societies*. Tesis doctoral, Universitat Autonoma de Barcelona, 2003.

[27] Kimaya Sanzgiri, Bridget Dahill, Brian Levine, Clay Shields y Elizabeth Belding-royer. A secure routing protocol for ad-hoc networks. Informe técnico, Electrical Engineering and Computer Science, University of Michigan,UM-CS-2001-037, 2001.

[28] M. Schillo, P. Funk y M. Rovatsos. Who can you trust: Dealing with deception. págs. 81–94, Seattle, USA, May 1999. Autonomous Agents '99 Workshop on "Deception, Fraud, and Trust in Agent Societies.

[29] Glenn Shafer. *A Mathematical Theory Of Evidence*. Princeton University Press, 1976.

[30] W.H. Winsborough, K.E. Seamons y V.E. Jones. Automated trust negotiation. *DARPA Information Survivability Conference and Exposition, 2000. DISCEX '00. Proceedings*, 1:88–102 vol.1, 2000.

[31] Manel Guerrero Zapata. Secure ad hoc on-demand distance vector routing. *SIGMOBILE Mob. Comput. Commun. Rev.*, 6(3):106–107, 2002.

[32] Lidong Zhou y Z.J. Haas. Securing ad hoc networks. *Network, IEEE*, 13(6):24–30, 1999.