UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

EL ACERTIJO DE LOS SABIOS.
RAZONANDO SOBRE CONOCIMIENTO EN SISTEMAS DISTRIBUIDOS

TESIS

QUE PARA OBTENER EL TITULO DE

MATEMATICO

PRESENTA:

ERNESTO MANUEL ESPINOSA ASUAR

DIRECTOR DE TESIS:
DR. SERGIO RAJSBAUM GORODEZKY

FACULTAD DE CIENCIAS
SECCION ESCOLAR

VNIVERSIDAD NACIONAL
AVFNOMA DE
MEXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comuniçamos a usted que hemos revisado el trabajo de Tesis:
El acertijo de los sabios. Razonando sobre conocimiento en sistemas distribuidos.

realizado por Ernesto Manuel Espinosa Asuar

con número de cuenta 9013045-6 , pasante de la carrera de matemáticas

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dr. Sergio Rajsbaum Gorodezki	
Propietario	Dra. Atocha Aliseda Llera	Atocha Aliseda.
Propietario	M. en C. José Alfredo Amor Montaño	
Suplente	Dr. Raymundo Morado Estrada	R. Raymundo Morado E.
Suplente	Mat. Carlos Velarde Velázquez	Carlos Velarde

Consejo Departamental de Matemáticas
Mat. César Guevara Bravo

Agradecimientos

A mi director de tesis, Sergio Rajsbaum, gracias por todo el apoyo y por la paciencia; porque, cuando era posible, siempre había disponibilidad y atención; por todos los comentarios y todas las discusiones, que fueron indicando el rumbo de este trabajo.

A los sinodales: Atocha Aliseda, José Alfredo Amor, Raymundo Morado y Carlos Velarde. Gracias a todos por apoyarme en estos momentos de prisa. Los comentarios que me hicieron enriquecieron la calidad de la tesis. Agradezco a Raymundo Morado la sugerencia del nombre de los colores inevitables.

A los del cubículo 203: Efrén, Diana, Maribel, Manuel, Eduardo, Gerardo, Armando y Thomas. Por compartir el espacio; por la convivencia; por los juegos de pizarrón.

A mis cuates: Aarón, Guille, Pável, Paty, Charles, Zirán, Juan Gabriel, Ianna, Genaro, Carlos Oliva, Francisco Aguayo.

A Karim y a Miguel. Por todas las confrontaciones; por la confianza; por la búsqueda del cuarto. A Miguel por ser el cunca.

A Manolo, Carmen, Nacho, Javier, Rodolfo, Irma, Fernando, Bárbara, Ana, José Manuel, Andrea, Javier hijo, Rodolfo hijo, Mariana. Por el ejemplo; por la honestidad; por el cariño de ser familia.

A Alejandro. Porque no sólo formalmente eres parte de la familia, hemos creado estrechos lazos afectivos y de amistad. Por las idas y venidas al fut en el Ajusco.

A mis abuelos. A mi abuelo Félix por su energía y vitalidad. A Quique y a Santiago desde la distancia. A la familia Espinosa por todo lo festivo.

A mis padres. A ambos por criarme; por enseñarme; por estar siempre presentes. Y además, en los últimos meses, por soportarme. A mi padre, gracias por todas las cosquillas, las físicas y las emocionales. A mi madre, gracias por ser tan mamá.

A Laura. Porque compartimos mucho más que el ser hermanos. Por cuidarme. Por tantas pláticas esclarecedoras. Porque siempre estaremos ahí para el otro, ya lo sabemos y nos gusta repetirlo.

A dania. Que coqueta ella me miraba. Por tu luz, azul de mar, verde de selva. Por la claridad al caminar a tu lado.

Una de tantas veces Achcauhtli [2], *Bataboob* [3] *y Cuauhcóatl* [4] *son llamados al palacio real debido a su reputación de enorme sabiduría e intachable honestidad, para participar en uno de los famosos experimentos intelectuales del rey Nezahualcóyotl. Terminado el protocolo usual de bienvenida, el rey les enseña una caja en la que hay tres tocados de plumas rojas y dos tocados de plumas blancas y los invita a sentarse a su mesa redonda. Pide a los tres sabios que cierren los ojos, toma tres tocados de la caja y pone uno a cada uno. Acto seguido esconde los dos tocados que sobran. El rey avisa a los tres sabios que pueden abrir los ojos y pregunta a Achcauhtli: "¿Sabes de qué color son las plumas del tocado que te he puesto?" Achcauhtli observa los tocados de los otros dos, piensa un momento y responde al rey: "No, su majestad, no lo sé". Entonces Nezahualcóyotl pregunta a Bataboob: "¿Sabes tú de qué color son las plumas de tu tocado?" Bataboob mira rápidamente a Achcauhtli, luego mira a Cuauhcóatl y responde: "No, su majestad, tampoco yo sé de qué color son". Por último Nezahualcóyotl, agudizando la mirada, pregunta a Cuauhcóatl: "¿Sabes el color de las plumas de tu tocado?" Cuauhcóatl piensa un instante y afirma: "Sí mi soberano, ya sé de que color son las plumas del tocado que llevo puesto". ¿De qué color son las plumas del tocado de Cuauhcóatl?*

---

[2](Del náhuatl *achtli* primero y *cáhutl* tiempo) "primero en tiempo", decano, el más anciano. Sacerdote principal entre los antiguos mexicanos

[3]Los *bataboob* eran los jefes de los pueblos y aldeas mayas nombrados por el *halach uinic*, el rey o emperador. Presidían los consejos locales y en su carácter de juez sentenciaban a los criminales y resolvían las causas civiles

[4](Del náhuatl *cuauhtli* águila y *coátl* serpiente). Cuenta la leyenda que los antiguos mexicanos en su larga travesía desde Aztlán en busca de un nuevo lugar donde asentarse, llegaron a un lago en medio del cual había una isleta; tres sacerdotes fueron a examinar el lugar, entre ellos iba Cuauhcóatl. Encontraron un tunal en cuyo vértice estaba parada un águila y al pie, sobre el agua, estaba el nido de la majestuosa ave, fabricado de diferentes y hermosas plumas de pájaro. Cuahucóatl se sumergió en el carrizal donde se hallaba el tunal y se hundió desapareciendo completamente. Todos pensaron que se había ahogado; sin embargo regresó al día siguiente y contó que se le había aparecido el dios Tláloc y le había dicho: "ha llegado mi hijo querido Huitzilopochtli y este lugar será su asiento y domicilio, el será el protector de vuestra vida en la tierra". En ese lugar, donde estaba el águila sobre un tunal, se fundó Tenochtitlán

Índice

Capítulo 1

Introducción

Un problema fundamental en sistemas distribuidos es la complejidad de diseñar, analizar y entender las distintas partes que interactúan entre sí. Una herramienta muy útil que ha incrementado su importancia en el proceso de diseño y análisis es el uso del concepto de **conocimiento**. Aunque, como veremos más adelante, esto no es en un sentido amplio o filosófico. Se ha argumentado que una manera de razonar acerca de protocolos distribuidos es en términos de cómo cambia el conocimiento que tienen los procesadores sobre el sistema.

Aunque las tareas que un sistema distribuido realiza normalmente se piensan en términos del comportamiento global del sistema, las acciones que lleva a cabo cada uno de los procesadores dependen únicamente de su información local. Por ejemplo, si queremos enfocarnos en lo relevante a la comunicación, podemos pensar en un *sistema de mensajes* en el que las acciones de cada proceso son únicamente *mandar* o *recibir* un mensaje y posiblemente un conjunto de acciones locales (como cambiar el valor de una variable). Regularmente, al analizar un protocolo, se piensa en el *estado del conocimiento* que un proceso tiene sobre el sistema en determinado momento de una ejecución. Existe una relación muy cercana entre el conocimiento y las acciones que se realizan dentro de un sistema; intuitivamente las acciones de un proceso dependen de su conocimiento y su conocimiento cambia a raíz de sus acciones. En los sistemas actuales, con muchos procesadores y gran cantidad de información, es complicado darse cuenta de esta relación; por ello la importancia de formalizar la descripción del conocimiento.

Razonar acerca del conocimiento en sistemas distribuidos es parte medular de los argumentos intuitivos usados para el diseño de protocolos. Muchas veces oímos expresiones como "Una vez que el proceso ha recibido el mensaje de reconocimiento, *sabe* que el paquete que ha estado enviando ha llegado a su destino, puede desechar este paquete y leer el siguiente para enviarlo" o "el procesador i *sabe* que el procesador j *sabe* que el mensaje m fue enviado antes que el mensaje n, entonces ..." o "el procesador 1 no sabe si el procesador 2 está fallando porque considera posible que ...". En los últimos años ha habido artículos en los que se ha formalizado estos argumentos usando el trabajo de filósofos de finales de los cincuentas y principios de los sesentas. En 1962 J. Hintikka publicó el primer libro que se ocupó enteramente de un análisis lógico para razonar sobre conocimiento (ver [48]) La formalización de la noción de conocimiento en sistemas distribuidos ha permitido entender mejor qué es lo que ocurre dentro de un sistema y en algunos casos se ha facilitado la tarea de diseño y verificación de un protocolo.

Es importante notar que pensamos en el conocimiento como una noción "externa", asignada a los agentes por alguien que está analizando el sistema. No imaginemos a un procesador rascándose la cabeza tratando de averigüar si sabe o no algún hecho φ (ver [36]). Un programador o un diseñador de algún protocolo puede decir, desde afuera, que un procesador sabe un hecho φ porque en todos los estados globales consistentes con su estado actual φ es verdadero y entonces debe realizar determinada acción. Esta noción de conocimiento está basada en la información; no toma en cuenta, por ejemplo, las dificultades que envuelve el intentar calcular el conocimiento. Un procesador no debe necesariamente responder preguntas basadas en su conocimiento. La importancia de esta definición de conocimiento es que se captura la manera informal en la que generalmente pensamos sobre protocolos y programas; un diseñador de un sistema puede pensar que "no se va a detener después de tres rondas porque el procesador 1 podría no saber que el procesador 2 sabe que el procesador 3 está fallando." Este uso informal de la palabra "sabe" es capturado por la definición de conocimiento que estamos usando. En términos prácticos esta noción de conocimiento ha resultado ser útil, dando importantes pistas para el diseño y verificación de protocolos distribuidos.

Se puede dar una semántica a la lógica del conocimiento usando el modelo de los *mundos posibles*. La idea es que además del estado real de una situación, un agente[1] considera que existen otros mundos, o estados de la situación, posibles. Se dice que un agente *sabe* un hecho representado por una fórmula φ si φ es verdadero en todos los mundos que él considera posibles. Expresamos esto utilizando a los operadores modales K_i, donde la fórmula $K_i\varphi$ se lee "el agente i sabe φ". Por ejemplo, en un juego de dominó, estos mundos posibles tienen una interpretación concreta: son todas las reparticiones de las 28 fichas entre los cuatro jugadores, tocándole siete fichas a cada uno; durante el transcurso de la partida cada jugador va adquiriendo más información; esta información permite ir eliminando algunos de los mundos que consideraba posibles. Por ejemplo, Alicia y Beto forman una pareja en una partida; en algún momento si Alicia sabe que Beto tiene la mula de doses entonces en todos los mundos que considera posibles, Beto tiene esta ficha. Y si Alicia pasa a cuatros, entonces Beto sabe que Alicia no tiene cuatros; esto se representa con el hecho de que Beto eliminará todos los mundos que consideraba posibles en los que Alicia tenía algún cuatro. Intuitivamente, mientras más mundos considere posibles un agente, mayor será su incertidumbre y menor será su conocimiento. Esto es, estamos viendo a la posibilidad como un dual de conocimiento.

El modelo de los mundos posibles comúnmente es formalizado usando técnicas desarrolladas inicialmente por S. Kripke en 1963. Kripke definió a las *estructuras Kripke* pero sólo asumía que había un agente. Sin embargo es posible extender este modelo y obtener estructuras Kripke para n agentes. Se asume que los n agentes viven en un mundo que puede ser descrito en términos de un conjunto Φ no vacío de proposiciones primitivas. Una estructura Kripke es una tupla $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ en donde S es un conjunto de *estados* o *mundos posibles*, π es una interpretación sobre los estados y las proposiciones primitivas y cada $\mathcal{K}_i$ es una relación de posibilidad entre los estados de S. Intuitivamente para dos estados $s, t \in S$ si $(s, t) \in \mathcal{K}_i$; esto quiere decir que estando en el estado s el agente i considera posible que el estado t sea la situación real.

[1] Un agente puede ser un jugador en una partida de dominó, un robot en una línea de ensamblaje de autopartes o un procesador en un sistema distribuido.

La interpretación que usaremos del modelo de los mundos posibles en sistemas distribuidos captura la interacción entre conocimiento y acción; esta interpretación y algunas variantes han aparecido constantemente en la literatura; aquí usaremos básicamente el modelo que apareció por vez primera en 1984 en una versión preliminar de [41] presentada en el *Third ACM Symposium on Principles of Distributed Computing*. En los dos años siguientes fueron presentados algunos artículos en congresos que utilizaban el mismo modelo (ver [38, 73, 74]); en 1986 se organizó la *First Conference on Theoretical Aspects of Reasonong About Knowledge* en la que se siguió discutiendo la noción de conocimiento en sistemas distribuidos (ver [26, 25, 57, 64]) y a partir de entonces se ha organizado este congreso cada dos años, ahora llamado *Conference on Theoretical Aspects of Rationality and Knowledge*. La versión más completa del modelo apareció en 1995 en el libro *Reasoning About Knowledge* de R. Fagin, J. Halpern, Y. Moses y M. Vardi y en 1997 en [24], de los mismos autores, se hicieron algunas modificaciones. El lector interesado en conocer un panorama del estado del arte en razonamiento sobre conocimiento puede consultar [36].

El modelo que utilizamos ha sido motivado por trabajos en análisis de protocolos basado en conocimiento (ver por ejemplo [9, 14, 20, 26, 38, 41, 45, 68, 70, 74]). Consideremos a un sistema distribuido como un conjunto de n procesadores o procesos conectados por medio de una red de comunicación. Cada procesador está en un *estado local*, en cada momento determinado; se asume que el estado local guarda toda la información a la que tiene acceso el procesador. Todo el sistema estará en un *estado global*, donde cada estado global es una $(n+1)$-ada que consiste de un estado local para cada procesador en el sistema y un estado correspondiente al *medio ambiente*. El medio ambiente es todo aquello relevante para el sistema que no está contenido en los estados de los procesadores. El medio ambiente variará dependiendo de lo que estemos modelando y de lo que queramos analizar.

Cada estado global describe al sistema en un determinado momento. Sin embargo, el análisis de un sistema consiste en estudiar cómo cambia con el transcurso del tiempo y cuáles son sus diferentes estados globales, así que necesitamos incorporar al tiempo en nuesto modelo. Identificaremos al sistema con un conjunto de *ejecuciones* donde una ejecución es una función de los números naturales a un conjunto de *estados globa-*

les. Intuitivamente una ejecución describe todos los eventos relevantes que ocurren en el sistema en un periodo de tiempo (para nuestros fines será suficiente si se toma el tiempo corriendo sobre los números naturales, aunque puede ampliarse este concepto a los números reales). Si r es una ejecución, entonces $r(t)$ es el estado global del sistema en la ejecución r al tiempo t. Un *punto* es una pareja (r,t) que consiste en una ejecución r y el tiempo t. Esta definición es útil porque si consideramos el estado global $r(t)$ asociado al punto (r,t), puede ocurrir que este estado global aparezca en otra ejecución r'; entonces $r(t) = r'(t')$; no podemos distinguir entre ambos estados globales; sin embargo los puntos (r,t) y (r',t') sí son distintos. Formalmente definiremos a un *sistema* como un conjunto de ejecuciones. Esto abstrae la idea de un sistema como una colección de procesadores que interactúan entre sí, lo que estamos modelando son los posibles comportamientos del sistema.

Para incorporar al conocimiento y al modelo de los mundos posibles en el modelo de un sistema distribuido, supondremos que tenemos un conjunto de proposiciones llamadas *proposiciones primitivas* que representarán la descripción de hechos básicos del sistema. Cada estado global induce una asignación de verdad $\pi(g)$ para las proposiciones primitivas, donde π es una *interpretación*, una función de los estados globales a asignaciones de verdad. Nosotros podemos decidir cómo es esta interpretación; sin embargo lo natural es que una proposición primitiva sea verdadera en un estado global si el hecho que representa ocurre en ese estado. Por ejemplo si p es "el valor de la variable x es 1" se espera que p sea verdadera sólo en los estados globales en los que x vale uno. Construiremos un lenguaje cerrando el conjunto de proposiciones bajo negación, conjunción y los operadores modales K_i.

Un *sistema interpretado* $\mathcal{I}$ es una pareja $(\mathcal{R},\pi)$ donde $\mathcal{R}$ es un sistema y π es una interpretación. Se asocia a un sistema interpretado con las estructuras Kripke de manera que los puntos de las ejecuciones del sistema son los mundos posibles y para cada procesador se define una relación entre los puntos; diremos que dos puntos del sistema $\mathcal{I}$, (r,t) y (r',t'), son *indistinguibles* para el procesador i si el estado local del procesador i es el mismo en ambos puntos. Las relaciones así definidas son relaciones de equivalencia y entonces obtenemos las llamadas estructuras Kripke $S5$.

También se ha pensado en modelos en los que sí interesa que los

agentes o procesadores hagan cálculos de su conocimiento. Hay que tomar en cuenta entonces qué tan complicado es hacer estos cálculos y de hecho el modelo semántico varía; para trabajos relacionados ver por ejemplo [18, 43, 52, 66].

Una manera de trabajar en el diseño de programas es yendo de arriba hacia abajo; primero se diseña un protocolo de alto nivel que no sea dependiente de las características del sistema y luego se hace la implementación pensando en las particularidades del entorno. Este estilo de programación en general nos permitirá hacer cambios cuando consideremos entornos diferentes.

Motivados por estas consideraciones se presentó la noción de *protocolos basados en conocimiento* (ver [38], [39] y ver también [70] en donde se presentó una versión simplificada). En estos protocolos las acciones de los procesadores dependían explícitamente de su conocimiento. El objetivo era obtener una semántica formal para programas con pruebas de conocimiento de la forma

$$\text{if } K(x = 0) \text{ do } y := y + 1$$

donde $K(x = 0)$ debe ser leído como "sabes que el valor de x es 0".

Sin embargo la definición de estos protocolos tenía deficiencias que la hacían difícil de manejar como herramienta para diseño de programas. Por un lado los protocolos basados en conocimiento estaban definidos como funciones de estados locales y sistemas a conjuntos de acciones. Entonces no se capturaba directamente la intuición de que los protocolos o programas tuvieran pruebas de conocimiento. Además no se hacía una distinción clara entre el protocolo y el entorno en el que se estaba ejecutando, perdiendo la idea de hacer un razonamiento de alto nivel independiente del modelo. De cualquier manera los protocolos basados en conocimiento fueron usados (formal o informalmente) en algunos artículos (ver [14, 45, 68]).

En el libro de Fagin, Halpern, Moses y Vardi ([21]) se presentó una nueva formalización que resuelve las deficiencias de la definición anterior. Se introdujo la noción de *programas basados en conocimiento*, que son lo que se quería que fueran los protocolos basados en conocimiento: programas sintácticos con pruebas de conocimiento. Se incluye también

la noción de *contexto* que captura el entorno en el que va a ser ejecutado el programa. Haciendo la distinción entre programas y contextos, y dando un significado a los programas en diferentes contextos, se logra hacer un razonamiento basado en conocimiento de alto nivel que sea independiente del modelo .

1.1 Los acertijos

En este trabajo presentamos el *acertijo de los tres sabios* y el *acertijo de las esposas infieles.* Estos acertijos han sido utilizados con frecuencia para mostrar el tipo de razonamiento que se está modelando y en general, para ilustrar nociones de conocimiento en sistemas distribuidos, (ver [7, 21, 22, 24, 32, 39, 41, 44, 62, 67, 76, 83]).

En el acertijo de las esposas infieles tenemos n matrimonios; en cada matrimonio la mujer es fiel o es infiel a su respectivo esposo. Todos los hombres casados están enterados de qué mujeres son infieles pero no saben si su propia esposa le es fiel. Generalmente se sitúa al acertijo en un país que vive bajo un régimen monárquico con un rey o sultán; se asume también que todos los hombres son perfectos razonadores, muy inteligentes y son completamente honestros. El rey anuncia un día que se sabe que hay infidelidades en el reino y que aquel que descubra que su esposa lo engaña debe tomar alguna acción esa misma noche; en general esta acción tiene que ver con algún castigo severo para la esposa. Se sabe que todos los hombres casados escuchan el anuncio y que todos van a obedecer el mandato. Se sabe también que todos se enterarán cuando alguien tome la acción. Pasan $k-1$ noches sin que nadie haga nada hasta que a la noche siguiente alguien realiza la acción. El acertijo está en descubrir cuántos esposos estaban siendo engañados. Es fácil ver por inducción que hay k esposas infieles y que todos los esposos engañados se dan cuenta a la k-ésima noche y entonces todas las esposas infieles son castigadas esa misma noche. Este acertijo también es conocido como *el acertijo de los esposos infieles* si se escoge que las mujeres sean más inteligentes que los hombres y entonces se piensa en un país bajo un matriarcado en el que los papeles se invierten.

Martin Gardner afirma (ver [30] y [31]) que el acertijo de las esposas infieles apareció por primera vez en 1958; esta primera versión habla

de 40 mujeres infieles en una ciudad gobernada por un sultán (ver [27]). En [30] Gardner presenta una versión de los esposos infieles; sitúa el acertijo en el planeta Womensa y asume que las mujeres del país son descendientes de mujeres alemanas de inteligencia extremadamente superior y mucho más inteligentes que los hombres. La reina del lugar ordena que aquella esposa que descubra que su esposo le es infiel debe avisar a las autoridades y como castigo el esposo será castrado ese mismo día. Una versión muy similar de este acertijo aparece en [67]; esta versión se sitúa en el país de Mamajorca y ahí el decreto de la reina es que las esposas engañadas que se han dado cuenta de la infidelidad de su marido deben matarlo con una escopeta a medianoche de ese mismo día. Recomendamos al lector interesado consultar este artículo en el que se hace una revisión bastante completa de distintas versiones del acertijo pensando en limitaciones posibles en la comunicación de mensajes.

Una versión isomorfa al acertijo de las esposas infieles es *el acertijo de los niños enlodados* (ver por ejemplo [21]). En esta versión tenemos a un conjunto de n niños, todos hermanos, que están jugando en el patio de su casa; después de un rato algunos de ellos se ensucian la cara, lo que supone un severo castigo por parte de la madre. Un niño no tienen forma de saber si está sucio pero como ve a todos sus hermanos sí sabe quienes se han ensuciado. Se sabe que ninguno de los niños va a avisarle a alguno de sus hermanos que está sucio. Llega el padre de los niños que quiere evitar que la madre regañe a los pequeños, pero no quiere señalar directamente a los niños sucios porque siente que los estaría sobreprotegiendo. Sin embargo, sabedor de la gran inteligencia de todos sus hijos, decide darles suficiente información para que los niños que se han enlodado lo descubran por si mismos. Así el padre anuncia a todos los niños que alguno de ellos está sucio y después pide que de un paso al frente aquel que ya sepa que está sucio. Cuando el padre ha hecho esta petición k veces todos los niños enlodados dan un paso al frente. El acertijo consiste en descubrir cuántos niños están sucios. La respuesta, similar a lo que ocurre en el acertijo de las esposas infieles, es que hay k niños enlodados. El acertijo de los niños enlodados es usado en [41] y en [21] como ejemplo para explicar qué es el conocimiento común. Puede consultarse una discusión más detallada del acertijo en [21] y en [44].

Una versión muy similar al acertijo de los sabios que presentamos, con tres sabios y tres sombreros rojos y dos sombreros blancos, apareció en 1985 (ver [69]), en un libro del que Martin Gardner dice que es "un libro delicioso para niños muy inteligentes, basado completamente en acertijos de inducción matemática acerca de sombreros de colores" (ver [31]). Martin Gardner a su vez utiliza el acertijo de los sabios como ejemplo de inducción matemática (ver [29] y [30]). En [29] Gardner generaliza el problema a n sabios y a una caja con n sombreros rojos y $n-1$ sombreros blancos. Gardner concluye que en esa situación da lo mismo si los sabios estan sentados en fila, cada uno viendo sólo a aquellos que no han respondido a la pregunta. Gardner discute otra situación para esta versión en la que los sabios están sentados en fila. Supone que el rey lleva $n-1$ sombreros negros y un sombrero rojo. La pregunta es si el sabio con sombrero rojo siempre va a saber el color de su sombrero. Resulta que va a saberlo sólo si está en una posición impar en la fila.

Hay otra versión del acertijo de los sabios en la que no se asume que los tres sabios sean igual de sabios; el rey lleva también tres sombreros rojos y dos sombreros blancos y pone un sombrero rojo a cada sabio; pasado un tiempo el sabio más sabio decubre el color de su sombrero.

Hay otros artículos que utilizan el acertijo de los tres sabios como ejemplo de sistemas formales de axiomas o de sistemas deductivos; presentamos un comentario sobre estos artículos en orden cronológico:

En [62] McCarthy da una descripción de un sistema formal; los axiomas estan escritos en lógica de primer orden y usan los mundos posibles al estilo de Kripke. Se utiliza este sistema formal para expresar el acertijo.

En [76] Stark presenta una formalización de la lógica del conocimiento en la que da una regla de deducción para la teoría de modelos; al final del artículo menciona que esta regla surgió a raíz de que McCarthy le comentó el problema de formalizar la respuesta de cada sabio en el acertijo.

En [1] Attardi y Simi presentaron otra solución para el acertijo utilizando el sistema *OMEGA*. Se representó el conocimiento de los sabios utilizando los *puntos de vista* (*viewpoints*) y se hacen las deducciones utilizando un conjunto de reglas que especifican las propiedades generales de estos puntos de vista.

En [51, 54] se estudia el acertijo desde la perspectiva del estado final del razonamiento. En estos artículos se utiliza el conocimiento común y la creencia común en la formalización del problema; se utiliza un metalenguaje para describir el razonamiento de los sabios. En el artículo de Konolige [51] se formaliza al acertijo con lógica modal proposicional. La formalización consiste en una teoría modal que describe al acertijo desde el punto de vista de un observador externo.

En [58] Lehman discute la formalización del conocimiento en sistemas distribuidos. Utilizando un lenguaje proposicional presenta un método para la descripción formal de un sistema. El lenguaje es interpretado en un modelo de estructuras Kripke. El acertijo es utilizado como ejemplo de la formalización.

En [6] Carlucci Aiello, Nardi y Schaerf dan una solución para el acertijo por medio de la lógica de primer orden, utilizando una arquitectura de meta nivel. El objetivo de ese artículo es mostrar que dicha arquitectura es idónea para resolver problemas de representación de conocimiento que requieran de razonamiento sobre el conocimiento. Los mismos autores en [7] dan una implementación del acertijo en un sistema llamado *FOL*.

En [12] Coscia, Franchesci, Levi, Sardu y Torre proponen al acertijo como ejemplo de las herramientas para manejar las bases de conocimiento descritas en el artículo. Utilizan un programa escrito en *PROLOG* extendido.

En [53] Konolige discute el problema de asignar creencias a un agente. Desarrolla un modelo de creencia llamado *modelo derivacional*. Formaliza el modelo y su teoría deductiva y los utiliza para resolver una versión del acertijo para dos sabios.

En [16] Elgot-Drapkin afirma que en las formalizaciones que se han dado del acertijo de los tres sabios no se ha tomado en cuenta al tiempo. Para modelar que el tiempo pasa mientras los sabios hacen sus razonamientos utiliza un formalismo llamado *lógica de pasos* (*step-logic*). La misma autora en [15] presenta una solución para una versión del acertijo con sólo dos sabios, basada en la lógica de pasos.

En [32] Gmytrasiewicz y Durfee presentan un modelo basado en las estructuras Kripke. Anidando recursivamente se define lo que son las actitudes proposicionales de un agente para distinguir los conceptos de *saber* y *creer*. El acertijo de los tres sabios se utiliza como ejemplo para

mostrar que el modelo que se presenta es útil para hacer razonamiento decuctivo.

En [8] Cimatti y Serafini discuten el uso de contextos de creencia para la formalización de razonamiento de multi-agentes. Discuten una solución del acertijo que utiliza estos contextos.

1.2 Nuestras contribuciones

Como ya mencionamos, los acertijos han sido utilizados como ejemplos para ilustrar nociones como la inducción, ilustrar el tipo de razonamiento que se quiere capturar con el modelo de los mundos posibles o como ejemplo de sistemas formales de axiomas o de sistemas deductivos. Sin embargo no sabemos de algún trabajo en el que se haya intentado estudiar lo que ocurre en versiones más generales de los acertijos. Por ejemplo podemos pensar que hay n sabios y que el rey lleva una caja con sombreros de varios colores.

Ambos acertijos son versiones de un acertijo general al que llamamos el *acertijo de los sabios*. En este acertijo un rey tiene una caja con r colores y un número de etiquetas de cada color. El rey va a asignar un color a cada sabio; entonces el rey necesita al menos n etiquetas. Supondremos que en la caja habrá en total $n+k$ etiquetas, donde $k \geq 0$. El rey se reúne con los n sabios del reino, les enseña la caja y asigna un color a cada uno; después el rey va preguntando a algún conjunto de sabios, que él escoge cada vez, si saben su color. El objetivo del rey es que los sabios no puedan adivinar este color.

Los colores representan los tocados, sombreros o etiquetas que se pone a los sabios en las versiones que hemos presentado. Por ejemplo, el acertijo de las esposas infieles y el acertijo de los niños enlodados se pueden ver como un acertijo en el que el rey lleva una caja $C = (n, n-1)$ y en el que pregunta simultáneamente a todos los sabios. Después de que el rey les enseña la caja, los n sabios saben que al menos a uno de ellos le tocará color negro ("al menos uno de ellos está siendo engañado" o "al menos uno de ellos está sucio").

En este trabajo analizamos dos versiones de este acertijo general, una versión en la que el rey va preguntando uno por uno a los sabios, haciendo en total n preguntas y otra versión en la que el rey cada vez le

pregunta a todos los sabios. Para ambos acertijos decimos qué es lo que ocurre. Respondemos a preguntas como cuándo un sabio sabe su color, si el número de etiquetas de cada color tiene que ver con las respuestas de los sabios, si para alguna caja el rey logra que ningún sabio sepa su color. Todos estos resultados están en el capítulo 4 y hasta donde sabemos son originales.

Se ha presentado el acertijo de los esposos infieles y el de los niños enlodados como ejemplos de un protocolo o un programa basado en conocimiento (ver [21, 24, 39, 83]); en dichos protocolos y programas, cada agente decide las acciones que tomará en función de su conocimiento del sistema. Siguiendo con estas ideas, pensaremos que el acertijo se desarrolla en un sistema distribuido. Usaremos el modelo que relaciona sistemas interpretados con estructuras Kripke para analizar desde el punto de vista del conocimiento lo que ocurre en las dos versiones del acertijo que nos interesan. Pensaremos en un sistema con $n + 1$ agentes o procesadores, uno para cada sabio y otro que representará al rey. Definiremos el estado local de cada agente. Las acciones que podrán tomar los agentes será enviar mensajes. El rey enviará mensajes que representen sus preguntas a los sabios y los sabios enviarán mensajes que representen sus respuestas. Definiremos un programa basado en conocimiento para el rey y para los sabios.

El programa del rey definirá un protocolo en el que el rey decidirá sus acciones sólo en función de su estado local. En cada ronda en la que le toque preguntar, el rey decidirá a qué subconjunto de los sabios les va a preguntar y entonces su acción será mandar un mensaje de pregunta a todos los sabios en este subconjunto. Sin embargo los programas basados en conocimiento de los sabios, que esencialmente es el mismo para todos los sabios porque todos se comportan igual, definen protocolos basados en conocimiento en los que un sabio decide sus acciones, es decir, decide si sabe o no su color, basándose en su estado local y en los mundos que considera posibles. El problema ahí es que estos mundos son todas las posibles ejecuciones del acertijo. Es decir que un sabio no puede determinar sus acciones sólo viendo su estado local y esto es lo que deseamos que pase en un protocolo. Esto no va a representar un problema porque los protocolos basados en conocimiento nos van a servir para que nosotros, como observadores externos, hagamos un análisis de lo que ocurre en una ejecución del

protocolo y entonces podremos determinar un protocolo para cada sabio en el que sus acciones sólo dependan de su estado local.

El objetivo será obtener, por medio del análisis del conocimiento en el sistema, un protocolo para cada sabio en el que no intervenga la noción de conocimiento sino que el sabio haga algún tipo de operación sobre su estado locales. Este protocolo implementará al programa basado en conocimiento, es decir, determinará las mismas acciones en cada ronda.

1.3 El acertijo de los sabios

Al inicio de este trabajo narramos el acertijo de los tres sabios; si el lector no se acuerda del acertijo lo invitamos a la primera página para que lo vuelva a leer y para que procure resolverlo; si ya se dio por vencido lo alentamos para que lo vuelva a intentar antes de continuar.

Veamos qué es lo que ocurre: los tres sabios razonan perfectamente y siempre responden a las preguntas del rey con toda honestidad. Para facilitar la lectura en vez de los sabios Achcauhtli, Bataboob y Cuahucóatl hablaremos de los sabios A, B y C; además asumiremos que el rey asigna un color a cada sabio, blanco o rojo. Bajo estos supuestos, se hace el siguiente razonamiento: el sabio A habría sabido su color sólo si los otros dos sabios tuvieran color blanco. Como A respondió "no", entonces alguno de los otros dos tiene color rojo. Si el sabio C tuviera blanco entonces B habría sabido que su color es rojo, pero B no supo su color. Por lo tanto el color de C es rojo.

Es interesante observar que B sólo se fijó en el color del sabio C; su gran inteligencia le permitió deducir que no necesitaba ver el color de A para responder, le fue suficiente con oir su respuesta. C ni siquiera miró a los otros dos, al oir sus respuestas respondió al rey inmediatamente. Habría sido lo mismo si los sabios hubiesen estado sentados de manera tal que cada sabio viese solamente a los que no han respondido; en particular, aún si C fuera ciego, habría sabido su color.

1.3.1 Yo sé que tu sabes: conocimiento común

Supongamos un experimento en el que el rey asigna color rojo a los tres sabios. Al preguntarle a A éste responde "no". Como ya dijimos, A sólo respondería sí cuando los otros dos sabios tuvieran color blanco. Cuando A responde "no" a la pregunta del rey, la información que se gana es que al menos un sabio, B o C, tiene rojo. B considera posible que su color sea rojo o que sea blanco, ya que está viendo que C tiene rojo, entonces B responde "no" al rey. Entonces C sabe que su color es rojo y responde "sí". Sea p la proposición "Entre B y C al menos uno tiene rojo". Martin Gardner plantea en [29] una pregunta interesante para este caso: ¿Por qué es importante que el rey le pregunte a A? B y C saben que no es posible que A sepa el color de su sombrero; esto

es, los dos saben p porque están viendo que el otro tiene color rojo. Si B y C ya tienen la información que da la respuesta negativa de A ¿de que les sirve oirla? La proposición p no puede ser la única información que da oir el "no" de A, ya que si el rey no le pregunta a éste y decide preguntar primero a B, C no podría saber su color: como B no sabe que el rey decidió no preguntarle a A, respondería "sí" sólo si los otros dos sabios tuvieran blanco; B responde "no" por lo tanto; para C, que tampoco sabe que el rey decidió no preguntarle a A, esto quiere decir que entre A y él, al menos uno tiene rojo; C ve que el color de A es rojo, para él es posible entonces que su color sea rojo o que sea blanco. No sabe, por lo tanto, el color de su sombrero. Hay una diferencia importante entre las dos situaciones; antes de que el rey haga alguna pregunta, B ve el color rojo de C, sabe que A contestará "no" a la pregunta del rey, por lo tanto sabe la proposición p, pero no sabe si C sabe p: B piensa que es posible que él mismo tenga blanco y entonces C no sabría p. En el caso de C es lo mismo, C sabe p, pero no sabe si B sabe p. La importancia de que el rey le pregunte a A es que depués de que éste responde "no", B ya sabe que C sabe p, y C sabe que B sabe p. Aún más, B sabe que C sabe que B sabe p y C sabe que B sabe que C sabe p, etc. A esto se le llama **conocimiento común** entre B y C de p [2]. Antes de que el rey le pregunte a A, B y C no tienen conocimiento común de p; la respuesta negativa de A agrega este conocimiento común. Así, cuando el rey le pregunta a C, C sabe que B sabe p. Si el rey no le pregunta a A, para C es posible que B no sepa p. En general un grupo tiene conocimiento común de una proposición p, si todos saben p, todos saben que todos saben p, todos saben que todos saben que todos saben p, etc.

1.3.2 La reunión de los sabios

Algún tiempo después, varios sabios del reino se reúnen en el patio mayor de los palacios, junto a los archivos reales. El rey Nezahualcóyotl, todavía algo molesto por la facilidad con la que Achcauhtli, Bataboob y Cuauhcóatl resolvieron el acertijo, se percata de la presencia de

[2] El conocimiento común es mencionado frecuentemente en la literatura, ver por ejemplo [59, 2, 63, 10, 14, 41, 71, 22]

los sabios y decide retarlos con un problema más complicado. Así que manda a un mensajero a proponer su juego: el rey llevará en una caja muchos tocados de plumas de colores, enseñará la caja a los sabios y pondrá un tocado a cada uno; luego esconderá los tocados que le sobren. Los sabios podrán ver el tocado de todos los demás pero no el suyo propio y el rey irá preguntando uno por uno si saben de qué color es el tocado que llevan puesto.

Los sabios, conociendo bien el carácter del rey, saben que intentará poner los tocados de manera que ninguno logre saber el color del que lleva puesto, por lo que antes de que llegue el emperador se juntan a estudiar el problema. Después de un par de horas de deliberaciones dan al mensajero la siguiente respuesta: "Su majestad, nos honrará participar en tan interesante experimento y le prometemos que alguno responderá afirmativamente, siempre y cuando la caja de tocados permita al menos una colocación inicial sobre nuestras humildes cabezas, en la cual el primero en ser interrogado tenga suficiente información para responder afirmativamente."

¿Cómo llegan los sabios a esta conclusión? En otras palabras, los sabios se dan cuenta de que existen solamente dos tipos de cajas de tocados (de acuerdo a cuántos hay y de qué color son). En el tipo *bueno,* no importa cuál sea la colocación inicial de tocados que el rey escoja, ni el orden en el que les pregunte, siempre existe un sabio que llega a saber el color de su tocado. En el tipo *malo,* para cualquier colocación inicial, ningún sabio llega a saber el color de su tocado.

Los resultados que hemos platicado en esta sección serán formalizados y demostrados en el capítulo 4.

1.3.3 Las esposas infieles

Nezahualcóyotl se caracterizó por poner a su reino en grandísimo orden; lo dividió en ocho mayordomías y para el buen gobierno estableció ochenta leyes; una de ellas mandaba en contra del adulterio: "Al adúltero, si le cogía el marido de la mujer en el adulterio con ella, morían ambos apedreados; si era por indicios o sospechas del marido, y se venía a averiguar la verdad del caso, morían ambos ahorcados, y después los

arrastraban hasta un templo fuera de la ciudad" [3].

Había un pueblo que se caracterizaba por la inteligencia de sus habitantes pero también por un alto número de casos de adulterio a pesar de las terribles sanciones. Los habitantes del pueblo consideraban que el adulterio era parte de sus costumbres y nunca denunciaban a nadie. Era un hecho conocido que todos eran bastante inteligentes así que se pensaba que un marido al que estaban engañando debía ser capaz de descubrirlo por sí mismo. También se sabía que todos los hombres casados sabían bien cuáles mujeres eran fieles y cuáles no, exceptuando, claro está, su propia mujer. El pueblo era pequeño y los chismes corrían rapidamente de boca en boca; si es que alguien no se enteraba de que habían agarrado a alguna esposa infiel se enteraba de todos modos cuando las ejecuciones se llevaban a cabo porque cuando arrastraban a los ejecutados hasta el templo la mayoría del pueblo asistía a esta procesión fúnebre y era imposible no enterarse por todo el revuelo que se armaba.

El mayordomo responsable de la región donde estaba el pueblo se encontraba muy preocupado porque Nezahualcóyotl había exigido que sus ordenanzas se cumplieran con todo rigor. Había intentado muchas cosas pero los del pueblo no entraban en razón. Finalmente un día se le ocurrió una solución. El mayordomo juntó a todos los hombres casados en la plaza central y les hizo un anuncio: "Sigue habiendo infidelidades en el pueblo; les ordeno que si alguno de ustedes se entera de que su esposa lo está engañando, debe denunciarlo inmediatamente y esa misma noche se llevará a cabo la ejecución que manda las ordenanzas". Treinta y nueve noches pasaron en las que nada ocurrió; a la noche siguiente, hubo ahorcados en el pueblo. ¿Cuántos maridos estaban siendo engañados?

No es muy complicado demostrar por inducción que si hay n maridos engañados, entonces descubren su situación después de $n-1$ noches sin ejecuciones, a la n-ésima noche todas las mujeres infieles son ahorcadas junto con sus amantes. En el caso de $n = 1$, el único marido engañado no sabe de mujeres infieles; como el mayordomo ha anunciado que hay al menos una mujer infiel, concluye que su propia esposa lo está engañando y la denuncia en la primera noche. Si hay dos maridos engañados cada

[3] Martínez, J.L. *Nezahualcóyotl*, Colección Lecturas Mexicanas 39, Fondo de Cultura Económica, México, 1984, p. 249

uno sabe del otro, en la primera noche no hacen nada pensando que la esposa del otro es la única infiel, pero entonces después de la primera noche cada uno piensa que si sólo hubiera una mujer infiel la habrían ejecutado en la primera noche y concluyen que tiene que haber otro marido engañado y ese marido debe ser él mismo, así ambos esposos engañados denuncian a sus esposas a la segunda noche. Continuando con este razonamiento concluiremos que en el pueblo había cuarenta maridos que estaban siendo engañados.

Ya vimos que el acertijo de las esposas infieles se representa en la versión general para el acertijo de los sabios suponiendo que el rey lleva una caja $C = (n, n-1)$. El rey asigna los colores y pregunta a todos los sabios "¿Alguno de ustedes sabe su color?". El rey repite esta pregunta k veces hasta que los k sabios con color 1 simultáneamente se dan cuenta de que ellos deben tener color 1 y simultáneamente responden "sí" al rey. Si pensamos en esta misma situación, pero volviendo a cuando el rey pregunta uno por uno, el primero que va a saber su color es el último al que le pregunte con color 1; esto es, si hay k sabios con color 1, a la k-ésima vez que el rey pregunte a un sabio con color 1, va a escucharse un "sí". Estos resultados los formalizaremos en el capítulo 4. Ahí analizamos el acertijo de los sabios para la versión de los tres sabios y la de las esposas infieles. Obtenemos resultados sobre la caja de colores que lleva el rey y sobre lo que ocurre en los acertijos dependiendo de la caja y de cómo asigne el rey los colores.

1.4 Organización de la tesis

En el capítulo 2 presentamos el modelo teórico de este trabajo; en la sección 2.1 hablamos sobre propiedades y características del modelo semántico utilizado para modelar el conocimiento en sistemas distribuidos y presentamos el modelo general en la sección 2.2. En el capítulo 3 hablamos sobre el modelo formal del acertijo de los sabios y presentamos los programas basados en conocimiento para el sistema del acertijo. En el capítulo 4 presentamos nuestros resultados sobre lo que ocurre en los acertijos y presentamos los protocolos que implementan a los programas basados en conocimiento.

Capítulo 2

Modelo Teórico

2.1 Lógica del conocimiento

Presentamos un modelo semántico para un sistema distribuido que se basa en el modelo de los mundos posibles. El objetivo es analizar el sistema por medio del conocimiento. Para formalizar este tipo de razonamiento necesitaremos un lenguaje. Describiremos la sintaxis del lenguaje y más adelante hablaremos sobre la semántica de nuestro modelo.

A la lógica proposicional se le añade los conectivos $\Box$ y $\Diamond$ de manera que si φ es una fórmula proposicional entonces $\Box\varphi$ y $\Diamond\varphi$ son fórmulas en la lógica proposicional modal. La semántica de la lógica modal se basa en la semántica de los mundos posibles, una noción que tiene sus origenes en Leibniz. La idea esencial es que existe un conjunto de mundos posibles que representan los posibles estados de la realidad y una relación entre los mundos.

En 1963 Kripke introdujo en [56] a las *estructuras Kripke* como un modelo formal de la semántica de los mundos posibles. Se asume que se tiene un conjunto de proposiciones atómicas o primitivas, que representan hechos básicos sobre los que queremos hacer razonamientos. Una estructura Kripke M es igual a la triada (W, R, V), donde W es un conjunto no vacío de mundos posibles, R es una relación binaria entre los mundos, llamada relación de accesibilidad y V es una función de valuación que asigna un valor de verdad a cada proposición atómica

en cada mundo posible. $\Box\varphi$ es verdadera en el mundo $s \in W$ de la estructrura M si φ es verdadera en todos los mundos s' tales que $(s, s') \in R$. $\Diamond\varphi$ es verdadera en s si φ es verdadera en algún mundo s' tal que $(s, s') \in R$. Es fácil verificar que $\Box\varphi$ es equivalente a $\neg\Diamond\neg\varphi$; es decir, que es suficiente con un sólo conectivo modal para expresar todas las fórmulas. Al lector interesado en lógica modal le recomendamos consultar [11, 50].

Se han dado varias interpretaciones de los operadores $\Box$ y $\Diamond$. En general en la lógica modal se les interpreta como necesidad y posibilidad respectivamente. En la lógica deóntica se les interpreta como obligatorio y permisible (ver [49]). También se les ha estudiado como operadores temporales, interpretándolos como siempre y eventualmente (ver [54, 40, 46]).

En lógica epistémica se interpreta a $\Box$ como saber. A partir de los trabajos de Hintikka se utilizó otra notación, sustituyéndo a $\Box$ por K_a, en referencia a *knowledge*, conocimiento en inglés (ver [48]). La interpretación es que $K_a\varphi$ es verdadera si *a sabe* φ. Una aportación importante de Hintikka fue la posibilidad de indexar los operadores de conocimiento a una persona, un agente o un procesador y entonces se puede hacer razonamientos sobre lo que alguien sabe y sobre lo que sabe sobre lo que saben los demás, etc.. En las correspondientes estructuras epistémicas de los mundos posibles, cada persona, agente o procesador puede tener una relación de accesibilidad distinta, con lo que se representa su propia visión de la realidad, es decir, que se representa lo que cada uno considera posible.

2.1.1 Sintaxis

Supongamos que tenemos un conjunto de n agentes (o procesadores, o robots). Como queremos razonar acerca del mundo en el que viven estos agentes, asumiremos que es un mundo que puede ser descrito en términos de un conjunto Φ no vacío de proposiciones primitivas, etiquetadas como $p, q, p', q', \ldots$. En un sistema distribuido, estas proposiciones primitivas representarán afirmaciones como "el valor de la variable x es 0", "el procesador 3 está fallando", "la entrada inicial del procesador 1 fue 17" o "el sistema está en *deadlock*". Asumimos que tenemos una sintaxis de lógica proposicional con los conectivos usuales

$(\neg, \rightarrow)$. Para poder hacer afirmaciones del tipo de "El procesador 1 sabe que el procesador 3 está fallando" aumentaremos el lenguaje proposicional con los operadores $K_1, \ldots, K_n$ (uno para cada agente). La afirmación $K_i\varphi$ se lee "el agente i sabe φ." Definimos $\mathcal{L}_n(\Phi)$ como el conjunto más pequeño de fórmulas que contiene Φ, cerrado bajo negación, conjunción y los operadores modales $K_1, \ldots, K_n$. Entonces si φ y ψ son fórmulas, también lo son $\neg\varphi$, $\varphi \wedge \psi$ y $K_i\varphi$, $i = 1, \ldots, n$. Usaremos las abreviaciones usuales $\varphi \vee \psi$ por $\neg(\neg\varphi \wedge \neg\psi)$ y $\varphi \rightarrow \psi$ por $\neg(\varphi \wedge \neg\psi)$. Tomaremos $\top$ como una constante que siempre es verdadera; $\bot$ es una constante que siempre es falsa.

2.1.2 Semántica de los mundos posibles

Estando en un mundo, a cada agente le asociamos el conjunto de mundos que el agente considera que, de acuerdo a su conocimiento, pudieran ser el mundo real. Decimos entonces que un agente *sabe* un hecho φ exactamente si φ es verdadero en todos los mundos que el agente considera podrían ser el mundo real. El agente *no sabe* φ si considera posible al menos un mundo en el que φ no es verdadera. El estado del conocimiento de un agente corresponde a su posibilidad de determinar en que mundo se encuentra.

Una *estructura Kripke* para n agentes es una tupla $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, donde S es un conjunto de *estados* o *mundos posibles*; en cada estado $s \in S$, $\pi(s)$ es una asignación de verdad a las proposiciones primitivas de Φ (i.e. $\pi(s) : \Phi \rightarrow \{verdadero, falso\}$ para cada estado $s \in S$); $\mathcal{K}_i$ es una relación binaria sobre los estados de S, para $i = 1, \ldots, n$. Estas relaciones capturan las relaciones de posibilidad de acuerdo al agente i, $(s,t) \in \mathcal{K}_i$ si en el mundo s de la estuctura M, el agente i considera a t un mundo posible.

Definimos $\mathcal{M}_n$ como la clase de todas las estructuras Kripke para n agentes.

Definición 1 *Sea (M,s) la pareja que consiste de una estructura M y un estado s de M, $\varphi \in \mathcal{L}_n(\Phi)$. $(M,s) \models \varphi$ se lee como "φ es verdadero en (M,s)", "(M,s) satisface φ" ó "φ se cumple en (M,s)".*

$$(M,s) \models p \text{ (para } p \in \Phi) \text{ syss } \pi(s)(p) = verdadero,$$

$(M,s) \models \top$,

No $(M,s) \models \bot$,

$(M,s) \models \neg\varphi$ *syss no* $(M,s) \models \varphi$,

$(M,s) \models \varphi \wedge \psi$ *syss* $(M,s) \models \varphi$ *y* $(M,s) \models \psi$,

$(M,s) \models \varphi \vee \psi$ *syss* $(M,s) \models \varphi$ *o* $(M,s) \models \psi$ *o ambos*,

$(M,s) \models \varphi \rightarrow \psi$ *syss si* $(M,s) \models \varphi$ *entonces* $(M,s) \models \psi$,

$(M,s) \models K_i\varphi$ *syss* $(M,t) \models \varphi$ *para toda* t *tal que* $(s,t) \in \mathcal{K}_i$.

Las primeras siete corresponden a la definición estándar de verdad en la lógica proposicional. La última captura la intuición de que estando en el estado s de la estructura M, el agente i sabe φ exactamente cuando φ es verdadera en todos los mundos que i considera posibles.

En lógica modal es común dar una interpretación gráfica de una estructura Kripke en la que los nodos corresponden a los mundos o estados. Las etiquetas en cada nodo describen si las proposiciones primitivas son verdaderas o falsas en el mundo que representa. Hay una arista dirigida de un mundo s a un mundo t etiquetada con el agente i, si estando en s el agente i no distingue a t, es decir si $(s,t) \in \mathcal{K}_i$.

2.1.3 Propiedades del conocimiento

Hemos descrito un lenguaje con operadores modales K_i y hemos definido una noción de verdad que, en particular, determina si una fórmula como $K_i\varphi$ es verdadera en un mundo posible. $K_i\varphi$ debe ser leída como "el agente i sabe φ". ¿Es esto razonable? La semántica que hemos dado, esto es, las estructuras Kripke y la definición de verdad, ¿capturan realmente las propiedades del conocimiento para procesadores que estamos tratando modelar? ¿Cómo podemos responder a esto? Una manera de responder es examinando cuáles son las propiedades del conocimiento bajo esta interpretación. Podemos hacerlo caracterizando las fórmulas que siempre son verdaderas. Dada una estructura $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, decimos que φ es *válida* en M, y escribimos

$M \models \varphi$, si $(M, s) \models \varphi$ para todo estado $s \in S$; decimos que φ es *satisfactible* en M si $(M, s) \models \varphi$ para algún estado $s \in S$. Decimos que φ es válida, y escribimos $\models \varphi$, si φ es válida en todas las estructuras, y decimos que φ es satisfactible si es satisfactible en alguna estructura.

Proposición 1 *φ es válida en M (resp. válida en $\mathcal{M}_n$) syss $\neg\varphi$ no es satisfactible en M (resp. no es satisfactible en $\mathcal{M}_n$).*

Prueba

φ es válida en M syss $(M, s) \models \varphi$ para todo estado s en M syss no $(M, s) \models \neg\varphi$ para todo s en M, por la definición de verdad, syss $\neg\varphi$ no es satisfactible en M.

φ es válida en $\mathcal{M}_n$ syss $M \models \varphi$ para toda estructura M en $\mathcal{M}_n$ syss no $M \models \neg\varphi$ para toda estructura, por la definición de verdad, syss $\neg\varphi$ no es satisfactible en $\mathcal{M}_n$.□

Diremos que un enunciado es *proposicionalmente atómico* si es una proposición primitiva o si es de la forma $K_i\varphi$.

Una *valuación* es una asignación de verdad a los enunciados proposicionalmente atómicos que extiende a las asignaciones de verdad para las proposiciones primitivas, es decir que asigna verdadero o falso a las proposiciones primitivas y a las fórmulas $K_i\varphi$. La valuación de verdad del resto de los enunciados del lenguaje está determinada por su estructura. Esto es, $\top$ es verdadera en toda valuación, $\bot$ es falsa en cada valuación, $\neg\varphi$ es verdadera en una valuación si y sólo si φ es falsa, $\varphi \wedge \psi$ es verdadera si y sólo si φ es verdadera y ψ es verdadera.

Entonces una valuación analiza a los enunciados semánticamente desde el punto de vista de su estructura funcional de verdad. Un enunciado es una tautología si es verdadera en cada valuación posible de los enunciados proposicionalmente atómicos que lo constituyen.

Hacemos notar que en cada estructura Kripke M cada estado $s \in S$ induce una valuación a cada enunciado proposicionalmente atómico φ, en la que φ es verdadero si ocurre que $(M, s) \models \varphi$ y falso en otro caso.

El siguiente teorema captura algunas de las propiedades formales de $\models$.

Teorema 1 *Para toda fórmula* φ, $\psi \in \mathcal{L}_n(\Phi)$, *estructura* $M \in \mathcal{M}_n$, *y agente* i, $i = 1, \ldots, n$.

(1) Si φ *es una tautología entonces* $\models \varphi$

(2) Si $M \models \varphi$ *y* $M \models \varphi \rightarrow \psi$, *entonces* $M \models \psi$ (modus ponens)

(3) $\models (K_i\varphi \wedge K_i(\varphi \rightarrow \psi)) \rightarrow K_i\psi$ (axioma K)

(4) Si $M \models \varphi$ *entonces* $M \models K_i\varphi$ (regla de generalización del conocimiento)

Prueba

(1) Si φ es una tautología entonces φ es verdadera en cada valuación posible. Entonces φ es verdadera en cada mundo posible de cada modelo porque es verdadera en la valuación inducida por cada estado s en M, entonces $(M, s) \models \varphi$ para toda estructura Kripke M y todo s en S. Entonces $\models \varphi$.

(2) Supongamos que $M \models \varphi$ y $M \models \varphi \rightarrow \psi$. Entonces para todo s en M, $(M, s) \models \varphi$ y $(M, s) \models \varphi \rightarrow \psi$. Entonces $(M, s) \models \psi$ para todo estado s en M por la definición de verdad de la implicación, por lo que $M \models \psi$.

(3) Si $(M, s) \models (K_i\varphi \wedge K_i(\varphi \rightarrow \psi))$, entonces $(M, s) \models K_i\varphi$ y $(M, s) \models K_i(\varphi \rightarrow \psi)$, entonces para todo estado $t \in S$ tal que $(s, t) \in \mathcal{K}_i$, se cumple $(M, t) \models \varphi$ y $(M, t) \models \varphi \rightarrow \psi$. Entonces por la definición de verdad se cumple $(M, t) \models \psi$. Por lo tanto $(M, s) \models K_i\psi$, para todo estado s y toda estructura M, por lo tanto $\models (K_i\varphi \wedge K_i(\varphi \rightarrow \psi))$

(4) Si $M \models \varphi$, entonces $(M, t) \models \varphi$ para todo estado t en M. En particular, si fijamos un estado s en M, sabemos que $(M, t') \models \varphi$ para todo t' tal que $(s, t') \in \mathcal{K}_i$. Por lo tanto $(M, s) \models K_i\varphi$ para todo estado s en M, y entonces $M \models K_i\varphi$.□

Los incisos (1) y (2) quieren decir que la lógica de proposiciones está incluida en la lógica modal que estamos considerando.

El axioma K nos dice que $(K_i\varphi \wedge K_i(\varphi \rightarrow \psi) \rightarrow K_i\psi)$ es una fórmula válida; a esta propiedad también se le conoce como el *axioma distributivo*, ya que permite distribuir el operador K_i sobre la implicación. La regla de generalización del conocimiento implica que cada agente sabe todas las fórmulas válidas en una estructura.

2.1.4 Conocimiento común

En sistemas distribuidos en muchas ocasiones se piensa en protocolos con los que se busca llegar a un acuerdo o a un consenso entre los procesadores del sistema (si se piensa que los procesadores pueden fallar el acuerdo se toma entre los procesadores que no han fallado). También hay protocolos que buscan coordinar a los procesadores para que realicen alguna tarea (lo que involucra un acuerdo). Un concepto que ha resultado muy útil para analizar este tipo de protocolos es el de *conocimiento común*. El conocimiento común se define como un operador modal entre un grupo de procesadores. Intuitivamente un grupo tiene conocimiento común de un hecho φ si todos saben φ y todos saben que todos saben φ y todos saben que todos saben que todos saben φ, etc. El conocimiento común fue estudiado por primera vez en 1969 (ver [59]) en el contexto de las convenciones, en donde para que algo sea una convención entre los miembros de un grupo tiene que ser conocimiento común en el grupo. También ha aparecido en la economía a raíz de resultados publicados por Aumann (ver [2]), en teoría de juegos (ver [5]), en inteligencia artificial (ver [63]), y en el campo de la comprensión del discurso (ver [10]). El conocimiento común ha resultado ser importante en el estudio de problemas en los que se requiere de acuerdos simultáneos, por ejemplo, se ha estudiado variantes del problema de *acuerdo bizantino* (ver [14, 68]). Aunque hay que señalar que no hay consenso en la definición de qué es exactamente el conocimiento común. Las definiciones utilizadas han sido diferentes. (Para ver una discusión al respecto de estas distintas definiciones recomendamos [3]).

Para capturar estas nociones en el modelo de los mundos posibles se aumenta el lenguaje $\mathcal{L}_n(\Phi)$ con dos nuevos operadores modales: E_G ("Todos en el grupo G saben") y C_G ("Es conocimento común entre los miembros de G") y se define un nuevo lenguaje $\mathcal{L}_n^C(\Phi)$ que incluye a los nuevos operadores.

La interpretación de los nuevos operadores modales es la siguiente:

$$(M,s) \models E_G\varphi \; syss \; (M,s) \models K_i\varphi \; para \; toda \; i \in G$$

Definamos a $E_G^0\varphi$ como φ. Y sea $E_G^{k+1}\varphi$ igual a $E_G E_G^k\varphi$. Entonces

$$(M,s) \models C_G\varphi \; syss \; (M,s) \models E_G^k\varphi \; para \; k = 1,2,\ldots$$

Esta definición de conocimiento común tiene una interpretación gráfica interesante.

Definición 2 *Sea M una estructura Kripke, s y t dos estados en M y G un subconjunto de agentes.*

El estado t es **G-accesible** *desde el estado s en k pasos si existen estados $s_0, s_1, \ldots, s_k$ en M, tales que $s_0 = s$ y $s_k = t$, y para toda $j = 0, \ldots, k-1$, existe $i \in G$, tal que $(s_j, s_{j+1}) \in \mathcal{K}_i$; t es G-accesible desde s, si lo es en algún número de pasos.*

Es decir que t es G-accesible desde s si hay un camino de s a t cuyas aristas estan etiquetadas sólo con miembros de G. Si el conjunto G consiste de todos los agentes diremos simplemente que t es *accesible* desde s. Entonces t es accesible desde s si y sólo si hay un camino dirigido de s a t.

Lema 1 *Para toda estructura M en $\mathcal{M}_n$*

(a) $(M,s) \models E_G^k \varphi$ syss $(M,t) \models \varphi$ para toda t G-accesible desde s en k pasos.

(b) $(M,s) \models C_G \varphi$ syss $(M,t) \models \varphi$ para toda t G-accesible desde s.

Prueba

(a) Por inducción sobre k. Si $k = 0$, entonces por definición $E_G^0 \varphi = \varphi$ y t es G-accesible desde s en 0 pasos si y sólo si $s = t$. Entonces $(M,s) \models E_G^0 \varphi$ syss $(M,t) \models \varphi$.

Supongamos que se cumple para toda $k' < k$. Por definición $E_G^k \varphi = E_G E_G^{k-1} \varphi$. Entonces $(M,s) \models E_G^k \varphi$ syss $(M,s) \models E_G E_G^{k-1} \varphi$ syss se cumple, por la definición de E_G, $(M,s) \models K_i E_G^{k-1} \varphi$ para todo i en G, syss $(M,t) \models E_G^{k-1} \varphi$ para todo t' tal que $(s,t') \in \mathcal{K}_i$ syss $(M,t) \models \varphi$ para toda t G-accesible desde s en k pasos, por hipótesis de inducción y porque un mundo es G-accesible desde s en k pasos syss es G-accesible desde t' en $k-1$ pasos para todo t' tal que $(s,t') \in \mathcal{K}_i$).

(b) Por definición $(M,s) \models C_G \varphi$ syss $(M,s) \models E_G^k \varphi$, para $k = 1, 2, \ldots$ syss (por el inciso a) $(M,t) \models \varphi$ para toda t G-accesible desde s.□

Veamos ahora tres propiedades válidas en todas las estructuras Kripke para los operadores E_G y C_G.

Teorema 2 *Para toda fórmula φ y ψ en $\mathcal{L}_n^C$, toda estructura M en $\mathcal{M}_n$ y todos los agentes $i = 1, \ldots, n$.*

(a) $M \models E_G\varphi \leftrightarrow \bigwedge_{i \in G} K_i\varphi$

(b) $M \models C_G\varphi \leftrightarrow E_G(\varphi \wedge C_G\varphi)$ (axioma del punto fijo)

(c) Si $M \models \varphi \rightarrow E_G(\psi \wedge \varphi)$ *entonces* $M \models \varphi \rightarrow C_G\psi$ (regla de inducción)

Prueba

(a) $M \models E_G\varphi$ syss para todo i en G se cumple $M \models K_i\varphi$ syss $M \models \bigwedge_{i \in G} K_i\varphi$, por la interpretación de la conjunción

(b) Supongamos que para un estado s en M, $(M, s) \models C_G(\varphi)$. Entonces por el lema 1 $(M, t) \models \varphi$ para todo estado t G-accesible desde s. En particular si u es accesible desde s en un paso se cumple que $(M, u) \models \varphi$ y además como todo estado t G-accesible desde u es G-accesible desde s entonces $(M, t) \models \varphi$ para todo estado t G-accesible desde u. Entonces $(M, u) \models \varphi \wedge C_G\varphi$ para toda u G-accesible desde s en un paso. Entonces $(M, s) \models E_G(\varphi \wedge C_G\varphi)$.

Para el converso supongamos que $(M, s) \models E_G(\varphi \wedge C_G\varphi)$. Supongamos que t es G-accesible desde s y sea s' el primer estado en la cadena que va de s a t definida por los miembros de G. Entonces $(M, s') \models \varphi \wedge C_G\varphi$ porque s' es G-accesible desde s en un paso. Si $s' = t$ entonces $(M, t) \models \varphi$. Si $s \neq t$ entonces t es G-accesible desde s'. Como $(M, s') \models C_G\varphi$ entonces por el lema 1 se cumple $(M, t) \models \varphi$. En cualquier caso se cumple que $(M, t) \models \varphi$ para todo estado t G-accesible desde s y por lo tanto $(M, s) \models C_G\varphi$.

(c) Supongamos que $M \models \varphi \rightarrow E_G(\psi \wedge \varphi)$ y sea s un estado en M tal que $(M, s) \models \varphi$. Por inducción sobre k se mostrará que $(M, t) \models \varphi \wedge \psi$ para todo estado t G-accesible desde s en k pasos. Para $k = 1$, supongamos que t es G-accesible desde s en un paso. Por hipótesis se cumple $M \models \varphi \rightarrow E_G(\psi \wedge \varphi)$, entonces en particular se cumple $(M, s) \models \varphi \rightarrow E_G(\psi \wedge \varphi)$ y también sabemos que se cumple $(M, s) \models$

φ, entonces por modus ponens $(M,s) \models E_G(\psi \wedge \varphi)$, y por el lema 1 $(M,t) \models \psi \wedge \varphi$ porque t es G-accesible desde s en un paso.

Supongamos que para todo t' G-accesible desde s en $k-1$ pasos se cumple que $(M,t') \models \psi \wedge \varphi$. Sea t G-accesible desde s en k pasos. Entonces exite un estado t' G-accesible desde s en $k-1$ pasos y G-accesible desde t en un paso. Por hipótesis de inducción $(M,t') \models \psi \wedge \varphi$ y entonces $(M,t') \models \varphi$. Como $M \models \varphi \rightarrow E_G(\psi \wedge \varphi)$, se cumple $(M,t') \models \varphi \rightarrow E_G(\psi \wedge \varphi)$. Por modus ponens se cumple $(M,t') \models E_G(\psi \wedge \varphi)$, y por el lema 1 $(M,t) \models \psi \wedge \varphi$ porque t es G-accesible desde t' en un paso.

Entonces $(M,t) \models \varphi \wedge \psi$ para toda t G-accesible desde s y por lo tanto $(M,t) \models \psi$ por lo que $(M,s) \models C_G\psi$.□

El axioma del punto fijo dice que un grupo tiene conocimiento común de un hecho φ exactamente en una situación en la que todos saben que φ es verdadero y todos saben que es verdadero que hay conocimiento común del hecho. Este axioma es una propiedad central que hace del conocimiento común un prerrequisito necesario para obtener acuerdo y coordinación. La regla de inducción proporciona una técnica para verificar si hay conocimiento común. La razón de su nombre es que si sabemos que $(\varphi \rightarrow E_G(\psi \wedge \varphi))$ es válida, entonces, como se hizo en la demostración, se demuestra por inducción sobre k que $(\varphi \rightarrow E_G^k(\psi \wedge \varphi))$ es válida para toda k, y se concluye que $(\varphi \rightarrow C_G\psi)$ es válida. En la demostración de hecho se prueba que si $M \models \varphi \rightarrow E_G(\psi \wedge \varphi)$ entonces $M \models \varphi \rightarrow C_G(\psi \wedge \varphi)$, pero se acostumbra más presentar a esta regla como se indica.

2.1.5 Estructuras Kripke $S5$

Hasta ahora no hemos dicho nada con respecto a las propiedades de las relaciones $\mathcal{K}_i$ de las estructuras Kripke. Como se verá en la sección 2.2, cuando pensamos en modelar al conocimiento en sistemas distribuidos, estas relaciones son relaciones de equivalencia.

Una *estructura Kripke S5* es aquella en la que las relaciones $\mathcal{K}_i$ son relaciones de equivalencia. Pensemos que el conjunto Φ de proposiciones primitivas está fijo. Hablaremos entonces del lenguaje $\mathcal{L}_n^C$ en

vez de $\mathcal{L}_n^C(\Phi)$ y definamos $\mathcal{M}_n^{rst}$ como la clase de todas las estructuras Kripke $S5$ para n agentes sobre $\mathcal{L}_n^C$.

Teorema 3 *Para toda fórmula φ y ψ en $\mathcal{L}_n$, toda estructura M en $\mathcal{M}_n^{rst}$ y todos los agentes $i = 1, \ldots, n$.*

(a) $M \models (K_i\varphi \wedge K_i(\varphi \rightarrow \psi)) \rightarrow K_i\psi$ (axioma K)

(b) Si $M \models \varphi$ *entonces* $M \models K_i\varphi$ (regla de generalización del conocimiento)

(c) $M \models E_G\varphi \leftrightarrow \bigwedge_{i \in G} K_i\varphi$

(d) $M \models C_G\varphi \leftrightarrow E_G(\varphi \wedge C_G\varphi)$ (axioma del punto fijo)

(e) Si $M \models \varphi \rightarrow E_G(\psi \wedge \varphi)$ *entonces* $M \models \varphi \rightarrow C_G\psi$ (regla de inducción)

(f) $M \models K_i\varphi \rightarrow \varphi$ (axioma T)

(g) $M \models K_i\varphi \rightarrow K_iK_i\varphi$ (axioma 4)

(h) $M \models \neg K_i\varphi \rightarrow K_i\neg K_i\varphi$ (axioma 5)

Prueba

Ya vimos la prueba para (a), (b), (c), (d) y (e), éstas no dependen del tipo de relación. Son válidas en la clase de todas las estructuras Kripke.

Sea s un estado en M.

(f) Supongamos que se cumple $(M, s) \models K_i\varphi$ para algún agente i. Entonces para todo estado t tal que $(s, t) \in \mathcal{K}_i$, se cumple $(M, t) \models \varphi$. Como $\mathcal{K}_i$ es reflexiva, $(s, s) \in \mathcal{K}_i$. Entonces $(M, s) \models \varphi$. Por lo tanto $(M, s) \models K_i\varphi \rightarrow \varphi$

(g) Supongamos que se cumple $(M, s) \models K_i\varphi$ para algún agente i. Sea t tal que $(s, t) \in \mathcal{K}_i$, y sea u tal que $(t, u) \in \mathcal{K}_i$. Como $\mathcal{K}_i$ es transitiva entonces $(s, u) \in \mathcal{K}_i$. Por lo tanto $(M, u) \models \varphi$. Entonces se cumple que para toda t tal que $(s, t) \in \mathcal{K}_i$, $(M, t) \models K_i\varphi$ y entonces $(M, s) \models K_iK_i\varphi$. Por lo tanto $(M, s) \models K_i\varphi \rightarrow K_iK_i\varphi$

(h) Supongamos que $(M, s) \models \neg K_i\varphi$. Entonces existe u tal que $(s, u) \in \mathcal{K}_i$ y $(M, u) \models \neg\varphi$. Sea t tal que $(s, t) \in \mathcal{K}_i$; como $\mathcal{K}_i$ es

simétrica, entonces $(t,s) \in \mathcal{K}_i$; como también es transitiva, entonces $(t,u) \in \mathcal{K}_i$. Entonces $(M,t) \models \neg K_i\varphi$ y entonces $(M,s) \models K_i\neg K_i\varphi$. Por lo tanto $(M,s) \models \neg K_i\varphi \rightarrow K_i\neg K_i\varphi$.□

Al axioma T también se le llama el *axioma del conocimiento* debido a que si bien un agente puede no conocer todos los hechos verdaderos, si un agente sabe algo, entonces ese hecho es verdadero. Las dos últimas propiedades dicen que los agentes pueden hacer introspección con respecto a su conocimiento. Al axioma 4 también se le conoce como el *axioma de la introspección positiva.* Al axioma 5 también se le conoce como el *axioma de la introspección negativa.*

Un axioma es una fórmula y una regla de inferencia es de la forma "De $\varphi_1,\ldots,\varphi_k$ se infiere ψ", donde $\varphi_1,\ldots,\varphi_k,\psi$ son fórmulas. Un sistema de axiomas consiste en una colección de *axiomas* y de *reglas de inferencia.* En lógica modal se identifica a los sistemas de axiomas por los nombres de los axiomas que los componen. Los sistemas en general incluyen a todas las tautologías, a modus ponens, al axioma K y a la regla de la generalización del conocimiento. Si añadimos el axioma T obtenemos el sistema T_n. Si añadimos a T y a 4 obtenemos $S4_n$ y si añadimos a T, a 4 y a 5 obtenemos al sistema $S5_n$. Si añadimos las propiedades que tienen que ver con los operadores E_G y C_G, obtenemos los sistemas T_n^C, $S4_n^C$ y $S5_n^C$. Los sistemas de axiomas se refieren a la sintaxis de la lógica modal; en ésta tenemos que hacer demostraciones a partir de los axiomas y las reglas de inferencia. Nosotros estamos interesados en la semántica, pero de hecho hay una fuerte relación entre un sistema de axiomas y el modelo semántico. Fijémonos en las demostraciones que hemos hecho para las propiedades del conocimiento. Para la propiedad que llamamos axioma K no asumimos nada sobre las relaciones de la estructura Kripke. Para el axioma T utilizamos que la relación es reflexiva. Para el axioma 4 utilizamos la transitividad y para 5 utilizamos que la relación es simétrica y transitiva. Esto nos da una relación entre los axiomas y las propiedades de las relaciones entre los estados de las estructuras Kripke.

Decimos que un sistema de axiomas AX es *correcto* con respecto a una clase de estructuras $\mathcal{M}$ si todo lo que es demostrable a partir de AX es válido en $\mathcal{M}$. El sistema de axiomas AX es *completo* si todo lo que es válido en la clase $\mathcal{M}$ es demostrable a partir de AX. La relación

entre la clase de las estructuras Kripke y los sistemas de axiomas se da a través de la relación entre las propiedades de las relaciones en las estructuras y los axiomas. Resulta entonces que los sistemas de axiomas T_n y T_n^C son correctos y completos con respecto a la estructuras Kripke reflexivas. Los sistemas $S4_n$ y $S4_n^C$ son correctos y completos respecto a las estructuras relexivas y transitivas, y los sistemas $S5_n$ y $S5_n^C$ son correctos y completos con respecto a la estructuras Kripke reflexivas, simétricas y transitivas, esto es, con respecto a la clase de las estructuras Kripke $S5$.

2.2 Conocimiento en sistemas distribuidos

Queremos utilizar el modelo de conocimiento que hemos descrito para analizar sistemas distribuidos; usaremos básicamente el modelo que apareció de forma más completa en [21], los mismos autores dan una versión revisada en [24]. El lector interesado puede consultar [34] y [39] para ver un panorama del razonamiento sobre conocimiento en sistemas distribuidos, y una amplia discusión sobre las propiedades del conocimiento en sistemas distribuidos en [20].

2.2.1 Ejecuciones y sistemas

Pensemos en un sistema distribuido como un conjunto, $\{1, \ldots, n\}$, de n agentes o procesadores. En cualquier momento de tiempo cada agente está en algún *estado local*. Pensemos que el estado local de un agente captura toda la información a la que tiene acceso. En una partida de dominó, el estado local de un jugador puede incluir las fichas que tiene y las fichas que han sido jugadas por cada jugador; también podría incluir cierta información sobre la forma de juego de los adversarios, por ejemplo Beto podría saber que a Alicia le gusta tirar sus mulas rápidamente o que Carlos ocasionalmente sale en falso. En un sistema distribuido, el estado local de un procesador pude incluir algunos valores iniciales que haya leído, la lista de mensajes que ha enviado, que ha recibido, y posiblemente el valor de un reloj.

Además de los agentes, es útil considerar al *medio ambiente*, que sería todo lo que es relevante para el análisis del sistema y que no está en los estados locales de los agentes. En algunos casos será conveniente pensar también en el medio ambiente como otro agente, aunque generalmente tendrá un papel especial. El medio ambiente tiene también un estado local, *el estado local del medio ambiente*. Si estamos analizando un *sistema de mensajes* donde los procesadores están mandando mensajes a través de una red de líneas de comunicación, el estado local del medio ambiente puede estar guardando el conjunto de mensajes en tránsito, que ya han sido enviados pero todavía no son entregados. Muchas veces el medio ambiente se modela para que tenga un compor-

tamiento no determinístico o probabilístico.

El *estado global* de un sistema con n agentes es una $(n+1)$-ada de la forma $(\ell_{ma}, \ell_1, \ldots, \ell_n)$ donde ℓ_{ma} es el estado local del medio ambiente y ℓ_i es el estado local del agente i, $i = 1, \ldots, n$.

Cada estado global describe al sistema en un determinado momento. Sin embargo, el análisis de un sistema consiste en estudiar sus cambios con el transcurso del tiempo, estudiar cuáles son los diferentes estados globales que va adquiriendo, así que necesitamos incorporar al tiempo en nuestro modelo. Una *ejecución* es una función que va del tiempo a los estados globales. Por simplicidad asumiremos que el tiempo corre sobre los números naturales[1]. Entonces $r(0)$ describe el *estado global inicial* del sistema en una posible ejecución r; el siguiente estado global es $r(1)$, etc. Puede pensarse que una ejecución r es una sucesión $r(0), r(1), \ldots$ de los estados globales que va tomando el sistema. El tiempo se va midiendo en algún reloj externo al sistema, no se asume necesariamente que los agentes tengan acceso a este reloj.

Un sistema puede tener muchas posibles ejecuciones, su estado global puede ir evolucionando de muchas maneras. Pensando en el ejemplo de la partida de dominó, los estados globales iniciales serían todas las posibles reparticiones de las 28 fichas entre los cuatro jugadores, donde el estado ℓ_i de cada jugador guarda las fichas que le tocaron. Para cada estado global inicial fijo hay muchas posibles "ejecuciones" del juego. En un sistema de mensajes, un mensaje en particular puede o no ser entregado; entonces también para un estado inicial fijo hay muchas posibles ejecuciones.

Formalmente definamos un *sistema* como un conjunto no vacío de ejecuciones. Esta definición modela los posibles comportamientos del sistema; el requerimiento de que sea un conjunto no vacío implica que el sistema tiene algún comportamiento.

Resumimos toda la discusión anterior en la siguiente definición:

[1] Esta decisión no es crucial pero así se hace normalmente. De igual manera se podría modelar el tiempo sobre los números reales; sin embargo, se quiere asegurar que en cualquier tiempo sólo ocurra un número finito de eventos (que corresponden a cambios en el estado global); esto es siempre cierto si escogemos que el tiempo corre sobre los naturales.

Definición 3 *Sea L_{ma} el conjunto de posibles estados del medio ambiente y sea L_i el conjunto de posibles estados locales de cada agente i, $i = 1, \ldots, n$. El conjunto de estados globales es $\mathcal{G} = L_{ma} \times L_1 \times \ldots \times \cdots \times L_n$. Una ejecución r sobre $\mathcal{G}$ es una función $r: \mathcal{N} \rightarrow \mathcal{G}$. Un* punto *es una pareja (r, t) que consiste en una ejecución r y el tiempo t. Si $r(t) = g = (\ell_e, \ell_1, \ldots, \ell_n)$ es el estado global en el punto (r, t), definimos $r_{ma}(t) = g_{ma} = \ell_{ma}$ y $r_i(t) = g_i = \ell_i$, para $i = 1, \ldots, n$. Una* ronda *tiene lugar entre dos tiempos. La ronda t en la ejecución r ocurre entre el tiempo $t-1$ y el tiempo t. Un* sistema *$\mathcal{R}$ sobre $\mathcal{G}$, es un conjunto de ejecuciones sobre $\mathcal{G}$. Decimos que (r, t) es un* punto del sistema *$\mathcal{R}$ si $r \in \mathcal{R}$.*

2.2.2 Algunos sistemas

Presentamos dos ejemplos de sistemas que nos serán útiles. Estos sistemas aparecieron como ejemplo en [21].

Sistemas síncronos

En muchos sistemas es común asumir que los procesadores o agentes tienen acceso a un reloj global, o que las acciones del sistema se llevan a cabo en rondas y en todo momento todos los procesadores saben cuál es la ronda que se está llevando a cabo. Visto de otra manera se asume implícitamente que el tiempo es conocimiento común por lo que todos los procesadores ejecutan sus acciones en sincronía. Muchos protocolos estan diseñados para ser ejecutados en rondas, de manera que ningún procesador inicia la ronda $t+1$ hasta que todos los demás hayan terminado la ronda t.

Para capturar la idea de un sistema síncrono en el modelo que hemos presentado, debemos recordar que el conocimiento de los procesadores es determinado por su estado local. En un sistema síncrono asumimos que los procesadores saben en qué ronda se encuentran; entonces el tiempo tiene que ser parte de su estado local, aunque este tiempo no necesariamente debe corresponder al "tiempo real".

Formalmente:

Definición 4 *$\mathcal{R}$ es un* **sistema síncrono** *si para todo procesador i y todo punto (r, t) y (r', t'), si $r_i(t) = r'_i(t')$ entonces $t = t'$.*

Esto indica que en el punto (r,t) cada punto que un procesador considera posible tiene el tiempo t. En particular esto quiere decir que los procesadores distinguen los puntos del presente de los puntos del futuro, y entonces en cada punto de una ejecución un procesador se encuentra en un estado local diferente.

Sistemas de mensajes

En muchas ocasiones, cuando analizamos un protocolo en sistemas distribuidos, queremos concentrarnos únicamente en los aspectos del sistema que tengan que ver con la comunicación. Capturamos esta idea en un *sistema de mensajes* en el cual las acciones fundamentales son las de mandar y recibir mensajes. En un sistema de mensajes pensaremos que el estado local de los procesadores contiene información acerca de su estado inicial, los mensajes que ha recibido y los que ha enviado y las acciones internas que ha llevado a cabo.

Pensemos que tenemos determinado a un conjunto Σ_i de posibles estados locales iniciales, un conjunto INT_i de acciones internas y un conjunto MSG de mensajes para el procesador i. Representaremos con $enviar_i(msg,j)$ al evento de "el procesador i envía el mensaje msg al procesador j" y con $recibir_i(msg,j)$ al evento "el procesador i recibe el mensaje msg del procesador j"; $int_i(a)$ representa "el procesador i ejecuta la acción interna a". Definimos a $historia_i$, la *historia de mensajes del procesador* i, como la sucesión cuyo primer elemento es el conjunto vacío y los siguientes elementos son conjuntos no vacíos cuyos elementos son de la forma $enviar_i(msg,j)$, $recibir_i(msg,j)$ ó $int_i(a)$.

Intuitivamente la historia del procesador i en el punto (r,t) consiste de todos los eventos de los que ha tenido noticia el procesador hasta la ronda t. Si el procesador i no realizó ningún evento en la ronda t podemos pensar que su historia en el punto $(r,t+1)$ es la misma que la del punto (r,t) (y entonces no distinguirá ambos puntos). Si queremos que el sistema sea síncrono, pensaremos que el procesador realizó una acción especial Λ que representa a la acción nula y sirve para indicar que ha pasado el tiempo.

Definición 5 *Un sistema $\mathcal{R}$ sobre los conjuntos Σ_i y INT_i para $i = 1, \ldots, n$ y un conjunto MSG de mensajes es un* **sistema de mensajes** *si cada punto $(r,t) \in \mathcal{R}$ cumple con:*

SM1. $r_i(t)$ incluye la historia de mensajes de i sobre Σ_i, INT_i y MSG,

SM2. Para cada evento $recibir_i(msg,j)$ en $r_i(t)$ existe un evento correspondiente $enviar_j(msg,i)$ en $r_j(t)$,

SM3. $r_i(0)$ consiste del estado inicial del procesador i y de la historia vacía y $r_i(t+1)$ es idéntica a $r_i(t)$ o es el resultado de añadir a la historia en $r_i(t)$ un conjunto de eventos.

$SM1$ dice que el estado local de un procesador incluye a su historia de mensajes. $SM2$ garantiza que todos los mensajes recibidos en la ronda t corresponden a mensajes que fueron enviados en alguna ronda anterior. $SM3$ garantiza que las historias no pierden información.

Podemos añadir requerimientos a un sistema de mensajes, por ejemplo si queremos que sea un sistema de mensajes confiable, entonces debemos pedir que todos los mensajes que son enviados sean recibidos en algún momento; esto puede pedirse con la siguiente condición:

SM4. Para todo procesador i, j, y todo punto $(r,t) \in \mathcal{R}$. Si $enviar_i(msg,j)$ está en $r_i(t)$, entonces existe algún tiempo $t' \geq t$, tal que $recibir_j(msg,i)$ está en $r_j(t')$.

Otros requerimientos que podemos pedir a un sistema de mensajes es que los mensajes lleguen en el orden que fueron enviados (que los canales de comunicación funcionen como una *cola*), que el sistema sea síncrono o que el retardo en los canales de comunicación esté acotado.

2.2.3 Incorporando conocimiento

Hemos pensado que un sistema distribuido es un conjunto de ejecuciones; para incluir al conocimiento veamos como podemos asociar a los sistemas con las estructuras Kripke. Asumamos que tenemos un conjunto Φ de proposiciones primitivas que describen hechos básicos

sobre el sistema. Recordemos que este conjunto define a un lenguaje $\mathcal{L}_n(\Phi)$. Los mundos posibles son los puntos del sistema. Nos hace falta definir la interpretación π que es una función de los estados a asignaciones de verdad para las proposiciones primitivas y nos hace falta definir las relaciones entre los mundos.

Definición 6 *Un* **sistema interpretado** $\mathcal{I}$ *es una pareja* $(\mathcal{R}, \pi)$, *donde* $\mathcal{R}$ *es un sistema sobre un conjunto* $\mathcal{G}$ *de estados globales y* π *es un interpretación que define una asignación de verdad de las proposiciones de* Φ *para cada elemento en* $\mathcal{G}$. *Entonces, para todo* $p \in \Phi$ *y todo estado global* $g \in \mathcal{G}$, *tenemos* $\pi(g)(p) \in \{verdadero, falso\}$.

Claramente π introduce una asignación de verdad sobre los puntos de $\mathcal{R}$; simplemente tomemos $\pi(r,t)$ como $\pi(r(t))$. Es importante notar que Φ y π no son intrínsecos al sistema $\mathcal{R}$, constituyen una estructura adicional por encima de $\mathcal{R}$ que nosotros, como observadores, añadimos a nuestra conveniencia para ayudarnos a entender mejor el sistema. Nos referiremos a los puntos y estados del sistema $\mathcal{R}$ como puntos y estados, respectivamente, del sistema interpretado $\mathcal{I}$. Esto es, diremos que el punto (r,t) está en el sistema interpretado $\mathcal{I} = (\mathcal{R}, \pi)$ si $r \in \mathcal{R}$.

Para definir las relaciones $\mathcal{K}_i$ diremos que dos puntos del sistema $\mathcal{I}$, (r,t) y (r',t'), son indistinguibles para el agente i si el estado local del agente i es el mismo en ambos puntos, esto es si $r_i(t) = r'_i(t')$; escribiremos entonces que $(r,t) \sim_i (r',t')$. Claramente la relación $\sim_i$ es una relación de equivalencia.

Podemos asociar a un sistema interpretado $\mathcal{I} = (\mathcal{R}, \pi)$ una estructura Kripke $M_{\mathcal{I}} = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ de manera directa: S es el conjunto de puntos en $\mathcal{I}$. Las relaciones $\mathcal{K}_i$ están definidas por la relación $\sim_i$ y son entonces, relaciones de equivalencia. Estamos asociando a los sistemas interpretados con las estructuras Kripke $S5$.

Podemos definir qué significa que una fórmula $\varphi \in \mathcal{L}_n(\Phi)$ sea verdadera en un punto (r,t) del sistema interpretado $\mathcal{I}$, esto es $(\mathcal{I}, r, t) \models \varphi$, usando las definiciones de la sección anterior:

Definición 7 $(\mathcal{I}, r, t) \models \varphi$ *syss* $(M_{\mathcal{I}}, (r,t)) \models \varphi$

2.2.4 Acciones

No se ha mencionado hasta ahora de dónde vienen las ejecuciones. Pensamos en una ejecución como una sucesión de estados globales, ¿qué es lo que hace que un sistema cambie de estado global?. Intuitivamente lo que provoca estos cambios son las "acciones" realizadas por cada agente y por el medio ambiente. Es conveniente pensar que las acciones se realizan durante una ronda.

Consideraremos a las acciones como elementos de un conjunto específico. Asumiremos que para cada agente i hay un conjunto ACT_i de acciones que puede realizar. Por ejemplo, en un sistema distribuido, una acción sería $enviar_i(msg, j)$; intuitivamente esta acción corresponde al evento de i enviando a j el mensaje msg. Manteniendo el criterio de ver al medio ambiente como un agente, existe un conjunto ACT_{ma} de las acciones que puede realizar el medio ambiente. Para los agentes y para el medio ambiente consideramos la posibilidad de que no realicen acción alguna, representada por la *acción nula* Λ.

En ocasiones no será suficiente ver por separado las acciones que realizan los componentes de un sistema para saber cómo cambia el estado global del sistema; habrá que considerar las acciones realizadas simultáneamente, esto es, en la misma ronda. Si tres procesadores tratan de escribir el valor de una variable en el mismo registro no será muy claro saber lo que sucederá a partir de lo que está haciendo cada uno de ellos. Para resolver lo que ocurre con acciones que interactúen entre sí consideraremos a las *acciones conjuntas.* Una acción conjunta es una tupla de la forma $(a_{ma}, a_1, \ldots, a_n)$, donde $a_{ma} \in ACT_{ma}$ y $a_i \in ACT_i$, $i = 1, \ldots, n$.

Definición 8 *Un* **transformador de estados globales** $\mathcal{T}$ *es una función de estados globales a estados globales,* $\mathcal{T} : \mathcal{G} \to \mathcal{G}$*, asociada a cada acción conjunta. La* **función de transición** τ *es un mapeo que asocia un transformador de estados globales* $\tau(a_{ma}, a_1, \ldots, a_n)$ *a cada acción conjunta* $(a_{ma}, a_1, \ldots, a_n)$.

Esta definición requiere que $\tau(a_{ma}, a_1, \ldots, a_n)(\ell_{ma}, \ell_1, \ldots, \ell_n)$ esté definido para cada acción conjunta $(a_{ma}, a_1, \ldots, a_n)$ y cada estado global $(\ell_{ma}, \ell_1, \ldots, \ell_n)$. En la práctica no todas las combinaciones de estados globales y acciones conjuntas serán de interés cuando analice-

mos el sistema, ya que muchas de ellas no serán realmente accesibles. En esos casos definiremos $\tau(a_{ma}, a_1, \ldots, a_n)(\ell_{ma}, \ell_1, \ldots, \ell_n)$ arbitrariamente. También definiremos $\tau(\Lambda, \ldots, \Lambda)$ como el *transformador nulo*, donde $\tau(\Lambda, \ldots, \Lambda)(\ell_{ma}, \ell_1, \ldots, \ell_n) = (\ell_{ma}, \ell_1, \ldots, \ell_n)$, para todo estado global.

2.2.5 Protocolos

En un sistema distribuido, los procesadores generalmente realizan sus acciones a partir de un protocolo. Intuitivamente el protocolo del procesador i es una descripción de las acciones que debe llevar a cabo en función de su estado local. Se deja abierta la posibilidad de que sea una función de estados locales a conjutos de acciones para poder modelar protocolos no determinísticos. Estando en un estado local, un procesador sólo ejecuta una de las acciones del conjunto; la elección de la acción se hace de forma no determinística. Un *protocolo determinístico* es aquel en el que el mapeo se hace sobre conjuntos de una sola acción. Por simplicidad si el protocolo es determinístico escribiremos $P_i(\ell_i) = a$, en vez de $P_i(\ell_i) = \{a\}$.

Definición 9 *Un* **protocolo** P_i *es una función del conjunto* L_i *de estados locales del agente* i *a conjuntos no vacíos de acciones en* ACT_i.
$P_i : L_i \rightarrow 2^{ACT_i} - \emptyset$

Ya que consideramos al conjunto de acciones ACT_{ma} para el medio ambiente, también consideraremos que el medio ambiente está ejecutando un protocolo $P_{ma} : L_{ma} \rightarrow 2^{ACT_{ma}} - \emptyset$. Por ejemplo, en un sistema de mensajes, el protocolo del medio ambiente puede determinar la posibilidad de que un mensaje se pierda, que se corrompa o que los mensajes no sean entregados en el orden en que fueron enviados.

Si todos los procesadores y el medio ambiente están ejecutando protocolos determinísticos, entonces hay una única ejecución a partir de un estado global inicial.

Los procesadores no ejecutan protocolos por sí mismos; la combinación de todos los protocolos es lo que causa que un sistema se comporte de determinada manera.

Definición 10 *Un* **protocolo conjunto** *es una tupla* $P = (P_1, \ldots, P_n)$, *donde* P_i *es el protocolo del agente* i.

No se incluye al protocolo del medio ambiente como parte del protocolo conjunto porque el medio ambiente juega un rol especial; usualmente se diseña y analiza los protocolos de los agentes, suponiendo que el protocolo del medio ambiente está fijo. En ocasiones se visualiza al medio ambiente como un adversario al que hay que vencer y que está tratando de que el sistema se comporte de forma indeseable.

2.2.6 Contextos

El protocolo conjunto P y el protocolo del medio ambiente describen al comportamiento de todos los participantes del sistema; sin embargo no determinan completamente el comportamineto del sistema en sí. Necesitamos saber el entorno o el "contexto" en el que se ejecuta el protocolo conjunto. El protocolo del medio ambiente P_{ma} debe ser parte del contexto ya que es lo que determina la injerencia del medio ambiente en el protocolo conjunto. También debemos incluir a la función de transición τ y el conjunto $\mathcal{G}_0$ de estados globales *iniciales*. Estos componentes del contexto nos dan una descripción del comportamiento del medio ambiente en cada paso de una ejecución.

En ocasiones queremos restringir más el comportamiento del medio ambiente, por ejemplo, si queremos establecer una condición como "todos los mensajes enviados eventualmente serán entregados". Una manera de hacerlo es estableciendo una condición de *admisibilidad* Ψ sobre las ejecuciones; esta condición nos dice cuáles son las ejecuciones aceptables. Formalmente Ψ es un conjunto de ejecuciones.

Definición 11 *Un* **contexto** *es una tupla* $(P_{ma}, \mathcal{G}_0, \tau, \Psi)$, *donde* P_{ma} *es el protocolo del medio ambiente,* $\mathcal{G}_0$ *es el conjunto de estados globales iniciales,* τ *es la función de transición y* Ψ *es un conjunto de ejecucioneses que representa una condición de admisibilidad donde para una ejecución* r, $r \in \Psi$ *si* r *satisface la condición* Ψ.

Al incluir τ dentro del contexto estamos incluyendo implícitamente a los conjuntos $L_{ma}, L_1, \ldots, L_n$ de estados locales y los conjuntos

$ACT_{ma}, ACT_1, \ldots, ACT_n$ de acciones ya que el conjunto de las acciones conjuntas es el dominio de τ y el conjunto de estados globales es el dominio de las transformaciones de estados globales indicadas por τ.

Hablemos ahora de las ejecuciones de un protocolo en un contexto dado

Definición 12 *Una ejecución r es* **débilmente consistente** *con un protocolo conjunto $P = (P_1, \ldots, P_n)$ en un contexto $\gamma = (P_{ma}, \mathcal{G}_0, \tau, \Psi)$ si*

1. $r(0) \in \mathcal{G}_0$
2. *para toda $t \geq 0$, si $r(t) = (\ell_{ma}, \ell_1, \ldots, \ell_n)$, entonces existe una acción conjunta $(a_{ma}, a_1, \ldots, a_n) \in P_{ma}(\ell_{ma}) \times P_1(\ell_1) \times \cdots \times P_n(\ell_n)$ tal que $r(t+1) = \tau(a_{ma}, a_1, \ldots, a_n)(r(t))$.*

   La ejecución r es **consistente** *con P en el contexto γ si además satisface.*
3. $r \in \Psi$

Siempre habrá ejecuciones que sean débilmente consistentes con un protocolo P en un contexto γ, pero es posible que no haya alguna ejecución consistente con P en γ. Esto sucede si ninguna ejecución en Ψ es débilmente consistente con P en γ. Entonces diremos que P es *inconsistente* con γ, de otra manera P es *consistente* con γ.

Diremos que un sistema $\mathcal{R}$ es *consistente* con el protocolo P en el contexto γ si todas las ejecuciones $r \in \mathcal{R}$ son consistentes con P en γ. En general va a haber muchos sistema consistentes con un protocolo en un contexto. Sin embargo cuando ejecutamos un protocolo en un contexto quisieramos saber cuál es el conjunto de todas las ejecuciones posibles del protocolo.

Definición 13 *El sistema $\mathcal{R}$* **representa** *al protocolo P en el contexto γ, si $\mathcal{R}$ es el conjunto de todas las ejecuciones consistentes con P en el contexto γ. Denotamos entonces $\mathcal{R} = \mathbf{R}^{rep}(P, \gamma)$.*

Nótese que un sistema $\mathcal{R}$ es consistente con el protocolo P en el contexto γ si y sólo si $\mathcal{R} \subseteq \mathbf{R}^{rep}(P, \gamma)$. Esto quiere decir que $\mathbf{R}^{rep}(P, \gamma)$ es el conjunto maximal consistente con P en γ.

Igual que definimos a los sistemas interpretados, definimos a un contexto interpretado como la pareja (γ, π) donde γ es un contexto y π es una interpretación. Ampliamos las definiciones diciendo que el sistema interpretado $\mathcal{I} = (\mathcal{R}, \pi)$ es *consistente* con un protocolo P en el contexto interpretado (γ, π) si $\mathcal{R}$ es consistente con P en γ. Se define a $\mathbf{I}^{rep}(P, \gamma, \pi) = (\mathbf{R}^{rep}(P, \gamma), \pi)$ como el sistema interpretado que representa al protocolo P en el contexto interpretado (γ, π).

2.2.7 Programas basados en conocimiento

Hemos definido a un protocolo como una función de estados locales a conjuntos de acciones. Generalmente se describe a un protocolo por medio de un *programa* escrito en algún lenguaje de programación. A continuación se describe un lenguaje de programación muy simple pero que es lo suficientemente rico para describir protocolos y cuya sintaxis pone énfasis en el hecho de que los agentes ejecutan sus acciones basados solamente en sus estados locales.

Programas estándar

Definición 14 *Un* programa estándar Pg_i *para el agente* i *es un enunciado de la forma:*

```
case of
    if t_1 do a_1
    if t_2 do a_2
    ...
end case
```

Donde t_j es una fórmula proposicional a la que llamaremos *prueba estándar*; a_j es una acción, esto es $a_j \in ACT_j$. Usaremos una interpretación π que nos dirá cómo evaluar las pruebas sobre el estado local de los procesadores. Pediremos que esta interpretación dependa únicamente del estado local de cada procesador, ya que sería inapropiado que las acciones que un procesador va a tomar dependan del valor de verdad de una prueba que no puede ser determinada desde su estado local.

Definición 15 *Para una proposición primitiva p que aparece en las pruebas del programa Pg_i, diremos que p es* **local** *para el agente i si para todos los estados g, g' en $\mathcal{G}$, tales que $g \sim_i g'$, se cumple $\pi(g)(p) = \pi(g')(p)$.*

La interpretación π sobre un conjunto de estados globales $\mathcal{G}$ es **compatible** *con el programa Pg_i para el agente i si cada proposición primitiva que aparece en Pg_i es local para el agente i.*

Definición 16 *Sea φ una fórmula proposicional tal que todas las proposiciones primitivas que la componen son locales al agente i, y sea $\ell \in L_i$, escribiremos $(\pi, \ell) \models \varphi$ si la asignación de verdad $\pi(g)$ satisface a φ, donde $g = (\ell_1, \ldots, \ell_n)$ es un estado global que cumple $\ell_i = \ell$.*

Dado un programa para el agente i queremos encontrar el protocolo correspondiente al programa. Intuitivamente el protocolo determina, dado un estado local, que se realice alguna de las acciones especificadas por las pruebas que son satisechas bajo ese estado local y selecciona la acción nula Λ si ninguna prueba es satisfecha. En general obtendremos un protocolo no determinístico porque más de una prueba podría ser satisfecha en cada estado.

Definición 17 *Dado el programa Pg_i para el agente i y la interpretación π compatible con Pg_i, definimos el protocolo Pg_i^π:*

$$Pg_i^\pi(\ell) = \begin{cases} \{a_j : (\pi, \ell) \models t_j\} & si\ \{j : (\pi, \ell) \models t_j\} \neq \emptyset \\ \{\Lambda\} & si\ \{j : (\pi, \ell) \models t_j\} = \emptyset \end{cases}$$

Muchas de las definiciones que dimos para protocolos tienen su análogo para programas.

Definición 18 *Un* **programa conjunto** *es una tupla $Pg = (Pg_1, \ldots, Pg_n)$, donde Pg_i es el programa del agente i. Una interpretación π es* **compatible** *con Pg si π es compatible con cada Pg_i. De Pg y la interpretación π obtenemos el protocolo conjunto $Pg^\pi = (Pg_1^\pi, \ldots, Pg_n^\pi)$.*

Definición 19 *El sistema interpretado $\mathcal{I} = (R, \pi)$* **representa** *al programa conjunto Pg en el contexto interpretado (γ, π) si π es compatible con Pg y R representa al protocolo Pg^π en el contexto γ. Denotamos como $I^{rep}(Pg, \gamma, \pi)$ al sistema interpretado que representa a Pg.*

Se incorpora el conocimiento

La noción de programa estándar es bastante simple; los procesadores realizan acciones basadas en las pruebas que llevan a cabo sobre sus estados locales. Sin embargo los programas estándar no pueden ser utilizados para describir la relación entre conocimiento y acción que se quiere capturar. Extendamos la noción de programa estándar a *programas basados en conocimiento* en los cuales no sólo hay pruebas estándar que los procesadores pueden llevar a cabo sobre su estado local, sino que también hay pruebas de conocimiento que no van a depender solamente del estado local sino de todos los estados que los procesadores no distinguen del estado global real porque su estado local es el mismo en todos ellos.

La noción que presentamos es la que apareció en [21] y en [24]. En [24] se afirma que esta idea ya había aparecido de manera informal en [38], en [55] y en [75].

Formalmente:

Definición 20 *Un* **programa basado en conocimiento** Pg_i *para el procesador* i *tiene la forma*

case of
    if $t_1 \wedge k_1$ do a_1
    if $t_2 \wedge k_2$ do a_2
    ...
end case

Donde las t_j *son pruebas estándar que ya hemos definido y las* k_j *son* **pruebas de conocimiento** *para el procesador* j. *Las* a_j *son acciones del conjunto* ACT_j.

Una prueba de conocimiento para el procesador i es una combinación booleana de fórmulas de la forma $K_i\varphi$, donde φ es una fórmula en $\mathcal{L}_n$. Intuitivamente el procesador selecciona una acción basándose en el resultado de las pruebas estándar sobre su estado local y de las pruebas de conocimiento sobre todos los puntos en los que su estado local es el mismo.

Definición 21 *Un* **programa conjunto** *basado en conocimiento es una tupla* $\boldsymbol{Pg} = (\ \boldsymbol{Pg}_1\ ,\ldots,\ \boldsymbol{Pg}_n\ )$ *en la cual* $\boldsymbol{Pg}_i$ *es un programa basado en conocimiento para el procesador* i, $i = 1,\ldots,n$.

2.2.8 Sistemas, protocolos y programas

Hace falta definir la semántica de los programas basados en conocimiento. Así como pensamos que los programas estándar inducen un protocolo que determina las acciones de los procesadores, queremos también que los programas basados en conocimiento induzcan un protocolo. Sin embargo no es obvio cómo puede hacerse esto. Un protocolo es una función de estados locales a conjuntos de acciones; para ir de un programa estándar a un protocolo lo que tenemos que hacer es evaluar las pruebas estándar sobre un estado local; esto se hace utilizando los sistemas intepretados. En un programa basado en conocimiento tenemos que evaluar las pruebas de conocimiento; sin embargo estas pruebas dependen de todo el sistema interpretado y no de un único estado local. Pudiera ser que un procesador estuviera en el mismo estado local ℓ en dos sistemas interpretados distintos $\mathcal{I}_1$ y $\mathcal{I}_2$, que la prueba $K_i(p)$ resulte verdadera en el estado local ℓ en $\mathcal{I}_1$ y que resulte falsa en el estado local ℓ en $\mathcal{I}_2$.

Para resolver este problema lo que se hace es que dado un sistema interpretado $\mathcal{I} = (\mathcal{R}, \pi)$ y un programa conjunto basado en conocimiento $\boldsymbol{Pg} = (\ \boldsymbol{Pg}_1\ ,\ldots,\ \boldsymbol{Pg}_n\ )$, se asocia a $\boldsymbol{Pg}$ un protocolo conjunto que denotaremos $Pg^{\mathcal{I}} = (Pg_1^{\mathcal{I}},\ldots,Pg_n^{\mathcal{I}})$. Intuitivamente evaluamos las pruebas estándar en $\boldsymbol{Pg}$ de acuerdo a π y evaluamos las pruebas de conocimiento de acuerdo a $\mathcal{I}$. Como en el caso de los programas estándar, pediremos que la asignación π sea compatible con $\boldsymbol{Pg}$; esto es, que cada proposición que aparezca en una prueba estándar en $\boldsymbol{Pg}_i$ sea local a i. Este requisito de localidad sólo se pide para las pruebas estándar y no para las pruebas de conocimiento. Queremos definir $Pg_i^{\mathcal{I}}(\ell)$ para todos los estados locales ℓ del procesador i. Para hacer esto primero definimos lo que quiere decir que una prueba φ se satisface en el estado local ℓ con respecto al sistema interpretado $\mathcal{I}$.

Definición 22 *Sea φ una prueba, ℓ un estado local del procesador i, $\mathcal{I} = (\mathcal{R}, \pi)$ un sistema interpretado.*

Si φ es una prueba estándar entonces análogamente a lo definido para programas estándar.

$$(\mathcal{I}, \ell) \models \varphi \; syss \; (\pi, \ell) \models \varphi$$

Si φ es una prueba de conocimiento de la forma $K_i\psi$

$$(\mathcal{I}, \ell) \models K_i\psi \; syss \; (\mathcal{I}, r, t) \models \psi p.t \; (r, t) \; tq \; r_i(t) = \ell$$

Para conjunciones y negaciones de pruebas de conocimiento se sigue el procedimiento usual y los demás conectivos son definidos a partir de estos dos.

$$(\mathcal{I}, r, t) \models \neg K_i\varphi \; syss \; no \; (\mathcal{I}, r, t) \models K_i\varphi$$

$$(\mathcal{I}, r, t) \models K_i\varphi \wedge K_i\psi \; syss (\mathcal{I}, r, t) \models K_i\varphi \; y \; (\mathcal{I}, r, t) \models K_i\psi$$

Nótese que definimos $(\mathcal{I}, \ell)$ aunque ℓ sea un estado local que no ocurre en ningún punto de $\mathcal{I}$. En este caso es inmediato de las definiciones que se cumple $(\mathcal{I}, \ell) \models K_i(\bot)$, se cumple por vacuidad porque no hay algún punto en el sistema en el que el estado local del procesador i sea ℓ. Esto quiere decir que el axioma del conocimiento, $(K_i\varphi \rightarrow \varphi)$, o axioma T, falla. Sin embargo si ℓ es un estado que ocurre en $\mathcal{I}$ entonces K_i se comporta según los axiomas de $S5_n$; veamos esto formalmente.

Proposición 2 *Si $\ell = r_i(t)$ para algún punto (r, t) de $\mathcal{I}$, entonces $(\mathcal{I}, \ell) \models K_i\varphi$ syss $(\mathcal{I}, r, t) \models K_i\varphi$.*

Prueba

$(\mathcal{I}, \ell) \models K_i\varphi$ syss $(\mathcal{I}, r', t') \models \varphi$ para todo punto (r', t') de $\mathcal{I}$ tal que $r_i'(t') = \ell$ syss $(\mathcal{I}, r', t') \models \varphi$ para todo punto (r', t') de $\mathcal{I}$ tal que $(r, t) \sim_i r_i'(t')$ syss $(\mathcal{I}, r, t) \models K_i\varphi$.□

Podemos dar ahora la definición del protocolo asociado a un programa basado en conocimiento.

Definición 23

$$Pg_i^{\mathcal{I}}(\ell) = \begin{cases} \{a_j : (\mathcal{I},\ell) \models t_j \wedge k_j\} & si\ \{j : (\mathcal{I},\ell) \models t_j \wedge k_j\} \neq \emptyset \\ \{\Lambda\} & si\ \{j : (\mathcal{I},\ell) \models t_j \wedge k_j\} = \emptyset \end{cases}$$

Las acciones que el protocolo $Pg_i^{\mathcal{I}}$ define para el procesador i son precisamente las que define el programa $\boldsymbol{Pg}_i$ en el sistema interpretado $\mathcal{I}$.

Esta definición incluye a los programas estándar, ya que si $\boldsymbol{Pg}$ es un programa estándar, entonces $\boldsymbol{Pg}$ es también un programa basado en conocimiento, aunque uno sin pruebas de conocimiento. Si consideramos al sistema interpretado $\mathcal{I}$ hay dos manera de asociar un protocolo a $\boldsymbol{Pg}$. Podemos pensar en $\boldsymbol{Pg}$ como en un programa estándar y asociarle Pg^{π}, o podemos pensar que es un programa basado en conocimiento y asociarle el protocolo $Pg^{\mathcal{I}}$; de las definiciones se desprende que ambos protocolos son idénticos.

Este mapeo de programas basados en conocimiento a protocolos permite definir lo que quiere decir que un sistema interpretado represente a un programa basado en conocimiento en un contexto interpretado utilizando la definición correspondiente para protocolos.

Definición 24 *Un sistema interpretado* $\mathcal{I} = (\mathcal{R},\pi)$ **representa** *al programa basado en conocimiento* $\boldsymbol{Pg}$ *en el contexto* (γ,π) *si* π *es compatible con* $\boldsymbol{Pg}$ *y* $\mathcal{R}$ *representa a* $Pg^{\mathcal{I}}$ *en* γ.

Para verificar si un sistema interpretado $\mathcal{I}$ representa a un programa $\boldsymbol{Pg}$, debemos verificar si $\mathcal{I}$ representa al protocolo $Pg^{\mathcal{I}}$ obtenido evaluando las pruebas de conocimiento de $\boldsymbol{Pg}$ en el mismo sistema interpretado $\mathcal{I}$. Debido a la circularidad de esta definición, no necesariamente habrá un único sistema interpretado que represente a un programa basado en conocimiento; pudiera ser que haya más de uno o que no haya ninguno.

Es conveniente hablar de que un protocolo *implementa* a un programa basado en conocimiento. La primera definición usada en [21] decía que el protocolo P implementa al programa basado en conocimiento $\boldsymbol{Pg}$ en el contexto interpretado (γ,π), si $P = Pg^{I^{rep}(P,\gamma,\pi)}$. La idea básica atrás de esto puede verse como sigue. Para un protocolo P y un contexto (γ,π) sea el sistema interpretado $\mathcal{I}_P = I^{rep}(P,\gamma,\pi)$.

Supongamos que ejecutamos a P y que además $P = Pg^{\mathcal{I}_P}$. Por definición las acciones que dictamina el protocolo $Pg^{\mathcal{I}_P}$ son exactamente las que dictamina el programa $\boldsymbol{Pg}$ en el sistema $\mathcal{I}_P$. Como el sistema $\mathcal{I}_P$ representa al protocolo P, entonces al ejecutar P nos estamos adhiriendo a lo que dice el programa $\boldsymbol{Pg}$. Esto sugiere que si $P = Pg^{\mathcal{I}_P}$ podemos decir que el protocolo P implementa al programa $\boldsymbol{Pg}$. Sin embargo en [24] se hace ver que esta definición es demasiado restrictiva; la razón es que el protocolo $Pg_i^{\mathcal{I}_P}$ está definido en todos los estados locales en L_i, incluidos aquellos estados que no aparecen en el sistema $\mathcal{I}$. En estos estados el comportamiento de $Pg_i^{\mathcal{I}_P}$ es completamente arbitrario. Consideremos un protocolo P' que es igual a P en todos los estados que aparecen en I_P pero posiblemente se comporta distinto en los demás estados. Bajo la primera definición P implementa a $\boldsymbol{Pg}$ en (γ, π) pero P' no. En [24] se da una nueva definición de implementabilidad que modifica la anterior.

Definición 25 *Sea P un protocolo y sea el sistema $\mathcal{I} = I^{rep}(P, \gamma, \pi)$.* P **implementa** *al programa $\boldsymbol{Pg}$ en el contexto (γ, π) si (1) $\mathcal{I} = I^{rep}(Pg^{\mathcal{I}}, \gamma, \pi)$ y (2) P y $Pg^{\mathcal{I}}$ concuerdan en todos los estados globales que aparecen en I.*

Esta definición resuelve el problema anterior sobre los estados globales que no aparecen en $\mathcal{I}$. La parte (1) por si sola no es suficiente porque solamente habla de los sistemas que representan a los protocolos. No hace referencia directa a las acciones que realizan los agentes. Podría ocurrir, por ejemplo que haya dos acciones conjuntas distintas a y b tales que en un estado global determinado en $\mathcal{I}$ el efecto de la acción a es exactamente el mismo del de la acción b. Fijándonos solamente en el sistema interpretado $\mathcal{I}$ no podríamos saber cual fue la acción que se realizó. La idea de que el protocolo P implemente al programa $\boldsymbol{Pg}$ es que las acciones que determina P sean exactamente las mismas que determina $\boldsymbol{Pg}$. Esto es lo que se asegura con la parte (2).

Una consecuencia inmediata de esta definición es que si el sistema interpretado I representa al programa $\boldsymbol{Pg}$ en el contexto (γ, π), entonces el protocolo $Pg^{\mathcal{I}}$ implementa a $\boldsymbol{Pg}$ en (γ, π)

Veamos un ejemplo de un programa para el que hay más de un sistema interpretado que lo represente y un ejemplo de un programa para el que no hay ninguno. Estos ejemplos aparecieron como ejercicios en [21].

Ejemplo 1 Supongamos que tenemos un sistema con un sólo agente, el agente 1, que tiene en su estado local una variable x inicializada en 0. Supongamos que el agente ejecuta el siguiente programa basado en conocimiento NU (de No Único)

$$\textbf{if } K_1(t \neq 0 \rightarrow x = 1) \textbf{ do } x := 1$$

Las proposiciones primitivas son $(t \neq 0)$ y $(x = 1)$. Suponemos que el estado local del medio ambiente incluye al tiempo. π^{nu} es una interpretación tal que $\pi^{nu}(t,k)(t \neq 0) = verdadero$ syss $t > 0$ y $\pi^{nu}(t,k)(x = 1) = verdadero$ syss $k = 1$. La única acción para el agente 1 es $x := 1$ que tiene el efecto de cambiar el valor de x a 1.

Formalizaremos al sistema considerando al contexto γ^{nu} donde $\gamma^{nu} = (P_{ma}, \mathcal{G}_0, \tau, True)$, definido como sigue: el estado local del agente 1 es 0 ó 1, representando el valor de x; el estado del medio ambiente guarda el tiempo y es de la forma t, donde $t \geq 0$. Como el tiempo inicial es 0 y x inicialmente tiene valor 0 entonces $\mathcal{G}_0 = \{(0,0)\}$. El protocolo del medio ambiente siempre ejecutará la acción inc para actualizar el tiempo, por lo que $P_{ma}(t) = inc$. Las acciones del agente son Λ ó $(x := 1)$. El efecto de τ es el de asignar los valores apropiados a las variables, entonces $\tau(inc, \Lambda)(t,k) = (t+1, k)$, $\tau(inc, x := 1)(t,k) = (t+1, 1)$, donde $k \in \{0,1\}$.

Sea r^0 la ejecución que inicia en el estado $(0,0)$ y en la que el agente no hace nada, siempre realiza la acción Λ. Entonces $r^0(t) = (t,0)$ para toda $t \geq 0$. Sea r^j, para $j \geq 1$ la ejecución que inicia el el estado $(0,0)$ y en la que el agente 1 cambia el valor de x a 1 en la ronda j. Entonces $r^j(t) = (t,0)$ para toda $t < j$, $r^j(t) = (t,1)$ para toda $t \geq j$. Éstas son las únicas ejecuciones que podemos tener en el contexto γ^{nu}, porque para el medio ambiente no hay ninguna otra acción posible y el agente 1 en algún momento cambia el valor de la variable x y entonces el estado global ya no cambia o el agente 1 nunca cambia el valor de la variable.

Veamos que ninguna ejecución r^j, $j > 1$, puede estar en un sistema interpretado $\mathcal{I}$ consistente con NU. Si estuviera entonces como el agente 1 cambió el valor de x exactamente en la ronda j tuvo que ocurrir que $(\mathcal{I}, r^j, j-1) \models K_1(t \neq 0 \rightarrow x = 1)$. Entonces por el axioma de la introspección positiva $(\mathcal{I}, r^j, j-1) \models K_1 K_1(t \neq 0 \rightarrow x = 1)$. Como $r_1^j(0) = 0 = r_1^j(j-1)$ entonces $(r^j, 0) \sim_1 (r^j, j-1)$, entonces $(\mathcal{I}, r^j, 0) \models K_1(t \neq 0 \rightarrow x = 1)$. Pero como el sistema $\mathcal{I}$ es congruente con $NU^{\mathcal{I}}$ esto quiere decir que el agente 1 debió cambiar el valor de x desde la primera ronda de r^j, lo cual es una contradicción. Las ejecuciones r^0 y r^1 si pueden aparecer en el sistema interpretado. Entonces el conjunto de ejecuciones de cualquier sistema interpretado consistente con NU debe ser un subconjunto no vacío de $\{r^0, r^1\}$. Sea $\mathcal{R}^j$ el sistema que sólo consiste en la ejecución r^j, $j = 0, 1$, y sea $\mathcal{I}^j = (\mathcal{R}^j, \pi^{nu})$. Veamos que $\mathcal{I}^0$ y $\mathcal{I}^1$ representan a NU en el contexto interpretado (γ^{nu}, π^{nu}).

En cada punto (r^1, t) de $\mathcal{I}^1$, se cumple que si $\pi^{nu}(r^1(t))(t \neq 0) = \mathit{verdadero}$ entonces $\pi^{nu}(r^1(t))(x = 1) = \mathit{verdadero}$ y entonces se cumple $(\mathcal{I}^1, r^1, t) \models K_1(t \neq 0 \rightarrow x = 1)$ para toda t. Entonces la acción que siempre lleva a cabo el agente 1 es la de cambiar el valor de x a 1. Entonces la única ejecución posible que sea consistente con las acciones del agente 1 es precisamente r^1 por lo que $\mathcal{I}^1$ representa a NU en (γ^{nu}, π^{nu})

En cada punto en $\mathcal{I}^0$, es decir en cada punto de la ejecución r^0, el estado local del agente 1 es 0; entonces el agente no distingue entre ninguno de los puntos de la ejecución r^0. Para todo punto (r^0, t) tal que $t > 0$ siempre se cumple $\pi^{nu}(r^0, t)(t \neq 0) = \mathit{verdadero}$ y $\pi^{nu}(r^0, t)(x = 1) = \mathit{falso}$ y entonces para todo punto tal que $t > 0$ se cumple $(\mathcal{I}^0, r^0, t) \models \neg(t \neq 0 \rightarrow x = 1)$, por lo que para todo punto en $\mathcal{I}^0$ se cumple $(\mathcal{I}', r^0, t) \models \neg K_1(t \neq 0 \rightarrow x = 1)$, entonces el agente siempre ejecuta la acción nula Λ, resultando que la única ejecución consistente con las acciones del agente es r^0. Entonces $\mathcal{I}^0$ también representa a NU en (γ^{nu}, π^{nu})

La pregunta natural es si el sistema interpretado $\mathcal{I}^2 = (\{r^0, r^1\}, \pi^{nu})$ representa a NU en (γ^{nu}, π^{nu}). Este sistema interpretado no es consistente con $NU^{\mathcal{I}^2}$ en (γ^{nu}, π^{nu}). Esto es porque como $r_1^1(0) = 0 = r_1^0(t)$ para toda t, entonces $(r^1, 0) \sim_1 (r^0, t)$ para toda t. Se cumple que para toda $t > 0$ $(\mathcal{I}^2, r^0, t) \models \neg(t \neq 0 \rightarrow x = 1)$, entonces se cumple $(\mathcal{I}^2, r^1, 0) \models \neg K_1(t \neq 0 \rightarrow x = 1)$, entonces en la primera ronda de

r^1 el agente no cambia el valor de la variable x y entonces r^1 no es consistente con las acciones dictaminadas por $NU^{\mathcal{I}^2}$ en (γ^{nu}, π^{nu}).

Entonces hay exactamente dos sistemas interpretados distintos que representan a NU en el contexto (γ^{nu}, π^{nu}).□

Ejemplo 2 Tomemos exactamente el mismo contexto interpretado del ejemplo pasado (γ^{nu}, π^{nu}). Supongamos que ahora nuestro agente ejecuta el siguiente programa basado en conocimiento NR (de No Representa):

$$\textbf{if } \neg K_1(t \neq 0 \rightarrow x = 1) \textbf{ do } x := 1$$

Igual que en el caso anterior las únicas ejecuciones que pudieran aparecer son las de la forma r^j, pero veamos que ninguna r^j con $j > 1$ puede estar en un sistema interpretado $\mathcal{I}$ consistente con NR. Si estuviera, entonces tuvo que ocurrir que $(\mathcal{I}, r^j, j-1) \models \neg K_1(t \neq 0 \rightarrow x = 1)$ y entonces como $(r^j, 0) \sim_1 (r^j, j-1)$ debe ocurrir $(\mathcal{I}, r^j, 0) \models \neg K_1(t \neq 0 \rightarrow x = 1)$, pero entonces el agente debió cambiar el valor de x desde la primera ronda lo cual es una contradicción. Entonces los posibles sistemas interpretados que representen a NR son $\mathcal{I}^0$, $\mathcal{I}^1$ y $\mathcal{I}^2$. Pero veamos que ninguno de estos sistemas es consistente con NR.

En $\mathcal{I}^0$ se cumple $(\mathcal{I}^0, r^0, t) \models \neg(t \neq 0 \rightarrow x = 1)$ para toda $t > 0$ y como $(r^0, t) \sim_1 (r^0, 0)$ para toda t entonces se cumple $(\mathcal{I}^0, r^0, 0) \models \neg K_1(t \neq 0 \rightarrow x = 1)$, por lo que el agente debió haber cambiado el valor de x en la primera ronda; pero no lo hizo, por lo que r^0 no es consistente con $NU^{\mathcal{I}^0}$

En $\mathcal{I}^1$ ocurre que $(r^1, t) \not\sim_1 (r^1, 0)$ para toda $t > 0$ y como se cumple $(\mathcal{I}^1, r^1, 0) \models \neg(t \neq 0)$ entonces se cumple $(\mathcal{I}^1, r^1, 0) \models (t \neq 0 \rightarrow x = 1)$ y entonces se cumple $(\mathcal{I}^1, r^1, 0) \models K_1(t \neq 0 \rightarrow x = 1)$; entonces el agente 1 debe realizar la acción nula Λ en la primera ronda, pero lo que hace es cambiar el valor de la variable x, por lo que r^1 no es consistente con $NU^{\mathcal{I}^1}$.

En $\mathcal{I}^2$ se cumple $(\mathcal{I}^2, r^0, t) \models \neg(t \neq 0 \rightarrow x = 1)$ para toda $t > 0$ y entonces $(\mathcal{I}^2, r^0, 0) \models \neg K_1(t \neq 0 \rightarrow x = 1)$; por lo que el agente debió haber cambiado el valor de x en la primera ronda de r^0; pero no lo hizo, por lo que r^0 no es consistente con $NU^{\mathcal{I}^2}$ y entonces $\mathcal{I}^2$ no representa a NR. Esto quiere decir que no hay algún sistema interpretado que represente a NR en el contexto (γ^{nu}, π^{nu}).□

Con el ejemplo 1 podemos aclarar mejor lo que hablábamos en la sección anterior al respecto de que un protocolo implemente a un programa basado en conocimiento. En el sistema interpretado $\mathcal{I}^1$, como ya vimos, ocurre que la acción que siempre lleva a cabo el agente 1 es la de cambiar el valor de x a 1; esto genera la ejecución r^1. Pero supongamos un protocolo conjunto P tal que el protocolo del medio ambiente siempre dictamine la acción *inc* y para el agente 1 $P_1(0) = (x := 1)$ y $P_1(1) = \Lambda$. La única ejecución consistente con el protocolo P en el contexto γ^{nu} es r^1; entonces si consideramos al sistema $\mathcal{I} = (\mathbf{R}^{rep}(P, \gamma^{nu}), \pi^{nu})$ ocurre que $\mathcal{I} = \mathcal{I}^1$; sin embargo no podemos decir que el protocolo P implemente al programa NU porque el protocolo P no concuerda con el protocolo $NU^{\mathcal{I}^1}$ ya que las acciones que prescriben son diferentes, aunque tengan el mismo efecto sobre los estados globales de r^1.

Cuando no hay algún sistema que represente a un programa, entonces el programa está mal diseñado y no tiene ningún interés; sin embargo si hay más de un sistema que represente esto no es necesariamente un problema, podemos pensar que el programa es una especificación de alto nivel que es satisfecha por muchos sistemas interpretados. Cuando hablamos de los programas estándar vimos que siempre hay un único sistema que los representa en un contexto interpretado, entonces los programas estándar describen completamente el comportamiento de los agentes o procesadores; esto no siempre ocurre con los programas basados en conocimiento.

En otras situaciones, sin embargo, hay una fuerte intuición de que un programa basado en conocimiento describe completamente el comportamiento de los procesadores y en consecuencia debe haber un único sistema que represente al programa. Intuitivamente esto ocurre cuando fijamos el conjunto de estados iniciales y a partir de ahí podemos ir generando todas las ejecuciones paso a paso. Formalizemos estas ideas.

Definición 26 *Sea $\mathcal{G}$ un conjunto de estados globales y r una ejecución sobre $\mathcal{G}$. El* **prefijo** *de r al tiempo t o* **t-prefijo**, *$Pref_t(r)$ es la sucesión de los primeros $t+1$ estados globales en r, es decir $Pref_t(r) = (r(0), r(1), \ldots, r(t))$. Si $\mathcal{R}$ es un conjunto de ejecuciones entonces $Pref_t(\mathcal{R}) = \{Pref_t(r) | r \in \mathcal{R}\}$. Si $\mathcal{I} = (\mathcal{R}, \pi)$ es un sistema interpretado, $Pref_t(\mathcal{I}) = (Pref_t(\mathcal{R}), \pi)$*

Supongamos que en cada paso hemos generado todos los t-prefijos de un sistema. Si en cualquier punto del sistema (r, t) un agente es capaz de evaluar todas las pruebas de conocimiento en el programa, entonces podemos determinar en cada punto qué acciones van a tomar todos los agentes y podremos generar todos los $(t + 1)$-prefijos. Esto ocurre cuando estando en un punto (r, t) todos los puntos (r', t') que un agente no distingue cumplen con $t' \leq t$; entonces las pruebas de conocimiento siempre se van a realizar sobre puntos que ya hemos construido.

Capítulo 3

El modelo del acertijo de los sabios

Pensemos en el acertijo general: el rey se reúne con los n sabios y lleva una caja de colores; supongamos que lleva un cierto número de etiquetas de cada color y quiere asignar un color a cada sabio, así que al menos lleva n etiquetas en total. El rey coloca a los sabios de manera que todos pueden ver a todos. Asumiremos que esto es conocimiento común y también es conocimiento común que todos los sabios pueden oir lo que dicen los demás y lo que dice el rey.

En el acertijo lo primero que ocurre es que el rey escoge una caja de colores; luego el rey enseña la caja a los sabios y asigna un color a cada sabio. Entonces el rey va escogiendo por rondas a un conjunto de los sabios a los que pregunta si saben su color y espera la respuesta de todos para hacer la pregunta de la ronda siguiente.

Consideraremos dos versiones del acertijo: la versión de las esposas infieles en la que en cada ronda el rey pregunta a todos los sabios; y la versión de los tres sabios, en la que en el rey va preguntando a los sabios uno por uno.

Modelaremos al acertijo en un sistema distribuido en donde los sabios y el rey son los procesadores del sistema. Las preguntas del rey y las respuestas de los sabios se llevarán a cabo mediante intercambio de mensajes. Cada procesador tiene un conjunto de líneas de comunicación por el que manda sus mensajes y otro por el que los recibe.

Usaremos el modelo que presentamos en la sección 2.2. Asumimos que el sistema es un sistema de mensajes síncrono. Definiremos un protocolo basado en conocimiento para cada sabio, para el rey y también definiremos un contexto interpretado. El acertijo de los niños enlodados y el acertijo de las esposas infieles han sido utilizados en la literatura (ver [21, 24, 39]) como ejemplos de los programas basados en conocimiento; retomamos algunas de las ideas desarrolladas en esos artículos.

La comunicación en el sistema se lleva a cabo en rondas. En cada ronda cada procesador envía un conjunto (posiblemente vacío) de mensajes y recibe todos los mensajes que le fueron enviados en esa misma ronda y posiblemente realiza algún conjunto de acciones internas. En un sistema de mensajes podemos ver al medio ambiente como el que decide qué mensajes llegan a su destino. En este caso el medio ambiente dejará que todos los mensajes lleguen a su destino y que de hecho lo hagan en la misma ronda en que fueron enviados.

Hablaremos del sistema de mensajes A para un conjunto de $n+1$ procesadores $\mathcal{P} = \{rey, s_1, \ldots, s_n\}$. Para referirnos al procesador s_i hablaremos del sabio i y en general usaremos las letras i, j; para referirnos al procesador *rey* hablaremos del rey. Y cuando digamos procesador es que nos estamos refiriendo a cualquiera de ellos (alguno de los sabios o el rey) y usaremos las letras p, q.

Modelaremos el acertijo de manera que en los estados locales iniciales ya se haya escogido una caja y el rey ya haya escogido una configuración inicial.

3.1 Configuraciones y visiones

Consideremos un conjunto $\mathcal{S} = \{s_1, \ldots, s_n\}$ de n sabios. Numeremos los colores de 1 a r. Definimos un conjunto $\mathcal{C}_n$ de cajas de colores para n sabios.

Definición 27 *$C \in \mathcal{C}_n$ si $C = (\sharp c_1, \sharp c_2, \ldots, \sharp c_r)$ y $\sum_{i=1}^{r} \sharp c_i = n + k = Tot(C)$, $k \geq 0$. Supondremos además que los colores están ordenados de mayor a menor, esto es que $\sharp c_1 \geq \sharp c_2 \geq \ldots \geq \sharp c_r$.*

Estamos suponiendo que hay r colores en la caja C y que hay $\sharp c_a$ etiquetas de cada color. Es decir que en la caja C el rey puede asignar el color a a lo más a $\sharp c_a$ sabios.

Para representar cuando el rey asigna los colores utilizaremos las **configuraciones.**

Definición 28 *Una* **configuración** α *es un conjunto de n parejas*

$$\alpha = \{(s_1, c_1), (s_2, c_2), \ldots, (s_n, c_n)\}$$

donde c_i *es el color del sabio* s_i, $1 \leq i \leq n$, $1 \leq c_i \leq r$

Dada una caja C, *una* **configuración posible** *es una configuración que cumple con ser una asignación posible de colores en* C, *esto es que para todo color* a, $1 \leq a \leq r$

$$|\{(s_j, c_j) \in \alpha : c_j = a\}| \leq \sharp c_a$$

Hacemos notar que $\sharp c_a$ se refiere al número de objetos o etiquetas o copias de color a que el rey trae en la caja. En las configuraciones la pareja (s_i, c_i) indica que el sabio i tiene color c_i donde c_i puede ser alguno de los colores entre 1 y r, es por esto que decimos que $1 \leq c_i \leq r$.

La caja junto con el conjunto S de sabios definen al conjunto de todas las configuraciones posibles. Llamaremos $\mathcal{CP}_C$ a este conjunto.

En el caso que vimos al principio había tres sabios y una caja $C = (3, 2)$. Una configuración es

$$\alpha = \{(s_1, 2), (s_2, 2), (s_3, 2)\}$$

Sin embargo α no es una configuración posible porque asigna a tres sabios el color 2. En cambio la siguiente configuración sí es una configuración posible:

$$\beta = \{(s_1, 1), (s_2, 2), (s_3, 2)\}$$

Mientras que una configuración es una asignación de los colores, una visión es lo que ve un sabio en una de las configuraciónes.

Definición 29 *Para una configuración* $\alpha = \{(s_1, c_1), (s_2, c_2), \ldots, (s_n, c_n)\}$ *La visión* $vision_i(\alpha)$ *del sabio* i *en la configuración* α *es el conjunto*

$$\{(s_1, c_1), \ldots, (s_{i-1}, c_{i-1}), (s_i, *), (s_{i+1}, c_{i+1}), \ldots, (s_n, c_n)\}$$

En el que se representa que el sabio i *ve el color de todos los sabios menos el suyo propio.*

Denotaremos al número de sabios con color a en la configuración α como p_a^α. También denotaremos f_a^α como el número de etiquetas de color a que el rey no utilizó. Así que tenemos las siguientes definiciones:

Definición 30

$$p_a^\alpha = |\{(s_j, c_j) \in \alpha : c_j = a\}|$$

$$f_a^\alpha = \sharp c_a - p_a^\alpha$$

Hacemos notar que como $\sum_{a=1}^{r} \sharp c_a = n + k$ y $\sum_{i=1}^{r} p_i^\alpha = n$. Es decir que en total hay $n + k$ etiquetas y el rey asigna n. Entonces $\sum_{i=1}^{r} f_i^\alpha = \sum_{a=1}^{r} \sharp c_a - \sum_{i=1}^{r} p_i^\alpha = k$. Es decir que el rey deja de asignar k etiquetas.

3.2 Características del sistema

En el sistema habrá tres mensajes posibles: cuando el rey le pregunta a alguno de los sabios si sabe su color, cuando un sabio responde que no sabe y cuando un sabio responde que sí sabe.

Sea MSG un conjunto de mensajes de la forma $\langle rey \rangle, \langle no, i \rangle, \langle a, i \rangle$. El mensaje $\langle rey \rangle$ representa la pregunta del rey a algún sabio; $\langle no, i \rangle$ representa la respuesta negativa del sabio i; $\langle a, i \rangle$ representa la respuesta en la que el sabio i indica que sabe que su color es a.

En el acertijo nos interesa que cada sabio recuerde las respuestas de otros sabios que haya escuchado. En general en un sistema de mensajes es importante que en los estados locales de los procesadores y en el del medio ambiente se guarde una historia de mensajes al tiempo t.

Definición 31 *$hist_{\ell_p}(t)$ es una* **historia de mensajes** *en el estado local ℓ_p del procesador p al tiempo t tal que*

$$hist_{\ell_p}(0) = \emptyset$$

para $t > 0$

$$hist_{\ell_p}(t) = (M^1_{\ell_p}, M^2_{\ell_p}, \ldots, M^t_{\ell_p})$$

donde

$$M^k_{\ell_p} = \{ev_p | ev_p \in \{enviar_p(msg, q), recibir_p(msg, q)\}\} \text{ o } M^k_{\ell_p} = \emptyset$$

Definiremos ahora al estado local de los procesadores y del medio ambiente. Incluimos al tiempo en su estado local y para efectos de este acertijo no necesitamos que haya más información porque el medio ambiente no "tomará" ninguna decisión; en todas las rondas permitirá que todos los mensajes enviados lleguen a su destino. Para los procesadores incluiremos el tiempo y la historia de mensajes ya que estamos modelando un sistema de mensajes síncrono. También incluiremos la caja de colores. En el estado local del rey pondremos la configuración que ha escogido. En el de los sabios pondremos a su visión en vez de esta configuración.

Definición 32 estados locales y estados globales

$C \in \mathcal{C}_n$, $\alpha \in \mathcal{CP}_C$.

ℓ_{ma} *es el* **estado local del medio ambiente**

$$\ell_{ma} = (t)$$

ℓ_{rey} *es el* **estado local del rey**

$$\ell_{rey} = (t, C, \alpha, hist_{\ell_{rey}}(t))$$

ℓ_i *es el* **estado local del sabio** i

$$\ell_i = (t, C, vision_i(\alpha), hist_{\ell_i}(t))$$

Un **estado global** *es de la forma*

$$g = (\ell_{ma}, \ell_{rey}, \ell_1, \ldots, \ell_n)$$

Se define a $\mathcal{G}$ como el conjunto de todos los estados globales posibles.

Definamos al conjunto de los estados globales iniciales.

Definición 33 *$\mathcal{G}_0$ es el conjunto de* **estados globales iniciales,**

$$(\ell_{ma}, \ell_{rey}, \ell_1, \ldots, \ell_n) \in \mathcal{G}_0$$

donde para $C \in \mathcal{C}_n$, $\alpha \in \mathcal{CP}_C$

$$\ell_{ma} = (0)$$

$$\ell_{rey} = (0, C, \alpha, \emptyset)$$

$$\ell_i = (0, C, vision_i(\alpha), \emptyset)$$

Continuamos con la definición de las acciones que pueden llevar a cabo los procesadores y el medio ambiente.

Aquí cabe hacer la aclaración de que, como estamos pensando en un sistema de mensajes en cada ronda, las acciones de un procesador van a ser conjuntos de acciones debido a que en cada ronda cada procesador manda un conjunto de mensajes y posiblemente realiza algunas acciones locales.

ACT_{ma} es el conjunto de acciones que puede realizar el medio ambiente. ACT_{rey} es el conjunto de acciones que puede realizar el rey. ACT_i es el conjunto de acciones que puede realizar el sabio i.

Las acciones del medio ambiente son conjuntos de elementos del tipo $entregar_{ma}(p, q)$; esto representa que el medio ambiente entrega al procesador p cualquier mensaje que le haya enviado el procesador q.

Las acciones que realiza el procesador p son conjuntos de elementos del tipo $enviar_p(msg, q)$, representando que el procesador p envia el mensaje msg al procesador q. Una acción que el rey realiza puede incluir también la acción interna $escoge_{rey}(G)$ que significa que el rey escoge el conjunto G de sabios para hacerles la siguiente pregunta.

Si en una ronda un procesador o el medio ambiente no realiza ninguna acción, lo representaremos con la acción nula Λ. Para términos prácticos asumiremos que $\Lambda = \emptyset$.

Definición 34 Acciones

$$a_{ma} = \{e_{ma} | e_{ma} = entregar_{ma}(p, q)\}$$

$$a_{rey} = \{e_{rey} | e_{rey} = enviar_{rey}(\langle rey \rangle, i) \ o \ e_{rey} = escoge_{rey}(G);\ G \subseteq \mathcal{S}\}$$

$$a_i = \{e_i | e_i = enviar_i(msg, q)\}$$

$$ACT = ACT_{ma} \times ACT_{rey} \times ACT_1 \times \ldots \times ACT_n$$

Para facilitar la notación definimos algunos conjuntos de acciones.

Definición 35

$$entregar_{ma}(todo) = \{entregar_{ma}(p, q)\ p.t.\ p, q\ tal\ que\ p \neq q\}$$

$$enviar_{rey}(msg, G) = \{escoger_{rey}(G)\} \cup \{enviar_p(\langle rey \rangle, q)\ p.t.\ q \in G\}$$

$$enviar_i(msg, G) = \{enviar_i(msg, q)\ p.t.\ q \in G\}$$

Definamos entonces cómo influyen las acciones de los procesadores en el estado global del sistema. La interpretación será de la manera natural; cuando un procesador envía un mensaje su estado local debe registrar este envío en su historia de mensajes y cuando un procesador recibe un mensaje su estado local debe incluir el evento de la recepción del mensaje en su historia de mensajes.

Definición 36 Transformador de estados globales y función de transición

Sea $a = (a_{ma}, a_{rey}, a_1, \ldots, a_n)$ *una acción conjunta.*

$\mathcal{TR}_A$ *es el conjunto de todos los transformadores de estados globales.*

$$\mathcal{TR}_A = \{T : \mathcal{G} \mapsto \mathcal{G}\}$$

τ_A *es la función de transición.*

$$\tau_A : ACT \mapsto \mathcal{TR}_A$$

Sea $hist_{\ell_p}(t) = (M^1_{\ell_p}, M^2_{\ell_p}, \ldots, M^t_{\ell_p})$
Para $t \geq 0$. $C \in \mathcal{C}_n$, $\alpha \in \mathcal{CP}_C$, *sea* $g = (\ell_{ma}, \ell_{rey}, \ell_1, \ldots, \ell_n)$ *tal que*

$$\ell_{ma} = (t)$$

$$\ell_{rey} = (t, C, \alpha, hist_{\ell_{rey}}(t))$$

$$\ell_i = (t, C, vision_i(\alpha), hist_{\ell_i}(t))$$

Sea $g' = (\ell'_{ma}, \ell'_{rey}, \ell'_1, \ldots, \ell'_n,)$ *tal que*

$$\tau_A(a)(g) = g'$$

donde

$$\ell'_{ma} = (t+1)$$

$$\ell'_{rey} = (t+1, C, (\alpha), hist_{\ell_{rey}}(t+1))$$

donde

$$hist_{rey}(t+1) = (M^1_{\ell_{rey}}, M^2_{\ell_{rey}}, \ldots, M^t_{\ell_{rey}}, M^{t+1}_{\ell_{rey}})$$

$$M^{t+1}_{\ell_{rey}} = a_{rey} \cup \{recibir_{rey}(msg, j) | enviar_j(msg, rey) \in a_j \ y$$

$$entregar_{ma}(rey, j) \in a_{ma}\}$$

$$\ell'_i = (t+1, C, vision_i(\alpha), hist_i(t+1)$$

donde

$$hist_{\ell_i}(t+1) = (M^1_{\ell_i}, M^2_{\ell_i}, \ldots, M^t_{\ell_i}, M^{t+1}_{\ell_i})$$

$$M^{t+1}_{\ell_i} = a_i \cup \{recibir_i(msg, q) | enviar_q(msg, i) \in a_q \ y$$

$$entregar_{ma}(i, q) \in a_{ma}\}$$

Habíamos definido que un contexto está formado por el protocolo del medio ambiente; el conjunto de estados globales iniciales, la función de transición y posiblemente alguna restricción sobre los puntos del sistema. En este caso el protocolo del medio ambiente consistirá en escoger siempre la acción de entregar todos los mensajes que sean enviados.

Ya definimos a la función de transición τ_A y por el momento no pondremos ninguna restricción al conjunto de ejecuciones, esto es $\Psi = TRUE$.

Definamos al protocolo del medio ambiente P_{ma}. Es una función del conjunto de estados locales del medio ambiente al conjunto de acciones ACT_{ma}.

Definición 37 *P_{ma} es el* **protocolo del medio ambiente**

$$P_{ma} : L_{ma} \mapsto ACT_{ma}$$

Donde para todo estado $l_{ma} \in L_{ma}$ definimos

$$P_{ma}(l_{ma}) = entregar_{ma}(todo)$$

Ya podemos definir al contexto γ_A para el acertijo de los sabios

Definición 38 *El contexto γ_A es*

$$\gamma_A = (P_{ma}, \mathcal{G}_0, \tau_A, \Psi)$$

Hacemos notar que este contexto captura todas las posibles ejecuciones del acertijo en el que en cada ronda el rey escoge un conjunto de sabios al que le pregunta. Más adelante nos interesará estudiar la versión del acertijo de los tres sabios y la del acertijo de las esposas infieles para lo que definiremos dos nuevos contextos cambiando a la reestricción Ψ. En un contexto sólo permitiremos ejecuciones en las que el rey en cada ronda le pregunta a todos los sabios y en el otro sólo permitiremos ejecuciones en las que va preguntando uno por uno a los sabios.

3.3 Proposiciones primitivas

Utilizando la sintaxis del lenguaje de programación dada en la sección 2.2, definiremos los programas que ejecutan los sabios y el rey.

Primero debemos definir el conjunto de proposiciones primitivas que serán incluidas en las pruebas estándar y en las pruebas basadas en conocimiento; también definiremos la interpretación π_A que nos dirá como evaluar las pruebas.

Definición 39 *El conjunto Φ_A es el conjunto de proposiciones primitivas donde*

$$\begin{aligned}\Phi_A \;=\; & \{inicial, oyorespuesta, oyoalrey_i, \\ & (c_i = 1), \ldots, (c_i = r)\} \; donde \; i = 1, \ldots, n\end{aligned}$$

Debemos dar la interpretación de verdad de estas proposiciones primitiva; intuitivamente *inicial* es verdadera si el estado local del rey corresponde a su estado local en un estado global inicial; *oyorespuesta* es verdadera si el rey oyó alguna respuesta de un sabio en la ronda anterior; $oyoalrey_i$ es verdadera si el rey le pregunta al sabio i si sabe su color, esto es si el rey le mandó un mensaje en la ronda anterior; $(c_i = a)$ es verdadera si el sabio i tiene color a.

Definamos formalmente en cada estado global g la asignación de verdad $\pi_A(g)$ para las proposiciones primitivas.

Definición 40 *Sea $g = (\ell_{ma}, \ell_{rey}, \ell_1, \ldots, \ell_n)$ un estado global.*

$\pi_A(g)(inicial) = verdadero \; syss$

$$\ell_{rey} = (0, C, \alpha, \emptyset)$$

$\pi_A(g)(oyorespuesta) = verdadero \; syss$

$$\ell_{rey} = (t, C, \alpha, hist_{\ell_{rey}}(t)) \; y \; recibir_{rey}(msg, i)) \in M^{t-1}_{\ell_{rey}} \; p.a \; i$$

$\pi_A(g)(oyoalrey_i) = verdadero \; syss$

$$\ell_i = (t, C, vision_i(\alpha), hist_{\ell_i}(t)) \; y$$

$$recibir_i(\langle rey\rangle, rey)) \in M^{t-1}_{\ell_i}$$

$\pi_A(g)((c_i = a)) = verdadero\ syss$

$$\ell_i = (t, C, vision_i(\alpha), hist_{\ell_{rey}}(t))\ y\ (s_i, a) \in \alpha$$

Hay que notar que las proposiciones primitivas $(c_i = a)$ no son locales para el sabio i. Esto no va a representar un problema porque estas proposiciones sólo van a aparecer en las pruebas de conocimiento del programa del sabio i. Veamos formalmente que las demás proposiciones primitivas son locales para algún procesador.

Lema 2 *Las proposiciones primitivas inicial y oyorespuesta son locales para el procesador rey; oyoalrey$_i$ es local para el sabio i.*

Prueba

Sean dos estados g, g' en $\mathcal{G}$ tales que $g \sim_{rey} g'$, entonces $g_{rey} = g'_{rey}$.

(a) $\pi_A(g)(inicial) = verdadero$ syss $g_{rey} = (0, C, \alpha, \emptyset)$ syss $g'_{rey} = (0, C, \alpha, \emptyset)$ syss $\pi_A(g')(inicial) = verdadero$. Entonces $\pi_A(g)(inicial) = \pi_A(g')(inicial)$ y por lo tanto *inicial* es local para el *rey*.

(b) $\pi_A(g)(oyorespuesta) = verdadero$ syss $g_{rey} = (t, C, \alpha, hist_{g_{rey}}(t))$ y $recibir_{rey}(msg, i)) \in M^{t-1}_{g_{rey}}$ syss $g'_{rey} = (t, C, \alpha, hist_{g'_{rey}}(t))$ y $recibir_{rey}(msg, i)) \in M^{t-1}_{g'_{rey}}$ syss $\pi_A(g')(oyorespuesta) = verdadero$. Entonces $\pi_A(g)(oyorespuesta) = \pi_A(g')(oyorespuesta)$ y por lo tanto *oyorespuesta* es local para el *rey*.

Sean dos estados g, g' en $\mathcal{G}$ tales que $g \sim_i g'$, entonces $g_i = g'_i$.

(c) $\pi_A(g)(oyoalrey_i) = verdadero$ syss $g_i = (t, C, vision_i(\alpha), hist_{g_i}(t))$ y $recibir_i(\langle rey\rangle, rey)) \in M^{t-1}_{g_i}$ syss $g'_i = (t, C, vision_i(\alpha), hist_{g'_i}(t))$ y $recibir_i(\langle rey\rangle, rey)) \in M^{t-1}_{g'_{rey}}$ syss $\pi_A(g')(oyoalrey_i) = verdadero$. Entonces $\pi_A(g)(oyoalrey_i) = \pi_A(g')(oyoalrey_i)$ y por lo tanto $oyoalrey_i$ es local para el sabio i.□

3.4 Los programas

El rey ejecutará un programa estándar en el que en cada ronda decidirá a qué conjunto de los sabios les pregunta si saben su color.

El rey ejecuta el programa estándar $\boldsymbol{Pg}_{rey}$; $G \subseteq \mathcal{S}$ es el subconjunto de los sabios que el rey escoge con la acción interna $escoge_{rey}$. Pensaremos que en cada ronda el rey escoge al conjunto G de manera aleatoria.

Definición 41 *El programa* $\boldsymbol{Pg}_{rey}$ *es*

case of
  if *inicial* do $enviar_{rey}(\langle rey \rangle, G)$
  if *oyorespuesta* do $enviar_{rey}(\langle rey \rangle, G)$
end case

Recordamos que aunque el programa $\boldsymbol{Pg}_{rey}$ es un programa estándar también puede ser visto como un programa basado en conocimiento. Como demostramos en el lema 2 todas las proposiciones primitivas que aparecen en las pruebas estándar son locales para el rey, entonces π_A es compatible con $\boldsymbol{Pg}_{rey}$.

Los sabios deciden sus acciones con base en lo que saben de su color; si el sabio i sabe que su color es a y el rey le pregunta entonces responderá "sí, mi color es a" . Si no sabe responderá "no". Entonces es adecuado modelar el comportamiento del sabio i con un programa basado en conocimiento.

El sabio i ejecutará el programa basado en conocimiento $\boldsymbol{Pg}_i$ con el que decidirá que acciones tomar. Para facilidad de notación definimos al conjunto $R_i = \mathcal{P} \setminus \{i\}$ como el conjunto de todos los procesadores exceptuando al sabio i. Esto para representar que las respuestas del sabio i tienen que llegar a los demás sabios y al rey.

Definición 42 *El programa* $\boldsymbol{Pg}_i$*es*

case of
  if $oyoalrey_i \wedge K_i(c_i = 1)$ do
    $enviar_i(\langle 1, i \rangle, R_i)$
  ...

if $oyoalrey_i \wedge K_i(c_i = r)$ **do**
$enviar_i(\langle r, i\rangle, R_i)$
if $oyoalrey_i \wedge \neg K_i(c_i = 1) \wedge \ldots \wedge \neg K_i(c_i = r)$ **do**
$enviar_i(\langle no, i\rangle, R_i)$
end case

En el lema 2 demostramos que todas las proposiciones primitivas que aparecen en las pruebas estándar son locales para el sabio i, entonces π_A es compatible con Pg_i.

Una vez definidos los programas del rey y de los sabios definimos al programa conjunto $Pg_A = (\ Pg_{rey}, Pg_1\ , \ldots,\ Pg_n\ )$. En el programa conjunto Pg_A evaluaremos las pruebas estándar utilizando la asignación de verdad π_A. Esto puede hacerse porque π_A es compatible con Pg_A.

3.5 El sistema y los protocolos

Los programas que hemos dado dictan el comportamiento del rey y de los sabios para cualquier estado local en el que se encuentren. Lo que tienen que hacer es realizar las pruebas estándar y las pruebas de conocimiento para entonces elegir alguna de las acciones que correspondan a las pruebas que hayan sido satisfechas. Sin embargo esto es sólo una especificación formal de las ideas que ya sabíamos sobre el comportamiento de los participantes en el acertijo. Lo que nos interesa es analizar por medio del conocimiento qué es lo que ocurre en el acertijo. En particular nos interesa caracterizar cuándo un sabio sabe su color y responder a preguntas como cuántos sabios llegan a saberlo. También nos preguntaremos qué ocurre con las cajas de colores que lleva el rey. ¿Habrá alguna relación entre la caja y las respuestas de los sabios? ¿Para alguna caja podrá lograr el rey que ningún sabio sepa su color?

Hacen falta todavía dos cosas; por un lado tenemos que decir cuál es el protocolo que ejecutará el rey y cuál es el que ejecutará cada sabio; por otro lado debemos decir cuál es el sistema del acertijo, el conjunto de ejecuciones que nos interesan. Los protocolos serán aquellos asociados a los programas. Una vez que tengamos definidos los protocolos nos interesará el conjunto de todas las ejecuciones posibles de estos

protocolos, esto es, el sistema interpretado que representa al programa conjunto $\boldsymbol{Pg}_A$ en el contexto γ_A.

Definamos primero a los protocolos de los procesadores. Sea $\mathcal{I} = (\mathcal{R}, \pi_A)$ un sistema interpretado. Veamos que la interpretación π_A es compatible con el programa conjunto $\boldsymbol{Pg}_A$ y entonces es correcto pensar en los protocolos.

Lema 3 *π_A es compatible con $\boldsymbol{Pg}_A$*

Prueba

π_A es compatible con $\boldsymbol{Pg}_A$ si lo es con $\boldsymbol{Pg}_{rey}$, $\boldsymbol{Pg}_1$, ..., $\boldsymbol{Pg}_n$; esto se cumple si cada proposición primitiva que aparece en las pruebas estándar del programa basado en conocimiento $\boldsymbol{Pg}_p$ es local para el procesador p. Esto ya lo demostramos en el lema 2.□

Definamos ahora a los protocolos del rey y de los sabios asociados al programa conjunto $\boldsymbol{Pg}_A$.

Definición 43 *El protocolo del rey es*

$$Pg^{\mathcal{I}}_{rey}(\ell_{rey}) = \begin{cases} enviar_{rey}(\langle rey \rangle, G) \\ \qquad si\ (\mathcal{I}, \ell_{rey}) \models inicial \vee oyorespuesta \\ \Lambda\ en\ otro\ caso \end{cases}$$

El protocolo del sabio i es

$$Pg^{\mathcal{I}}_i(\ell_i) = \begin{cases} enviar_i(\langle 1, i \rangle, R_i)\ si\ (\mathcal{I}, \ell_i) \models oyoalrey_i \wedge K_i(c_i = 1) \\ \ldots \\ enviar_i(\langle r, i \rangle, R_i)\ si\ (\mathcal{I}, \ell_i) \models oyoalrey_i \wedge K_i(c_i = r) \\ enviar_i(\langle no, i \rangle, R_i)\ si \\ \qquad (\mathcal{I}, \ell_i) \models oyoalrey_i \wedge \\ \qquad\qquad \neg K_i(c_i = 1) \wedge \ldots \wedge \neg K_i(c_i = r) \\ \Lambda\ en\ otro\ caso \end{cases}$$

El protocolo del rey es un protocolo no determinístico debido a que el conjunto G de los sabios a los que va a preguntar se escoge de manera aleatoria. El protocolo del sabio i sí es determinístico porque para un estado local sólo una de las pruebas puede ser satisfecha.

Estos protocolos están definidos para cualquier sistema interpretado $\mathcal{I} = (\mathcal{R}, \pi_A)$. Definimos al protocolo conjunto $Pg^{\mathcal{I}}_A$ como

$$Pg^{\mathcal{I}}_A = (Pg^{\mathcal{I}}_{rey}, Pg^{\mathcal{I}}_1, \ldots, Pg^{\mathcal{I}}_n)$$

Por como está definido, las acciones dictaminadas por el protocolo conjunto $Pg_A^{\mathcal{I}}$ son las mismas que define el programa conjunto basado en conocimiento $\boldsymbol{Pg}_A$. Nos interesan los sistemas interpretados que representen a $\boldsymbol{Pg}_A$ en el contexto interpretado (γ_A, π_A).

3.6 El sistema que representa

Veamos que hay un único sistema que representa al programa $\boldsymbol{Pg}_A$ en el contexto interpretado (γ_A, π_A). Esto ocurre porque podemos ir construyendo paso a paso los t-prefijos del sistema y todos los puntos que ya hemos generado serán suficientes para construir el conjunto de los $(t+1)$-prefijos.

Supongamos que $\mathcal{I}_A = (\mathcal{R}_A, \pi_A)$ representa a $\boldsymbol{Pg}_A$. Veamos un resultado útil sobre el comportamiento del protocolo $Pg_A^{\mathcal{I}_A}$ en el contexto interpretado (γ_A, π_A).

Lema 4 *Sea $r \in \mathcal{R}_A$ una ejecución consistente con $Pg_A^{\mathcal{I}_A}$ en γ_A. Sea $r(t)$ el estado global de r al tiempo t.*

Si t es par entonces la acción conjunta que se realiza en la ronda $t+1$ es $a_{t+1} = (entregar_{ma}(todo), enviar_{rey}(\langle rey \rangle, G), \Lambda, \ldots, \Lambda)$.

Si t es impar entonces la acción conjunta que se realiza en la ronda $t+1$ es $a_{t+1} = (entregar_{ma}(todo), \Lambda, a_1, \ldots, a_n)$; donde $a_j = \Lambda$ o $a_i = enviar_i(msg_i, R_i)$, dependiendo de si el sabio recibió o no la pregunta del rey en la ronda anterior.

Prueba

Hagámoslo por inducción sobre el número de tiempo de la ejecución r. La base de la inducción es para $t = 0$ y $t = 1$.

Base

(a) Si $t = 0$ entonces como r es consistente con $Pg_A^{\mathcal{I}_A}$, $r(0)$ es un estado global inicial tal que $r_{ma}(0) = (0)$; $r_i(0) = (0, C, vision_i(\alpha), \emptyset)$ y $r_{rey}(0) = (0, C, \alpha, \emptyset)$, para alguna $\alpha \in \mathcal{CP}_C$.

Entonces $\pi_A(r_{rey}(0))(inicial) = verdadero$ y entonces $(\mathcal{I}, r_{rey}(0)) \models inicial$ y entonces $Pg_{rey}^{\mathcal{I}_A}(r_{rey}(0)) = enviar_{rey}(\langle rey \rangle, G)$, para algún conjunto G de sabios.

Por otro lado $\pi_A(r_i(0))(oyoalrey_i) = falso$, entonces $(\mathcal{I}, r_i(0)) \models \neg oyoalrey_i$. Por lo que $Pg_i^{\mathcal{I}_A}(r_i(0)) = \Lambda$.

Entonces $a = (entregar_{ma}(todo), enviar_{rey}(\langle rey\rangle, G), \Lambda, \ldots, \Lambda)$ es la acción conjunta que se realiza en la ronda 1.

(b) Si $t = 1$ entonces $r(1) = \tau_A(a)(r(0))$. Por lo que $r_{ma}(1) = (1)$; $r_{rey}(1) = (1, C, \alpha, hist_{r_{rey}}(1))$, donde $M^1_{r_{rey}} = enviar_{rey}(\langle rey\rangle, G)$ porque el rey no recibió ningún mensaje en la ronda 1 pero sí envió; $r_i(1) = (1, C, vision_i(\alpha), hist_{r_i}(1))$, donde $M^1_{r_i} = \{recibir_i(\langle rey\rangle, rey)\}$ para todo $i \in G$ y $M^1_{r_i} = \emptyset$ para todos los demás sabios.

Entonces $\pi_A(r_{rey}(1))(inicial) = falso$ y $\pi_A(r_{rey}(1))(oyorespuesta) = falso$. Entonces $(\mathcal{I}, r_{rey}(1)) \models \neg(inicial \vee oyorespuesta)$ y entonces $Pg^{\mathcal{I}_A}_{rey}(r_{rey}(1)) = \Lambda$.

Por otro lado $\pi_A(r_i(1))(oyoalrey_i) = verdadero$ para todo $i \in G$, entonces $(\mathcal{I}, r_i(1)) \models oyoalrey_i$. Entonces $Pg^{\mathcal{I}_A}_i(r_i(1)) = enviar_i(msg_i, R_i)$ donde msg_i corresponde a la respuesta del sabio, dependiendo de si sabe o no sabe su color.

En el caso de que $i \notin G$, entonces $\pi_A(r_i(1))(oyoalrey_i) = falso$ por lo que $(\mathcal{I}, r_i(1)) \models \neg oyoalrey_i$. Entonces $Pg^{\mathcal{I}_A}_i(r_i(1)) = \Lambda$.

Entonces $b = (entregar_{ma}(todo), \Lambda, a_1, \ldots, a_n)$ es la acción conjunta que se realiza en la ronda 2, donde $a_i = enviar_i(msg_i, R_i)$ para todo $i \in G$, $a_j = \Lambda$ para los demás sabios.

Hipótesis de inducción

Supongamos que el lema se cumple para toda $k' < k$. Demostrémoslo para k.

Paso Inductivo

(a) Si k es par entonces $k - 1$ es impar y por hipótesis de inducción $b = (entregar_{ma}(todo), \Lambda, a_1, \ldots, a_n)$ es la acción conjunta que se realizó en la ronda k, donde $a_i = enviar_i(msg_i, R_i)$ para todo $i \in G$, $a_j = \Lambda$ para los demás sabios, para G un subconjunto de los sabios. $r(k) = \tau_A(a)(r(k-1))$ por lo que $r_{ma}(k) = (k)$; $r_{rey}(k) = (k, C, \alpha, hist_{r_{rey}}(k))$, donde $M^k_{r_{rey}} = \{recibir_{rey}(msg_i, i) | i \in G\}$; $r_i(k) = (k, C, vision_i(\alpha), hist_{r_i}(k))$, donde $M^k_{r_i} = enviar_i(msg_i, R_i) \cup \{recibir_i(msg_j, j | j \in G, j \neq i\}$ para todo $i \in G$, $M^k_{r_j} = \{recibir_i(msg_j, j | j \in G\}$ para todos los demás.

Entonces $\pi_A(r_{rey}(k))(oyorespuesta) = verdadero$ por lo que $Pg^{\mathcal{I}_A}_{rey}(r_{rey}(k)) = enviar_{rey}(\langle rey\rangle, G')$, para algún conjunto G' de sabios.

Por otro lado $\pi_A(r_i(k))(oyoalrey_i) = falso$ para todo sabio i, entonces $(\mathcal{I}, r_i(k)) \models \neg oyoalrey_i$. Por lo que $Pg^{\mathcal{I}_A}_i(r_i(k)) = \Lambda$.

Entonces $a = (entregar_{ma}(todo), enviar_{rey}(\langle rey \rangle, G'), \Lambda, \ldots, \Lambda)$ es la acción conjunta que se realiza en la ronda $k+1$.

(b) Si k es impar entonces $k-2$ es impar y entonces $k-1$ es par y por hipótesis de inducción la acción conjunta que se realizó en la ronda k es $a = (entregar_{ma}(todo), enviar_{rey}(\langle rey \rangle, G), \Lambda, \ldots, \Lambda)$. $r(k) = \tau_A(a)(r(k-1))$ por lo que $r_{ma}(k) = (k)$; $r_{rey}(k) = (k, C, \alpha, hist_{r_{rey}}(k))$, donde $M^k_{r_{rey}} = enviar_{rey}(\langle rey \rangle, G)$; $r_i(k) = (k, C, vision_i(\alpha), hist_{r_i}(k))$, donde $M^k_{r_i} = \{recibir_i(\langle rey \rangle, rey)\}$ para todo $i \in G$ y $M^k_{r_i} = \emptyset$ para todos los demás sabios.

Entonces $\pi_A(r_{rey}(k))(inicial) = falso$ y $\pi_A(r_{rey}(k))(oyorespuesta) = falso$. Entonces $(\mathcal{I}, r_{rey}(k)) \models \neg(inicial \vee oyorespuesta)$ y entonces $Pg^{\mathcal{I}_A}_{rey}(r_{rey}(k)) = \Lambda$.

Por otro lado $\pi_A(r_i(k))(oyoalrey_i) = verdadero$ para todo $i \in G$, por lo tanto $(\mathcal{I}, r_i(k)) \models oyoalrey_i$.

Entonces $Pg^{\mathcal{I}_A}_i(r_i(k)) = enviar_i(msg_i, R_i)$ donde msg_i corresponde a la respuesta del sabio, dependiendo de si sabe o no sabe su color.

En el caso de que $i \notin G$, entonces $\pi_A(r_i(k))(oyoalrey_i) = falso$ por lo que $(\mathcal{I}, r_i(k)) \models \neg oyoalrey_i$. Entonces $Pg^{\mathcal{I}_A}_i(r_i(1)) = \{\Lambda\}$.

Entonces $b = (entregar_{ma}(todo), \Lambda, a_1, \ldots, a_n)$ es la acción conjunta que se realiza en la ronda k, donde $a_i = enviar_i(msg_i, R_i)$ para todo $i \in G$, $a_j = \Lambda$ para los demás sabios.□

Para que un sistema interpretado $\mathcal{I} = (\mathcal{R}, \pi_A)$ represente a $\boldsymbol{Pg}_A$ en (γ_A, π_A) se debe cumplir que π_A sea compatible con $\boldsymbol{Pg}_A$ y que $\mathcal{R}$ represente a $Pg^{\mathcal{I}}_A$ en γ_A. π_A es compatible con $\boldsymbol{Pg}_A$ por el lema 3. $\mathcal{R}$ representa a $Pg^{\mathcal{I}}_A$ en γ_A si $\mathcal{R}$ es el conjunto de todas las ejecuciones consistentes con $Pg^{\mathcal{I}}_A$ en γ_A. En este caso como $\Psi = True$, entonces todas las ejecuciones débilmente consistentes con $Pg^{\mathcal{I}}_A$ en γ_A son consistentes.

El lema 4 caracteriza cómo se va desarrollando el acertijo; en las rondas impares el rey pregunta a un conjunto de sabios si saben su color y los sabios no hacen nada; en las rondas pares contestan los sabios a los que el rey les preguntó en la ronda anterior; los demás sabios y el rey no hacen nada. Este lema será útil para demostrar el siguiente resultado.

Veamos que podemos construir paso a paso todos los t-prefijos de ejecuciones consistentes con el protocolo evaluado en el sistema que estamos construyendo. Las pruebas de conocimiento al tiempo t sólo

dependerán de los puntos que ya hemos construido y esto será lo que nos permita ir construyendo todos los prefijos.

Cada ejecución inicia en un estado global que pertenece a $\mathcal{G}_0$. Dada una caja C en $\mathcal{C}_n$ hay una correspondencia uno a uno entre el conjunto de configuraciones posibles en $\mathcal{CP}_C$ y el conjunto de estados globales iniciales para el acertijo con la caja C y n sabios.

A partir de un estado global inicial hay muchas posibles ejecuciones; una por cada posible subconjunto de los sabios a los que el rey puede preguntar. Hay entonces en cada ronda $2^n - 1$ posibles acciones que el rey puede tomar. La ejecución r^α es una ejecución tal que su estado global inicial, $r^\alpha(0)$, corresponde a la configuración α. No hay una única ejecución r^α.

Lema 5 *Para cualesquiera dos puntos* (r^α, t) *y* (r^β, u), $\alpha \in \mathcal{CP}_C$, $\beta \in \mathcal{CP}_{C'}$, $(r^\alpha, t) \sim_{rey} (r^\beta, u)$ *si y sólo si* $(r^\alpha, t) = (r^\beta, u)$. *Y si* $t \neq u$ *o* $C \neq C'$, *entonces para todo sabio* i, $(r^\alpha, t) \not\sim_i (r^\beta, u)$

Prueba

$(r^\alpha, t) \sim_{rey} (r^\beta, u)$ syss $r^\alpha_{rey}(t) = r^\beta_{rey}(u)$ syss $(t, C, \alpha, hist_{r^\alpha_{rey}}(t)) = (u, C', \beta, hist_{r^\beta_{rey}}(u))$ syss $t = u$, $C = C'$, $\alpha = \beta$ e $hist_{r^\alpha_{rey}}(t) = hist_{r^\beta_{rey}}(u)$ syss $(r^\alpha, t) = (r^\beta, u)$.

Si $t \neq u$ o $C \neq C'$ entonces $(r^\alpha_i(t)) = (t, C, vision_i(\alpha), hist_{r^\alpha_i}(t)) \neq (u, C', vision_i(\beta), hist_{r^\beta_i}(u)) = (r^\beta_i(u))$ por lo que $(r^\alpha, t) \not\sim_i (r^\beta, u)$.□

Corolario 1 *El sistema interpretado* $\mathcal{I}_A$ *es síncrono.*

Prueba

Por el lema anterior si $(r^\alpha, t) \sim_p (r^\beta, u)$ entonces $t = u$, para $p \in \{rey, 1, \ldots, n\}$. Entonces el sistema interpretado $\mathcal{I}_A$ es síncrono.□

Lema 6 *Existe un único sistema interpretado que representa a* Pg_A *en* (γ_A, π_A).

Prueba

Tenemos que ver que podemos generar todos los $t+1$-prefijos utilizando solamente los t-prefijos.

Tenemos todos los 0 prefijos porque $Pref_0(\mathcal{R}) = \{Pref_0(r) | r \in \mathcal{R}\} = \{r(0) | r \in \mathcal{R}\} = \mathcal{G}_0$.

Supongamos que tenemos todo el conjunto $Pref_t(\mathcal{R})$. Para generar al conjunto $Pref_{t+1}(\mathcal{R})$ el único problema podría presentarse en las

pruebas de conocimiento que realizan los sabios, las pruebas de la forma $K_i(c_i = a)$ para algún color a. Estando en un punto (r^α, t) el estado local del sabio i es $r_i^\alpha(t)$, la verdad de la prueba de conocimiento $K_i(c_i = a)$ depende del punto (r^α, t) y de todos los puntos (r^β, t') en los que el estado local del sabio i sea igual a $(r_i^\alpha(t))$. Como el sistema es síncrono entonces debe ocurrir que $t' = t$. Entonces las pruebas de conocimiento dependen únicamente de puntos de la forma (r^β, t), entonces si tenemos a todos los t-prefijos los sabios pueden hacer todas sus pruebas de conocimiento y entonces a partir de estos prefijos podemos generar todos los $t+1$-prefijos. El sistema que representa al programa Pg_A en (γ_A, π_A) se va conformando con los prefijos $Pref_t(\mathcal{R})$, para $t = 0, 1, \ldots$. □

Capítulo 4

Explorando el acertijo

Analizaremos las dos versiones del acertijo que ya hemos mencionado: la del acertijo de las esposas infieles, en el cual el rey pregunta simultáneamente a todos los sabios y la del acertijo de los tres sabios en el cual el rey va preguntando uno por uno a los sabios.

Definamos un contexto para cada uno de los acertijos; la diferencia con el contexto γ_A es la condición de admisibilidad.

Definición 44

$$\Psi_{sa} = \{r \mid M^{2t-1}_{r_{rey}} = enviar_{rey}(\langle rey \rangle, \{t\}))\}$$

para $t = 1, 2, \ldots, n$

$$\Psi_{ei} = \{r \mid M^{2t-1}_{r_{rey}} = enviar_{rey}(\langle rey \rangle, \mathcal{S})\}$$

para $t = 1, 2, \ldots$

Definimos dos nuevos contextos

Definición 45

$$\gamma_{sa} = (P_{ma}, \mathcal{G}_0, \tau, \Psi_{sa})$$

$$\gamma_{ei} = (P_{ma}, \mathcal{G}_0, \tau, \Psi_{ei})$$

En el acertijo de los sabios, en la ronda $t+1$ el rey escoge preguntarle al conjunto $\{t\}$, para $t = 1, \ldots, n$. En el acertijo de las esposas infieles siempre le pregunta al conjunto $\mathcal{S}$ de todos los sabios.

Cuando el rey llega a la reunión en el patio mayor, pide a los sabios que cierren los ojos y asigna un color a cada uno.

Por razones similares a la prueba de que hay un único sistema que representa a $\boldsymbol{Pg}_A$ en (γ_A, π_A), hay un único sistema interpretado que representa a $\boldsymbol{Pg}_A$ en (γ_{sa}, π_A) y un único sistema interpretado que representa a $\boldsymbol{Pg}_A$ en (γ_{ei}, π_A); en ambos sistemas el 0-prefijo es igual al conjunto de los estados globales iniciales $\mathcal{G}_0$ y a partir del t-prefijo generamos todos los estados globales del $t+1$-prefijo. Llamaremos $\mathcal{I}_{sa}$ y $\mathcal{I}_{ei}$ a estos sistemas interpretados. Podemos definir a dos nuevos protocolos conjuntos $Pg_A^{\mathcal{I}_{sa}}$ y $Pg_A^{\mathcal{I}_{ei}}$. En esta sección hablaremos de las ejecuciones consistentes con el protocolo del acertijo respectivo, en cada uno de los contextos interpretados (γ_{sa}, π_A) y (γ_{ei}, π_A).

La demostración del lema 4 también se aplica a estos sistemas, ya que son subconjuntos de las ejecuciones del sistema $\mathcal{I}_A$. En las rondas impares el rey pregunta a los sabios y en las rondas pares los sabios contestan. Es decir que para $t \geq 0$

En la ronda $2t+1$ en $\mathcal{I}_{sa}$ se lleva a cabo la acción

$$(entregar_{ma}(todo), enviar_{rey}(\langle rey \rangle, \{t\}), \Lambda, \ldots, \Lambda)$$

y en $\mathcal{I}_{ei}$ se lleva a cabo la acción

$$(entregar_{ma}(todo), enviar_{rey}(\langle rey \rangle, \mathcal{S}), \Lambda, \ldots, \Lambda)$$

En la ronda $2t+2$ en $\mathcal{I}_{sa}$ se lleva a cabo la acción

$$(entregar_{ma}(todo), \Lambda, \Lambda, \ldots, \Lambda, enviar_t(msg, R_t), \Lambda, \ldots, \Lambda)$$

y en $\mathcal{I}_{ei}$ se lleva a cabo la acción

$$(entregar_{ma}(todo), \Lambda, enviar_1(msg_1, R_1), \ldots, enviar_n(msg_n, R_n))$$

En cada uno de los sistemas interpretados $\mathcal{I}_{sa}$ y $\mathcal{I}_{ei}$ hay exactamente una ejecución correspondiente a cada una de las configuraciones en $\mathcal{CP}_C$; esto es porque ya no permitimos que el protocolo del rey escoga

al conjunto G de los sabios a los que pregunta en cada ronda, este conjunto de alguna manera ya está determinado de antemano con la condición de admisibilidad que define a cada uno de los contextos. A partir de un estado global inicial, correspondiente a la configuración α, hay una sóla ejecución que se genera en el sistema, porque hay una única acción conjunta posible en cada ronda; es decir, hay una única ejecución que se genera a partir de cada configuración posible α, identificamos a esta ejecución como r^α.

A partir de una configuración resultará útil hablar de nuevas configuraciones en las que modificamos el color de alguno de los sabios. Así que para una configuración α definimos como $\alpha_{i,b}$ a la configuración tal que para todo sabio $j \neq i$, $(s_j, c_j) \in \alpha_{i,b}$ syss $(s_j, c_j) \in \alpha$ y $(s_i, b) \in \alpha_{i,b}$. Cambiamos el color del sabio i en α por el color b y obtenemos la configuración $\alpha_{i,b}$. Es claro que si $(s_i, a) \in \alpha$ entonces $\alpha = (\alpha_{i,b})_{i,a}$. Veremos que $vision_i(\alpha) = vision_i(\alpha_{i,b})$ porque la única diferencia entre ambas configuraciones es el color del sabio i.

El rey escoge una configuración posible cada vez que se juega el acertijo; llamaremos la *configuración real* a esta configuración y la denotaremos por ρ.

Veamos algunos resultados que se cumplen en todos los sistemas del acertijo.

Lema 7 *Sea $\alpha \in \mathcal{CP}_C$ para todo sabio i y para toda configuración $\beta \in \mathcal{CP}_C$ se cumple que $vision_i(\alpha) = vision_i(\beta)$ si y sólo si $\beta = \alpha_{i,b}$ para algún color b.*

Prueba

Sea a tal que $(s_i, a) \in \alpha$ y sea b, no necesariamente distinto a a, tal que $(s_i, b) \in \beta$.

$vision_i(\alpha) = vision_i(\beta)$ syss la única diferencia posible entre α y β es el color del sabio i, syss para todo color c, si $j \neq i$ ocurre que $(s_j, c) \in \alpha$ syss $(s_j, c) \in \beta$ syss $\beta = \alpha_{i,b}$.□

Lema 8 *Supongamos que $(s_i, a) \in \alpha$, entonces para todo color $b \neq a$ en $\alpha_{i,b}$ se cumple que para $c \neq a, b$; $f_c^{\alpha_{i,b}} = f_c^\alpha$ y se cumple que $f_b^{\alpha_{i,b}} = f_b^\alpha - 1$; $f_a^{\alpha_{i,b}} = f_a^\alpha + 1$.*

Prueba

Para todo color $c \neq a, b$ se cumple que $(s_i, c) \in \alpha$ syss $(s_i, c) \in \alpha_{i,b}$, entonces $p_c^\alpha = p_c^{\alpha_{i,a}}$, entonces $f_c^{\alpha_{i,b}} = \sharp c_c - p_c^{\alpha_{i,a}} = \sharp c_c - p_c^\alpha = f_c^\alpha$.

Se cumple que $vision_i(\alpha) = vision_i(\alpha_{i,b})$ y $(s_i, a) \in \alpha$, $(s_i, b) \in \alpha_{i,a}$. Entonces si $a \neq b$ se cumple que $p_b^{\alpha_{i,b}} = p_b^{\alpha} + 1$ y $p_a^{\alpha_{i,b}} = p_a^{\alpha} - 1$.

Entonces $f_b^{\alpha_{i,b}} = \sharp c_b - p_b^{\alpha_{i,b}} = \sharp c_b - (p_b^{\alpha} + 1) = f_b^{\alpha} - 1$.

También $f_a^{\alpha_{i,b}} = \sharp c_a - p_a^{\alpha_{i,b}} = \sharp c_a - (p_a^{\alpha} - 1) = f_b^{\alpha} + 1$.□

Otro resultado

Lema 9 *Para una caja $C \in \mathcal{C}_n$, sea $\alpha \in \mathcal{CP}_C$, entonces para todo color b, $\alpha_{i,b} \in \mathcal{CP}_C$ syss $f_b^{\alpha} > 0$.*

Prueba

$\alpha_{i,b} \in \mathcal{CP}_C$ syss para todo color b, $1 \leq b \leq r$, se cumple $p_b^{\alpha_{i,b}} \leq \sharp c_b$, syss $f_b^{\alpha_{i,b}} \geq 0$, porque $f_b^{\alpha_{i,b}} = \sharp c_b - p_b^{\alpha_{i,b}} \geq 0$ syss $f_b^{\alpha} > 0$, porque por el lema anterior $f_b^{\alpha_{i,b}} = f_b^{\alpha} - 1$.□

En el lema 5 vimos algunas propiedades de distinguibilidad entre los puntos del sistema $\mathcal{I}_A$; luego vimos que este sistema es un sistema síncrono. Estos resultados se cumplen también para los sistemas $\mathcal{I}_{sa}$ y $\mathcal{I}_{ei}$ porque en las demostraciones sólo se hizo referencia a las propiedades de los puntos y los puntos en $\mathcal{I}_{sa}$ y $\mathcal{I}_{ei}$ también son puntos de $\mathcal{I}_A$. Es decir que los sistemas $\mathcal{I}_{sa}$ y $\mathcal{I}_{ei}$ son sistemas síncronos.

4.1 Una interpretación gráfica

La idea básica de la interpretación de conocimiento que estamos manejando es que un agente o procesador sabe un hecho si el hecho es verdadero en todos los mundos que considera posibles; en el caso de los sistemas interpretados serían todos los puntos que considera posibles.

Veamos qué pasa en $\mathcal{I}_{sa}$ y en $\mathcal{I}_{ei}$. Los puntos de los sistemas representan a los estados globales que se alcanzan en cada ejecución. Ambos sistemas son síncronos, debido a que todos los agentes tienen al tiempo en su estado local; es decir, un agente distingue entre dos estados globales que tengan distinto tiempo. Nos interesa saber cuándo un agente no distingue entre dos puntos. Veremos que los puntos correspondientes a los estados globales iniciales forman una sola componente conexa, bajo la relación de accesibilidad, si se cumple que el rey lleva al menos $n + 1$ etiquetas en total.

La diferencia entre dos estados globales iniciales puede estar en la caja o en la configuración. El rey tiene guardada en su estado local a la caja y a la configuración; el rey distingue a todos los puntos porque dos puntos tendrán a la misma caja y a la misma configuración, en el estado del rey, sólo si pertenecen a la misma ejecución, pero el rey distingue dos puntos distintos de una misma ejecución porque guarda el tiempo en su estado local.

Los sabios tienen a la caja y a su visión guardadas en su estado local; un sabio no distingirá entre dos estados globales iniciales si tienen el mismo tiempo y la misma caja y si su visión es la misma en ambos estados. Por el lema 7 entre dos configuraciones α y β, $vision_i(\alpha) = vision_i(\beta)$ si y sólo si $\beta = \alpha_{i,b}$ para algún color b. Es decir que el sabio i no distingue entre dos estados globales iniciales si y sólo si corresponden a dos configuraciones de este tipo. Pero para que haya un estado inicial correspondiente a una configuración de la forma $\alpha_{i,b}$ por el lema 9 debe ocurrir que $f_b^\alpha > 0$. Esta discusión será la clave para construir la gráfica correspondiente a los puntos de los sistemas interpretados que estamos estudiando.

Cada punto de una ejecución estará representado por un nodo en la gráfica. Entre dos puntos habrá una arista dirigida etiquetada con i si el estado local del sabio i es el mismo en ambos puntos. Como el rey distingue entre todos los puntos, las etiquetas sólo corresponderán a los sabios. Como la relación de indistinguibilidad es una relación de equivalencia, cada punto tiene un lazo dirigido etiquetado con todos los procesadores, pero por simplicidad omitiremos estos lazos sabiendo que todos los puntos los tienen. Por la simetría las aristas siempre son bidireccionales y se cumple la transitividad.

Como todos los procesadores distinguen dos puntos que pertenezcan a distintas cajas o a distinto tiempo, las componentes conexas de la gráfica corresponden a puntos de ejecuciones generadas por configuraciones de la misma caja y el mismo tiempo. Veremos que también las historias de mensajes tienen que ver en la conexidad de la gráfica.

Esta gráfica nos interesa porque representa los puntos sobre los que se hacen las pruebas de conocimiento en los protocolos de los sabios. Por el lema 4 los sabios siempre responden en las rondas pares; entonces las prueba de conocimiento las realizan sobre los puntos con tiempo impar. Estaremos interesados mayormente en estos puntos porque en

las rondas pares ya sabemos cuál es la acción conjunta que se lleva a cabo en cada sistema.

En los resultados siguientes, cuando hagamos referencia al sistema interpretado $\mathcal{I}$, estaremos pensando en cualquiera de los sistemas $\mathcal{I}_{ei}$ o en $\mathcal{I}_{sa}$. Lo que queremos es identificar algunas propiedades de la gráfica asociada a la estructura Kripke $M_{\mathcal{I}_{ei}}$ y la asociada a $M_{\mathcal{I}_{sa}}$.

Si dos puntos (r^α, t) y (r^β, t) son indistinguibles para el sabio i, es decir, si $(r^\alpha, t) \sim_i (r^\beta, t)$, ocurre que cualesquiera dos puntos (r^α, t') y (r^β, t') tales que $t' > t$ cumplen con que tienen el mismo tiempo y la misma caja; también se cumple que $vision_i(\alpha) = vision_i(\beta)$ y entonces lo único que puede hacer que el sabio i distinga a (r^α, t') de (r^β, t') es que la historia de mensajes en el estado local del sabio i en cada punto no sea la misma; esto va a ocurrir sólo si el sabio i recibe distintos mensajes de algún sabio, es decir si la respuesta de algún sabio es distinta. Como cada sabio distinto a i tiene el mismo color en ambos puntos sólo puede ocurrir que en una ejecución un sabio responda que sabe su color y que en la otra responda que no sabe. No pude ocurrir que un sabio responda en ambas que sabe su color y que reporte colores diferentes.

Nos interesa saber cuándo dos puntos correspondientes a estados globales iniciales son adyacentes; de hecho queremos saber cuando son accesibles. En general cuando digamos accesibilidad nos estamos refiriendo a la $\mathcal{S}$-accesibilidad, es decir a los puntos accesibles considerando las etiquetas de todos los sabios.

Veamos que ocurre si $\sum_a^r \sharp c_a = n$, es decir si $k = 0$.

Lema 10 *Si $k = 0$ para cualesquiera dos puntos distintos (r^α, t) y (r^β, t), para todo sabio i se cumple que $(r^\alpha, t) \not\sim_i (r^\beta, t)$.*

Prueba

Como $k = 0$ entonces $\sum_{a=1}^r \sharp c_a = n$ y entonces $\sum_{a=1}^r f_a^\alpha = 0$, para toda $\alpha \in \mathcal{CP}_C$. Entonces $f_a^\alpha = 0$ para $a = 1, \ldots, r$.

Supongamos que $(r^\alpha, t) \sim_i (r^\beta, t)$, por el lema 5 debe ocurrir que $\alpha, \beta \in \mathcal{CP}_C$ y además debe ocurrir que $vision_i(\alpha) = vision_i(\beta)$. Por el lema 7 entonces $\beta = \alpha_{i,b}$ para algún color b pero por el lema 9 $\alpha_{i,b} \in \mathcal{CP}_C$ syss $f_b^\alpha > 0$. Pero $f_b^\alpha = 0$. Una contradicción. Entonces $(r^\alpha, t) \not\sim_i (r^\beta, t)$.□

Veamos que ocurre si $\sum_{a=1}^r \sharp c_a > n$.

Lema 11 *Si $k > 0$ entonces para $(r^{\alpha}, 0)$ tal que $\alpha \in \mathcal{CP}_C$ para alguna caja C, $(r^{\beta}, 0)$ es accesible desde $(r^{\alpha}, 0)$ si y sólo si $\beta \in \mathcal{CP}_C$.*

Prueba

($\Rightarrow$) Si $(r^{\beta}, 0)$ es accesible desde $(r^{\alpha}, 0)$ entonces existen $(r^{\alpha_0}, 0), (r^{\alpha_1}, 0), \ldots, (r^{\alpha_m}, 0)$ tales que $(r^{\alpha}, 0) = (r^{\alpha_0}, 0)$; $(r^{\beta}, 0) = (r^{\alpha_m}, 0)$ y además $(r^{\alpha_{j-1}}, 0) \sim_{i_j} (r^{\alpha_j}, 0)$ para algún sabio i_j, $j = 1, \ldots, m$. Por el lema 5 $\alpha_{j-1}, \alpha_j \in \mathcal{CP}_{C'}$ para alguna caja C', para $j = 0, 1, \ldots, m$. Como $\alpha_0 \in \mathcal{CP}_C$ entonces debe ocurrir que $\alpha_j \in \mathcal{CP}_C$ para $j = 1, \ldots, m$, por lo que $\beta \in \mathcal{CP}_C$.

($\Leftarrow$) Supongamos que se cumple $\alpha, \beta \in \mathcal{CP}_C$.

Sea

$$\alpha = \{(s_1, c_1), (s_2, c_2), \ldots, (s_n, c_n)\}$$

$$\beta = \{(s_1, d_1), (s_2, d_2), \ldots, (s_n, d_n)\}$$

Vayamos colocando uno por uno el color de cada sabio. Sea $\alpha_0 = \alpha$. Para el sabio i supongamos que ya hemos colocado los $i-1$ colores de los sabios anteriores. Es decir que

$$\alpha_{i-1} = ((s_1, d_1), \ldots, (s_{i-1}, d_{i-1}), (s_i, c_i), \ldots, (s_n, c_n))$$

Si $f_{d_i}^{\alpha_{i-1}} > 0$ entonces sea $\alpha_i = (\alpha_{i-1})_{i,d_i}$, forzando a que $(s_i, d_i) \in \alpha_i$. Si $f_{d_i}^{\alpha_{i-1}} = 0$ no podemos hacer lo mismo ya que en α_{i-1} ocurre que $p_{d_i}^{\alpha_{i-1}} = \sharp c_{d_i}$. No puede ser que en los primeros $i-1$ sabios de α_{i-1} estén todos los sabios a los que se les asignó el color d_i, porque estos sabios tienen asignado el mismo color que en β, si en los primeros $i-1$ sabios de α_{i-1} estuvieran todos los sabios con color igual a d_i; esto ocurre también en β, pero entonces no sería posible $(s_i, d_i) \in \beta$. Entonces existe j tal que $j > i$ y $(s_j, d_i) \in \alpha_{i-1}$. Sea $\beta_i = (\alpha_{i-1})_{j,d}$ para d cualquier color tal que $f_d^{\alpha_{i-1}} > 0$, que siempre existe porque $k > 0$. Entonces $f_{d_i}^{\beta_i} > 0$ y sea $\alpha_i = (\beta_i)_{i,d_i}$. Entonces $\alpha_i = \{(s_1, d_1), (s_2, d_2), \ldots, (s_i, d_i), (s_{i+1}, c_{i+1}), \ldots, (s_j, d), \ldots, (s_n, c_n)\}$. α_i siempre es accesible desde α_{i-1}. Haciendo este proceso n veces ocurre que $\beta = \alpha_n$ y entonces β es accesible desde α.□

Entonces si en la caja ocurre que $\sum_{a=1}^{r} \sharp c_a \geq n+1$, es decir, si $k > 0$, todos los estados globales iniciales correspondientes a configuraciones de una misma caja forman una sola componente conexa. Si $k = 0$ todos los sabios distinguen cualesquiera dos puntos y ocurre que todos los sabios saben su color desde el inicio del acertijo.

Si dos puntos son G-accesibles, para algún conjunto G de sabios, los dos puntos deben tener el mismo tiempo y la misma caja en todos sus estados locales. De hecho dos puntos accesibles deben tener registradas las mismas acciones de todos los sabios. La prueba la hacemos sobre los puntos con tiempo impar que son los que nos interesan porque en los puntos con tiempo par ya sabemos lo que ocurre: es cuando el rey le pregunta a los sabios.

Lema 12 *Para cualesquiera dos puntos* $(r^\alpha, 2t-1)$ *y* $(r^\beta, 2t-1)$, *si* $(r^\alpha, 2t-1) \sim_i (r^\beta, 2t-1)$ *entonces* $hist_{r_j^\alpha}(2t-1) = hist_{r_j^\beta}(2t-1)$, *para todo sabio* j.

Prueba

Si $(r^\alpha, 2t-1) \sim_i (r^\beta, 2t-1)$ entonces $hist_{r_i^\alpha}(2t-1) = hist_{r_i^\beta}(2t-1)$ porque el estado local del sabio i es el mismo en ambos puntos.

Debemos mostrar que se cumple $hist_{r_j^\alpha}(2t-1) = hist_{r_j^\beta}(2t-1)$ para todo sabio $j \neq i$. esto ocurre si y sólo si

$$M^{t'}_{r_j^\alpha} = M^{t'}_{r_j^\beta}$$

para toda $t' < 2t-1$, donde $hist_{r_j^\alpha}(2t-1) = (M^1_{r_j^\alpha}, \ldots, M^{2t-1}_{r_j^\alpha})$ e $hist_{r_j^\beta}(2t-1) = (M^1_{r_j^\beta}, \ldots, M^{2t-1}_{r_j^\beta})$.

Los únicos eventos que pueden estar en $M^{t'}_{r_j^\alpha}$ y en $M^{t'}_{r_j^\beta}$ son

$$recibir_j(msg, h),\ recibir_j(\langle rey \rangle, rey)\ o\ enviar_j(msg, R_j)$$

Hay tres casos:

(a) $recibir_j(msg, h) \in M^{t'}_{r_j^\alpha}$ syss $enviar_h(msg, R_h) \in M^{t'}_{r_h^\alpha}$ syss $recibir_i(msg, h) \in M^{t'}_{r_i^\alpha}$ syss $recibir_i(msg, h) \in M^{t'}_{r_i^\beta}$, porque $(M^{t'}_{r_i^\alpha} = M^{t'}_{r_i^\beta})$ syss

$enviar_h(msg, R_h) \in M^{t'}_{r^\beta_h}$ syss $recibir_j(msg, h) \in M^{t'}_{r^\beta_j}$

(b) $recibir_j(\langle rey\rangle, rey) \in M^{t'}_{r^\alpha_j}$ syss $recibir_j(\langle rey\rangle, rey) \in M^{t'}_{r^\beta_j}$

porque las acciones del rey siempre son las mismas en todas las ejecuciones.

(c) $enviar_j(msg, R_j) \in M^{t'}_{r^\alpha_j}$ syss $recibir_i(msg, j) \in M^{t'}_{r^\alpha_i}$ syss

$recibir_i(msg, j) \in M^{t'}_{r^\beta_i}$ porque $(M^{t'}_{r^\alpha_i} = M^{t'}_{r^\beta_i})$ syss

$enviar_j(msg, R_j) \in M^{t'}_{r^\beta_j}$.

Entonces $(M^{t'}_{r^\alpha_j} = M^{t'}_{r^\beta_j})$, para toda $t' < 2t-1$, por lo que $hist_{r^\alpha_j}(2t-1) = hist_{r^\beta_j}(2t-1)$ para todo sabio j.□

Estos dos últimos lemas formalizan la idea de que dos puntos serán indistinguibles para un sabio i si y sólo si todos los sabios han respondido lo mismo en ambos puntos.

Los puntos de una misma componente conexa en la gáfica, corresponden a ejecuciones en las que las respuestas de cada sabio han sido las mismas.

Lema 13 *Para algún subconjunto G de los sabios, si un punto $(r^\beta, 2t-1)$ es G-accesible desde $(r^\alpha, 2t-1)$ entonces $\alpha, \beta \in \mathcal{CP}_C$ para alguna caja C e $hist_{r^\alpha_i}(2t-1) = hist_{r^\beta_i}(2t-1)$ para todo sabio i.*

Prueba

Si $(r^\beta, 2t-1)$ es G-accesible desde $(r^\alpha, 2t-1)$ entonces existen puntos $(r^{\alpha_0}, 2t-1), (r^{\alpha_1}, 2t-1), \ldots, (r^{\alpha_m}, 2t-1)$ tales que $(r^\alpha, 2t-1) = (r^{\alpha_0}, 2t-1)$, $(r^\beta, 2t-1) = (r^{\alpha_m}, 2t-1)$ y además $(r^{j-1}, 2t-1) \sim_{i_j} (r^j, 2t-1)$ para $i_j \in G$.

Por el lema 5 todas las ejecuciones corresponden a configuraciones que pertenecen a una misma caja. Es decir que $\alpha, \beta \in \mathcal{CP}_C$.

Por el lema 12 para $j = 1, \ldots, m$ se cumple $hist_{r^{\alpha_{j-1}}_i}(2t-1) = hist_{r^{\alpha_j}_i}(2t-1)$ para todo i. Entonces $hist_{r^{\alpha_0}_i}(2t-1) = hist_{r^{\alpha_n}_i}(2t-1)$ para todo sabio i. Entonces $hist_{r^\alpha_i}(2t-1) = hist_{r^\beta_i}(2t-1)$ para todo sabio i.□

Corolario 2 *Para cualesquiera dos puntos* (r^α, t) *y* (r^β, t) *y para un subconjunto* G *de los sabios, si* (r^β, t) *es* G*-accesible desde* (r^α, t) *entonces* (r^β, t') *es* G*-accesible desde* (r^α, t') *para toda* $t' < t$.

Prueba

Hagamoslo por inducción sobre en número de pasos que toma llegar de (r^β, t) a (r^α, t). Si (r^β, t) es G-accesible desde (r^α, t) en 1 paso entonces $(r^\beta, t) \sim_i (r^\alpha, t)$ para $i \in G$. Entonces $\alpha, \beta \in \mathcal{CP}_C$ y $vision_i(\alpha) = vision_i(\beta)$. Por el lema 12 $hist_{r_i^\beta}(t) = hist_{r_i^\alpha}(t)$, entonces $hist_{r_i^\beta}(t') = hist_{r_i^\alpha}(t')$ para todo $t' < t$ y entonces $(r^\beta, t') \sim_i (r^\alpha, t')$.

Supongamos que se cumple para todo (r^β, t) G-accesible desde (r^α, t) en $k-1$ pasos. Sea (r^β, t) G-accesible desde (r^α, t) en k pasos. Entonces existen $(r^{\alpha_0}, t), (r^{\alpha_1}, t), \ldots, (r^{\alpha_k}, t)$ tales que $(r^\alpha, t) = (r^{\alpha_0}, t)$, $(r^\beta, t) = (r^{\alpha_k}, t)$ y además $(r^{\alpha_{j-1}}, t) \sim_{i_j} (r^{\alpha_j}, t)$ para $i_j \in G$. Entonces $(r^{\alpha_{k-1}}, t)$ es G accesible desde (r^α, t) en $k-1$ pasos. Entonces por hipótesis de inducción para todo $t' < t$ $(r^{\alpha_{k-1}}, t')$ es G accesible desde (r^α, t') en $k-1$ pasos. Por el lema 13 para todo $t' < t$ $hist_{r_j^{\alpha_{k-1}}}(t') = hist_{r_j^\alpha}(t')$ para todo sabio j . Como $(r^{\alpha_{k-1}}, t) \sim_{i_k} (r^\beta, t)$ entonces por el lema 12 $hist_{r_j^{\alpha_{k-1}}}(t) = hist_{r_j^\beta}(t)$, para todo sabio j y entonces para todo $t' < t$, $hist_{r_j^{\alpha_{k-1}}}(t') = hist_{r_j^\beta}(t')$, para todo sabio j. Entonces $(r^{\alpha_{k-1}}, t') \sim_{i_k} (r^\beta, t')$. Entonces (r^β, t') es G-accesible desde (r^α, t') para toda $t' < t$.□

La importancia de la gráfica es que refleja las relaciones $\mathcal{K}_i$ de la estructura Kripke $M_\mathcal{I}$. Nos interesa para poder visualizar mejor como evalúa el sabio i la prueba de conocimiento que realiza en el programa Pg_A. El sabio i realiza la prueba $K_i(c_i = a)$, para todo color a.

Lema 14 *Se cumple* $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$ *syss* $(s_i, a) \in \alpha$ *y para todo otro punto* (r^β, t) *se cumple que* $(r^\alpha, t) \not\sim_i (r^\beta, t)$.

Prueba

($\Rightarrow$) Si se cumple $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$ entonces se cumple $(\mathcal{I}, r_i^\beta(t')) \models (c_i = a)$ para todo (r^β, t') tal que $(r^\alpha, t) \sim_i (r^\beta, t')$. Como $(r^\alpha, t) \sim_i (r^\alpha, t)$ se cumple $(\mathcal{I}, r_i^\alpha(t)) \models (c_i = a)$ por lo que $(s_i, a) \in \alpha$. Supongamos que existe otro punto (r^β, t') tal que $(r^\alpha, t) \sim_i (r^\beta, t')$, entonces $r_i^\beta(t') = r_i^\alpha(t)$, por lo que $t' = t$ y $vision_i(\beta) = vision_i(\alpha)$. Por el lema 7 debe ocurrir que $\beta = \alpha_{i,b}$ para algún color b, como $\beta \neq \alpha$

entonces $b \neq a$ y entonces $(s_i, b) \in \beta$ por lo que $(\mathcal{I}, r_i^\beta(t')) \models \neg(c_i = a)$, una contradicción. Entonces para todo otro punto (r^β, t') se cumple que $(r^\alpha, t) \not\sim_i (r^\beta, t')$.

($\Leftarrow$) Si $(s_i, a) \in \alpha$ y para todo otro punto (r^β, t') se cumple que $(r^\alpha, t) \not\sim_i (r^\beta, t')$, como se cumple $(\mathcal{I}, r_i^\alpha(t)) \models (c_i = a)$ y $(r^\alpha, t) \sim_i (r^\alpha, t)$ entonces se cumple $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$.□

Podemos relacionar al protocolo de cada sabio con la gráfica asociada a M_T, de manera que $Pg_i^{\mathcal{I}}(r_i^\alpha(2t-1)) = enviar_i(\langle a, i\rangle, R_i)$ syss en la gráfica no hay ninguna arista etiquetada con i que salga del punto $(r^\alpha, 2t-1)$.

Veamos que el conocimiento no se pierde:

Lema 15 *Si* $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$ *entonces* $(\mathcal{I}, r_i^\alpha(t')) \models K_i(c_i = a)$ *para todo* $t' > t$

Prueba

Si $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$ entonces $(s_i, a) \in \alpha$ y para todo otro punto (r^β, t) se cumple que $(r^\beta, t) \not\sim_i (r^\alpha, t)$. Entonces $(r^\beta, t') \not\sim_i (r^\alpha, t')$ para toda $t' > t$ entonces $(\mathcal{I}, r_i^\alpha(t')) \models K_i(c_i = a)$ para toda $t' > t$.□

Recordemos que la utilidad de la gráfica será para las respuestas de los sabios; nos fijaremos únicamente en los puntos que tengan tiempo impar, correspondientes a las rondas pares que es cuando los sabios responden al rey. La pregunta del rey en cada ronda impar nos indica en qué etiquetas debemos fijarnos.

Cuando $k = 0$ tenemos un acertijo en el que todos los sabios responden siempre que sí saben su color.

Lema 16 *Si* $k = 0$ *entonces* $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$ *para todo punto* (r^α, t) *y todo sabio* i *tal que* $(s_i, a) \in \alpha$.

Prueba

Por el lema 10 para cualquier punto (r^β, t) se cumple $(r^\alpha, t) \not\sim_i (r^\beta, t)$ para todo sabio i. Por el lema 14 entonces si $(s_i, a) \in \alpha$ se cumple $(\mathcal{I}, r_i^\alpha(t)) \models K_i(c_i = a)$.□

Al rey no le interesarán las cajas donde $k = 0$ porque el juego deja de tener sentido para él si todos los sabios saben siempre su color.

4.2 Colores inevitables

Dijimos que cuando los sabios se juntaron a deliberar la respuesta que mandarán al rey, se dieron cuenta de que existen dos tipos de caja: el tipo *bueno* y el tipo *malo*. Primero, sin mucha dificultad, los sabios se dan cuenta de que si en una caja no existe alguna asignación de los colores que permita que a la primera pregunta algún sabio sepa su color, entonces el juego no tiene sentido, nunca va a ser posible que algún sabio sepa su color por más que el rey pregunte. Un poco más adelante demostraremos formalmente este resultado. Estas cajas son las de tipo *malo*. Veremos que en el tipo de caja *bueno* debe existir alguna asignación inicial de colores en la cuál algún sabio responda afirmativamente a la primera pregunta del rey. Para que un sabio responda sí a la primera pregunta del rey, debe suceder que éste le asigne al sabio i un color d y que para cada color $b \neq d$ asigne ese color b a $\sharp c_b$ sabios; así el sabio i puede estar seguro de que su color es d porque si el rey le hubiera asignado otro color estaría violando la reestricción de la caja. Definamos como dif_a a la suma $\sum_{b \neq a} \sharp c_b$, es decir el número máximo de sabios que el rey puede asignar con un color disitinto a a. Llamemos *inevitable* al color d si $dif_d < n$, donde el número de sabios es n.

Definición 46 *Para cualquier color a,*

$$dif_a = \sum_{b \neq a} \sharp c_b$$

Definamos a los colores inevitables

Definición 47 *El color d es* **inevitable** *si*

$$dif_d < n$$

A los colores que no sean inevitables, es decir que $dif(a) \geq n$, los llamaremos *evitables*.

A continuacióbn dos resultados que motivan la definición de colores inevitables.

Lema 17 *Sea d un color inevitable en la caja C. Sea $\alpha \in \mathcal{CP}_C$ una configuración en la que $(s_i, d) \in \alpha$ y $p_b^\alpha = \sharp c_b$ para toda $b \neq d$. Entonces para toda $\beta \in \mathcal{CP}_C$ tal que $\alpha \neq \beta$, $vision_i(\alpha) \neq vision_i(\beta)$.*

Prueba

Como d es un color inevitable, entonces $dif_d < n$, por lo que es congruente pensar que es posible que en α el rey haya asignado a dif_d sabios un color distinto a d, esto es que ocurre que $p_b^\alpha = \sharp c_b$ para todo color $b \neq d$ y entonces $\sum_{b \neq d} p_b^\alpha = dif_d$, $f_d^\alpha = k$ y $f_b^\alpha = 0$ para toda $b \neq d$. Entonces por el lema 9 se cumple que $\alpha_{i,b} \notin \mathcal{CP}_C$ para todo color $b \neq d$. Entonces por el lema 7 para toda $\beta \in \mathcal{CP}_C$ se cumple que $vision_i(\alpha) \neq vision_i(\beta)$.□

En la interpretación gráfica, esto quiere decir que si a es un color inevitable entonces hay puntos $(r^\alpha, 0)$ para los que se cumple que $(s_i, a) \in \alpha$ y que el rey asignó a dif_a sabios un color distinto a a, esto es posible porque $dif_a < n$. Entonces se cumple que $(r^\beta, 0) \not\sim_i (r^\alpha, 0)$ para cualquier otro punto $(r^\beta, 0)$ y entonces por el lema 14 se cumple $(\mathcal{I}, r_i^\alpha(0)) \models K_i(c_i = a)$. Es decir que el sabio i sabe su color desde antes de que el rey inicie a preguntar a los sabios.

Lema 18 *Sea C una caja que no tiene colores inevitables, entonces para toda configuración $\alpha \in \mathcal{CP}_C$ hay al menos dos colores a, b tales que $f_a^\alpha, f_b^\alpha > 0$.*

Prueba En α no hay colores inevitables, entonces $dif_a \geq n$, para $a = 1, \ldots, r$. Debe ocurrir que $k > 0$ porque sino $\sum_{i=1}^r \sharp c_i = n$ y entonces $dif_a < n$ para todo a. Entonces como $\sum_{a=1}^r f_a^\alpha = k$ entonces hay al menos un color a tal que $f_a^\alpha > 0$. Supongamos que $(s_i, a) \in \alpha$ y que $f_b^\alpha = 0$ para todo color $b \neq a$, entonces $p_b^\alpha = \sharp c_b$ y entonces en α estarían asignados dif_a sabios con un color distinto al color a y entonces $dif_a < n$, una contradicción. Entonces debe ocurrir que para otro color b se cumple que $f_b^\alpha > 0$.□

Este resultado lo vamos a utilizar para ver que, dada cualquier configuración α en una caja que no tiene colores inevitables, para cualquier sabio es posible encontrar una configuración β tal que $vision_i(\alpha) = vision_i(\beta)$.

Corolario 3 *Sea C una caja que no tiene colores inevitables, entonces para toda configuración α y para todo sabio i siempre existe $\alpha_{i,b} \in \mathcal{CP}_C$ tal que $\alpha_{i,b} \neq \alpha$ para algún color b.*

Prueba

Supongamos que $(s_i, a) \in \alpha$. Por el lema 18 siempre hay un color $b \neq a$ tal que $f_b^\alpha > 0$, $\alpha_{i,b} \neq \alpha$ y por el lema 9 entonces $\alpha_{i,b} \in \mathcal{CP}_C$.□

Si tenemos una configuración α y queremos encontrar otra configuración β, tal que $vision_i(\alpha) = vision_i(\beta)$ para algún sabio i, es decir una configuración que el sabio i no distinga de α, debemos fijarnos en el color del sabio i y en todos los colores $b \neq a$ tales que $p_b^\alpha < \sharp c_b$, intercambiando el color del sabio i por alguno de estos colores obtenemos las configuraciones deseadas.

Lema 19 *Existe un color inevitable en una caja C si y sólo si existe una configuración que permite a algún sabio responder afirmativamente a la primera pregunta del rey*

Prueba

($\Rightarrow$) Supongamos que hay algún color inevitable d en la caja C, entonces $dif_d < n$. Sea $\alpha \in \mathcal{CP}_C$ una configuración en la que $(s_i, d) \in \alpha$, supongamos que el rey le pregunta al sabio i en la primera ronda. Supongamos también que el rey asignó a dif_d sabios un color distinto a d, es decir que $\sum_{b \neq d} p_b^\alpha = dif_d$ y entonces $p_b^\alpha = \sharp c_b$ para todo $b \neq d$. Entonces $f_b^\alpha = 0$ para todo $b \neq d$. Entonces por el lema 17 para toda $\beta \in \mathcal{CP}_C$ tal que $\alpha \neq \beta$, $vision_i(\alpha) \neq vision_i(\beta)$ y entonces $(r^\alpha, 1) \not\sim_i (r^\beta, 1)$. Por el lema 14 se cumple $(\mathcal{I}, r_i^\alpha(1)) \models K_i(c_i = d)$ y entonces $Pg^{\mathcal{I}_i}(r_i^\alpha(1)) = enviar_i(\langle d, i\rangle, R_i)$. Y entonces a la primera pregunta del rey el sabio i sabe su color.

($\Leftarrow$) Demostraremos la contrapuesta, supongamos que en la caja C no hay algún color inevitable. Entonces $\sharp c_a \leq k$ para toda a. Tenemos que demostrar que para toda configuración cualquier sabio respondería no a la primera pregunta del rey. Sea $r^\alpha \in \mathcal{I}$. Veamos que siempre se cumple $(\mathcal{I}, r_i^\alpha(1)) \models oyoalrey_i \wedge \neg K_i(c_i = 1) \wedge \ldots \wedge \neg K_i(c_i = r)$ para cualquier i y entonces la acción del sabio i en la ronda 2 es enviar a todos los demás el mensaje $\langle no, i\rangle$ si es que el rey le pregunta.

Supongamos que el rey envió su pregunta al conjunto G de sabios, que $i \in G$ y $(s_i, a) \in \alpha$.

Se cumple $(\mathcal{I}, r_i^\alpha(1)) \models oyoalrey_i$ porque la acción del rey en la ronda 1 fue $enviar_{rey}(\langle rey\rangle, G)$, como $i \in G$ entonces $recibir_i(\langle rey\rangle, rey)) \in M^1_{r_i^\alpha}(1)$.

También se cumple $(\mathcal{I}, r_i^\alpha(1)) \models \neg K_i(c_i = 1) \wedge \ldots \wedge \neg K_i(c_i = r)$. Para todo color $b \neq a$ se cumple que $(\mathcal{I}, r_i^\alpha(1)) \models \neg(c_i = b)$ porque $(s_i, a) \in \alpha$. Para ver que también se cumple $(\mathcal{I}, r_i^\alpha(1)) \models \neg K_i(c_i = a)$ fijémonos en algún color b tal que $f_b^\alpha > 0$, que siempre existe por el lema 18 y en la ejecución $r^{\alpha_{i,b}}$ en $\mathcal{I}$ tal que en la primera ronda el rey realizó la acción $enviar_{rey}(\langle rey\rangle, G)$. Entonces $r_i^{\alpha_{i,b}}(1) = (1, C, vision_i(\alpha_{i,b}), (M^1_{r_i^{\alpha_{i,b}}(1)}))$ donde $M^1_{r_i^{\alpha_{i,b}}(1)} = \{recibir_i(\langle rey\rangle, rey))\}$. Por el lema 18 $vision_i(\alpha_{i,b}) = vision_i(\alpha)$ y entonces $r_i^{\alpha_{i,b}}(1) = r_i^\alpha(1)$, por lo que $(r^{\alpha_{i,b}}, 1) \sim_i (r^\alpha, 1)$. Como $(s_i, b) \in \alpha_{i,b}$ entonces se cumple $(\mathcal{I}, r_i^{\alpha_{i,b}}(1)) \models \neg(c_i = a)$ y entonces $(\mathcal{I}, r_i^\alpha(1)) \models \neg K_i(c_i = a)$.

Entonces se cumple $(\mathcal{I}, r_i^\alpha(t)) \models oyoalrey_i \wedge \neg K_i(c_i = 1) \wedge \ldots \wedge \neg K_i(c_i = r)$ por lo que $Pg_i^{\mathcal{I}}(r_i^\alpha(1)) = enviar_i(\langle no, i\rangle, R_i)$.□

Este lema es muy importante porque caracteriza cuándo una caja tiene colores inevitables. Veremos que esto será determinante si queremos que algún sabio sepa qué color tiene.

Definimos a continuación los distintos tipos de caja.

Definición 48 *Una caja $C = (\sharp c_1, \sharp c_2, \ldots, \sharp c_r)$ es de tipo* **bueno** *si tiene algún color inevitable. Una caja es de tipo* **malo** *si no tiene colores inevitables. C_n^B es el conjunto de cajas de tipo bueno y C_n^M el de tipo malo.*

Veamos qué ocurre si el rey usa una caja de tipo malo para el acertijo

Lema 20 *Si el rey usa una caja de tipo malo en $\mathcal{I}$, ningún sabio sabrá su color.*

Prueba

Sea una caja del tipo malo, $C \in C_n^M$, entonces C no tiene colores inevitables y por el lema 19 para cualquier ejecución consistente con $Pg^{\mathcal{I}}$, la respuesta de todos los sabios a la primera pregunta del rey es que no saben su color.

Esto pasa en todas las preguntas, es decir que en todas las rondas pares la acción conjunta que se realiza es:

$$b = (entregar_{ma}(todo), \Lambda, act_1, \ldots, act_n))$$

$$donde\ act_i = enviar_i(\langle no, i\rangle, R_i)\ o\ act_i = \Lambda$$

Por el lema 4 sabemos que en todas las rondas impares se realiza la acción conjunta

$$a = (entregar_{ma}(todo), enviar_{rey}(\langle rey\rangle, G'), \Lambda, \ldots, \Lambda)$$

Lo haremos por inducción sobre el número de ronda, para la ronda $2t$.

Base

Cuando $t = 1$ la demostración está en el lema 19.

Hipótesis de inducción

Supongamos que para cualquier ejecución r^α consistente con $Pg^{\mathcal{I}}$ en todas las rondas pares $2t'$ tales que $t' < t$ todos los sabios han respondido que no, es decir se ha realizado una acción del tipo de la acción conjunta b.

Paso Inductivo

Sea r^α una ejecución consistente con $Pg^{\mathcal{I}}$ tal que $\alpha \in \mathcal{CP}_C$. Supongamos que $(s_i, a) \in \alpha$. Como C es una caja de tipo malo entonces por el lema 18 existe $\alpha_{i,b} \in \mathcal{CP}_C$ para algún color $b \neq a$. Consideremos también la ejecución $r^{\alpha_{i,b}}$ consistente con $Pg^{\mathcal{I}}$ en la que las acciones del rey en las primeras $2t$ rondas has sido las mismas que las de r^α.

Por hipótesis de inducción hasta la ronda $2t$ todos los sabios a los que el rey les ha preguntado han contestado que no saben su color tanto en r^α como en $r^{\alpha_{i,b}}$. Fijémonos en el estado local del sabio i en $r^\alpha(2t-1)$ y en $r^{\alpha_{i,b}}(2t-1)$. Sabemos que en las primeras $2t-1$ rondas en ambas ejecuciones en las rondas impares se realizó la acción conjunta a y en las rondas pares se realizó la acción b, también ocurre que en cada ronda el rey le ha preguntado al mismo subconjunto de los sabios en ambas ejecuciones. Entonces la historia de mensajes del sabio i al tiempo $2t-1$ en ambas ejecuciones es la misma. Por el lema 7 se cumple $vision_i(\alpha) =$

$vision_i(\alpha_{i,b})$, entonces $r_i^\alpha(2t-1) = (2t-1, C, vision_i(\alpha), hist_{r_i^\alpha}(2t-1)) = (2t-1, C, vision_i(\alpha_{i,b}), hist_{r_i^\alpha}(2t-1)) = r_i^{\alpha_{i,c}}(2t-1)$.

Como la ronda anterior fue impar se realizó la acción a y entonces $recibir_i(\langle rey\rangle, rey)) \in M_{\rho_i^{rey}}^{2t-1}$, entonces se cumple $(\mathcal{I}, r_i^\alpha(2t-1)) \models oyoalrey_i$.

Para todo color $b \neq a$ se cumple $(\mathcal{I}, r_i^\alpha(2t-1)) \models \neg(c_i = b)$, entonces se cumple $(\mathcal{I}, r_i^\alpha(2t-1)) \models \neg K_i(c_i = b)$.

Como $(s_i, b) \in \alpha_{i,b}$ entonces se cumple $(\mathcal{I}, r_i^{\alpha_{i,b}}(2t-1)) \models \neg(c_i = a)$ y como $(r^\alpha, (2t-1)) \sim_i (r^{\alpha_{i,b}}, (2t-1))$ entonces $(\mathcal{I}, r_i^\alpha(2t-1)) \models \neg K_i(c_i = a)$.

Entonces se cumple $(\mathcal{I}, r_i^\alpha(2t-1)) \models oyoalrey_i \wedge \neg K_i(c_i = 1) \wedge \ldots \wedge \neg K_i(c_i = r)$ por lo que $Pg_i^{\mathcal{I}}(r_i^\alpha(2t-1)) = enviar_i(\langle no, i\rangle, R_i)$ para todo i.

Entonces en cualquier ejecución para una caja mala todos los sabios responden en todas las rondas impares que no saben su color.□

Los sabios, astutamente, se dieron cuenta de que si no había colores inevitables en la caja, el rey estaría burlándose de ellos y por eso le pidieron que hubiera alguna configuración en la que algún sabio pudiera saber su color a la primera pregunta. Ya vimos que esto ocurre si y sólo si hay algún color inevitable en la caja si y sólo si la caja es del tipo bueno. El rey acepta la condición pensando que puede salirse con la suya, logrando que los sabios no averiguen su color.

De ahora en adelante supondremos que la caja C es del tipo bueno. Lo que quiere decir que cualquier caja debe de tener un color inevitable. Como definimos que $\sharp c_1 \geq \sharp c_2 \geq \ldots \geq \sharp c_r$, el color 1 siempre es inevitable.

4.3 Los mínimos

En cualquier asignación de colores cuando mucho hay dif_d sabios con un color distinto a d para cualquier color inevitable d, todos los demás sabios deben tener un color d; el rey siempre tiene que asignar al menos el color d al menos a $n - dif_d$ sabios.

Definición 49 *Sea d un color inevitable, el mínimo de d, min_d, se define como*

$$min_d = (n - dif_d)$$

Veamos otra caracterización del mínimo que nos será útil.

Lema 21

$$min_d = \sharp c_d - k$$

Prueba Sea $C \in \mathcal{C}_n$ una caja tal que el color d es inevitable. Entonces

$$min_d = n - dif_d$$

Entonces

$$min_d + dif_d = n$$

Como $Tot_C = n + k$

$$min_d + dif_d = Tot_C - k$$

Pero también $Tot_C = \sharp c_d + dif_d$ por lo tanto

$$min_d + dif_d = \sharp c_d + dif_d - k$$

Y entonces

$$min_d = \sharp c_d - k. \square$$

En toda configuración de la caja, hay al menos min_d sabios con color d. Demostrémoslo formalmente:

Lema 22 *Para toda $\alpha \in CP_C$ y para todo color inevitable d se cumple que $p_d^\alpha \geq min_d$.*

Prueba Sea d un color inevitable. Se cumple que $p_b^\alpha \leq \sharp c_b$ para todo color $b \neq d$, entonces $\sum_{b \neq d} p_b^\alpha \leq dif_d$. Entonces se cumple que $p_d^\alpha + \sum_{b \neq d} p_b^\alpha \leq p_d^\alpha + dif_d$. Como $p_d^\alpha + \sum_{b \neq d} p_b^\alpha = n$ entonces $p_d^\alpha \geq n - dif_d = min_d$.□

Demostremos otras propiedades que resultarán útiles

Lema 23 *$min_d > 0$, para todo color inevitable.*

Prueba Sabemos que $dif_d < n$. Por lo tanto $min_d = n - dif_d > n - n = 0$, entonces $min_d > 0$.□

El rey está obligado a asignar siempre un color inevitable a algún sabio. De hecho está obligado a asignar el color d a min_d sabios, para cada color inevitable d. Es decir que es inevitable que haya min_d sabios con color d, de ahí el nombre.

Lema 24 *El color a es inevitable si y sólo si $\sharp c_a > k$*

Prueba

($\Rightarrow$) Si el color a es inevitable, entonces $min_a = \sharp c_a - k$ y por el lema 23 $\sharp c_a - k > 0$ y entonces $\sharp c_a > k$.

($\Leftarrow$) Si $\sharp c_a > k$, entonces $\sharp c_a + dif_a > dif_a + k$ y entonces como $Tot_C = \sharp c_a + dif_a$ se cumple que $Tot_C > dif_a + k$, entonces $Tot_C - k > dif_a$; pero también $Tot_C = n + k$, entonces $dif_a < n$, por lo que a es un color inevitable.□.

Volvamos a la situación en la que $k = 0$; en esta situación todos los colores son inevitables. Veamos que el rey está obligado a asignar el color d a $\sharp c_d$ sabios si y sólo si ocurre esta situación.

Lema 25 *Para todo color inevitable d, $\sharp c_d = min_d$, si y sólo si $Tot_C = n$.*

Prueba Sea d un color inevitable, $min(d) = \sharp c_d - k$ y $Tot_C = n + k$, entonces $min_d = \sharp c_d$ si y sólo si $k = 0$ si y sólo si $Tot_C = n$.□

4.4 Incluimos al conocimiento común

El rey lleva una caja en la que hay al menos un color inevitable d; por el lema 22, para toda configuración $\alpha \in \mathcal{CP}_C$, se cumple que el número de sabios con color d es al menos min_d. En la sección 2.2 hablamos del conocimiento común. Se demostró que en un punto (r,t) hay conocimiento común de un hecho si el hecho es verdadero en todos los puntos accesibles desde (r,t) (lema 1). Vimos además que todos los puntos $(r^\alpha, 0)$, correspondientes a configuraciones de una misma caja, forman una sola componente conexa (lema 11). Y en todos los puntos se cumple que hay al menos min_d sabios con cada color inevitable de la caja. Por lo tanto en cada punto es conocimiento común entre todos los sabios que hay al menos min_d sabios con color d para todo color d inevitable. Veamoslo formalmente:

Vamos a necesitar nuevas proposiciones primitivas que nos permitan describir propiedades del sistema.

Definición 50 *Aumentamos las proposiciones primitivas con el conjunto* Φ_S *donde*

$\Phi_S = \{(\sharp 1 \geq p), \ldots, (\sharp r \geq p)\}$ *donde* $k = 0, 1, 2, \ldots,$

Definimos también unas abreviaciones: $(\sharp a < p)$ por $\neg(\sharp a \geq p)$; $(\sharp a = p)$ por $((\sharp a \geq p) \wedge (\sharp a < p+1))$; para $a = 1, \ldots, r$; $p = 0, 1, 2, \ldots$.

La interpretación de estas proposiciones primitivas es que $(\sharp a \geq p)$ es verdadera en el punto (r^α, t) si en α hay al menos p sabios con color a.

Extendamos la interpretación de verdad

Definición 51 *Sea* $g = (\ell_{ma}, \ell_{rey}, \ell_1, \ldots, \ell_n)$ *un estado global que corresponde a la configuración* α

$\pi_A(g)((\sharp a \geq k)) =$ *verdadero syss*

$$p_a^\alpha \geq k$$

Si en un punto $(\sharp a \geq k)$ es verdadero, entonces será verdadero en todos los puntos de la ejecución, porque la propiedad no depende del tiempo, sólo depende de la configuración.

Lema 26 *Si* $(\mathcal{I}, r_i^\alpha(t)) \models (\sharp a \geq k)$ *entonces* $(\mathcal{I}, r_i^\alpha(t')) \models (\sharp a \geq k)$ *para todo* t'

Prueba

Si $(\mathcal{I}, r_i^\alpha(t)) \models (\sharp a \geq k)$ entonces $\pi_A(r^\alpha(t))((\sharp a \geq k)) = verdadero$. Entonces $p_a^\alpha \geq k$ por lo que $\pi_A(r^\alpha(t'))((\sharp a \geq k)) = verdadero$ para toda t' y entonces $(\mathcal{I}, r_i^\alpha(t')) \models (\sharp a \geq k)$ para toda t'.□

Veamos el resultado que mencionamos al principio de la sección

Lema 27 *Para todo punto* $(r^\alpha, 0)$ *tal que* $\alpha \in \mathcal{CP}_C$ *se cumple* $(\mathcal{I}, r^\alpha, 0) \models C_\mathcal{S}((\sharp d \geq min_d))$ *para todo color inevitable* d.

Prueba

Sea el punto $(r^\alpha, 0) \in \mathcal{I}$ tal que $\alpha \in \mathcal{CP}_C$. Por el lema 11, $(r^\beta, 0)$ es accesible desde $(r^\alpha, 0)$ si y sólo si $\beta \in \mathcal{CP}_C$. Por el lema 22 se cumple $p_d^\beta \geq min_d$ para todo color inevitable d. Entonces se cumple $(\mathcal{I}, r^\beta, 0) \models (\sharp d \geq min_d)$. Entonces $(\sharp d \geq min_d)$ se cumple para todo punto accesible desde $(r^\alpha, 0)$ por lo que por el lema 1 se cumple $(\mathcal{I}, r^\alpha, 0) \models C_\mathcal{S}((\sharp d \geq min_d))$.□

Para un color b, que no es inevitable, ocurre que hay configuraciones en las que el rey no asigna a algún sabio el color b. Si pensamos en el conocimiento común, esto quiere decir que al tiempo 0 no hay conocimiento común de que haya al menos un sabio con color b. Formalmente

Lema 28 *Para todo punto* $(r^\alpha, 0)$ *tal que* $\alpha \in \mathcal{CP}_C$ *se cumple* $(\mathcal{I}, r^\alpha, 0) \models \neg C_\mathcal{S}((\sharp b \geq 1))$ *para todo color* b *que no sea inevitable.*

Prueba

Sea el punto $(r^\alpha, 0) \in \mathcal{I}$ tal que $\alpha \in \mathcal{CP}_C$. Como b no es un color inevitable entonces se cumple $dif_b \geq n$, entonces es posible asignar a los n sabios colores distintos a b. Sea β una configuración en la que se cumple $p_b^\beta = 0$. Por el lema 11, $(r^\beta, 0)$ es accesible desde $(r^\alpha, 0)$ porque $\beta \in \mathcal{CP}_C$. Entonces se cumple $(\mathcal{I}, r^\beta, 0) \models \neg(\sharp b \geq 1)$. Entonces se cumple $(\mathcal{I}, r^\alpha, 0) \models \neg C_\mathcal{S}((\sharp b \geq 1))$.□

Por el lema 19, siempre hay una configuración en la que algún sabio sabe su color a la primera pregunta del rey. Esto ocurre exactamente en las configuraciones en las que hay exactamente min_d sabios con color d. Todos los sabios con color d saben su color. Si en la primera pregunta el rey le pregunta a un sabio con color inevitable y este sabio no sabe su color, esto quiere decir que hay más de min_d sabios con color d. Después de la primera pregunta del rey, si nadie supo su color, lo que ocurre es que se vuelve conocimiento común que al menos hay min_d+1 sabios con color d para todo color d inevitable, si es que el rey le preguntó a algún sabio con ese color. En general, en las dos versiones del acertijo en las que estamos interesados, si se cumple que hay conocimiento común de que hay al menos p sabios con color d, donde $p > 0$ y el rey le pregunta a algún sabio con color d y éste no sabe su color, entonces se vuelve conocimiento común que hay al menos $p + 1$ sabios con color d.

En la configuración α hay p_d^α sabios con color d; si en algún momento se vuelve conocimiento común que hay al menos p_d^α sabios con color d, como cualquier sabio con ese color ve que hay $p_d^\alpha - 1$ sabios con color d y sabe que debe haber uno más, en ese momento estos sabios saben su color.

Puede parecer extraño que un sabio que ve $p_d^\alpha - 1$ sabios con color d, en un inicio sólo "sepa" que hay al menos min_d sabios con este color. Lo importante es lo que el sabio sabe sobre lo que los demás sabios saben sobre lo que los demas sabios saben, etc. El sabio sabe que al menos hay $p_d^\alpha - 1$ sabios con color d, considera posible un mundo en el que él tiene un color distinto; en este otro mundo cada sabio con color d está viendo $p_d^\alpha - 2$ sabios con color d, cada uno de ellos considera posible un mundo en el que hay $p^\alpha - 2$ sabios con color d; así podemos seguir hasta llegar a un mundo en el que haya exactamente min_d sabios con color d, si el color d es inevitable ya vimos que no puede haber menos sabios con ese color, y ocurre que en ese mundo los sabios con color d ya no consideran posible otro mundo. En un principio hay conocimiento común de que hay al menos min_d sabios con color d. Los sabios posiblemente saben que hay más pero no saben si los demas saben que los demas saben, etc. Para un color b, evitable, la cadena continúa hasta que llegamos a un mundo en el que no hay algún sabio con color b y en ese mundo ningún sabio sabrá que su color es b porque no hay sabios con ese color

A grandes rasgos este será el razonamiento que siguen los sabios; veremos que en la configuración α un sabio sabrá su color cuando hay conocimiento común de que hay al menos p_d^α sabios con color d. En un principio sólo los sabios con color inevitable estan habilitados para saber su color. Pero veremos también que los demás sabios, los que tienen un color evitable, no están del todo perdidos. Un punto que resultará importante y que hasta ahora omitimos mencionar, es en cuál conjunto de sabios nos fijaremos para el conocimiento común; en la versión del acertijo de las esposas infieles el conocimiento común que nos interesa es entre el conjunto de todos los sabios, en la versión del acertijo de los tres sabios nos interesará el conocimiento común entre aquellos sabios a los que el rey no les ha preguntado.

Lema 29 *Sea α una configuración, si se cumple $(\mathcal{I}, r^\alpha, 2t+1) \models C_G((\sharp a \geq p_a^\alpha))$ entonces $(\mathcal{I}, r^\alpha, 2t+1) \models K_i((c_i = a))$, para todo sabio i tal que $i \in G$ y $(s_i, a) \in \alpha$.*

Prueba

Si $(\mathcal{I}, r^\alpha, 2t+1) \models C_G((\sharp a \geq p_a^\alpha))$ entonces $(\mathcal{I}, r^\beta, 2t+1) \models (\sharp a \geq p_a^\alpha)$ para todo $(r^\beta, 2t+1)$ G-accesible desde $(r^\alpha, 2t+1)$, sea $i \in G$ tal que $(s_i, a) \in \alpha$. Si $(r^\alpha, 2t+1) \sim_i (r^\beta, 2t+1)$ entonces por el lema 7 $\beta = \alpha_{i,b}$, pero si $b \neq a$ se cumple que $p_a^\beta = p_a^\alpha - 1$, entonces $(\mathcal{I}, r^\beta, 2t+1) \models (\sharp a < p_a^\alpha)$, una contradicción. Entonces Si $(r^\alpha, 2t+1) \not\sim_i (r^\beta, 2t+1)$ para todo $(r^\beta, 2t+1)$. Entonces por el lema 14 $(\mathcal{I}, r^\alpha, 2t+1) \models K_i((c_i = a))$.□

4.5 Las esposas infieles

Analizaremos primero lo que ocurre en la versión de las esposas infieles. El objetivo es llegar a entender el comportamiento de los sabios, para poder obtener un protocolo que implemente al programa $\boldsymbol{Pg}_A$ en el contexto interpretado (γ_{ei}, π_A), en el que los sabios puedan determinar sus acciones haciendo operaciones sobre sus estados locales.

Estamos suponiendo que todos los sabios razonan igual, que esto es conocimiento común, que es conocimiento común que todos reciben la misma información sobre la caja y que todos ven a todos; veremos entonces que en cada pregunta todos los sabios con un mismo color responden lo mismo. Este hecho nos va a servir para utilizar otra interpretación gráfica de las configuraciones posibles que nos va a ayudar para hacer los razonamientos.

4.5.1 Nuevas gráficas

Aprovechemos la idea intuitiva de que todos los sabios con un mismo color se comportan igual, para analizar qué es lo que ocurre en el acertijo. Lo que va a interesarnos es cuántos sabios hay con cada color en una configuración y no tanto el color del sabio i o el sabio j en particular. Vamos a fijarnos en cómo se comporta un sabio con color a y no tanto en cómo se comporta el sabio i que tiene color a. Este análisis va a resultar útil porque el número de configuraciones posibles va a reducirse y porque una vez que sepamos cómo se comportan los sabios, podremos regresar a pensar en que tenemos una configuración posible en la que el sabio j tiene color a, pero entonces ya sabremos cuáles son sus acciones.

Veamos primero un resultado que nos dice que dos puntos tendrán la misma historia exactamente cuando las acciones de todos los sabios han sido las mismas.

Lema 30 *Para cualesquiera dos puntos* $(r^\alpha, 2t+1)$ *y* $(r^\beta, 2t+1)$, *para un sabio* i *se cumple que* $hist_{r_i^\alpha}(2t+1) = hist_{r_i^\beta}(2t+1)$ *si y sólo si para todo sabio* j *se cumple que para toda* $t' < t$, $(\mathcal{I}_{ei}, r^\alpha, 2t'+1) \models K_j((c_j = e))$ *si y sólo si* $(\mathcal{I}_{ei}, r^\beta, 2t'+1) \models K_j((c_j = e))$

Prueba

($\Rightarrow$) Sean dos puntos $(r^\alpha, 2t+1)$ y $(r^\beta, 2t+1)$. Supongamos que para un sabio i se cumple $hist_{r_i^\alpha}(2t+1) = hist_{r_i^\beta}(2t+1)$, donde $hist_{r_i^\alpha}(2t+1) = (M^1_{r_j^\alpha}, \ldots, M^{2t+1}_{r_j^\alpha})$; $hist_{r_i^\beta}(2t+1) = (M^1_{r_j^\beta}, \ldots, M^{2t+1}_{r_j^\beta})$. Entonces $M^{t'}_{r_j^\alpha} = M^{t'}_{r_j^\beta}$ para toda $t' \leq 2t+1$ entonces se cumple que

$$recibir_i(msg, j) \in M^{t'}_{r_i^\alpha} \; syss \; recibir_i(msq, j) \in M^{t'}_{r_i^\beta}$$

$$enviar_i(msg, R_i) \in M^{t'}_{r_i^\alpha} \; syss \; enviar_i(msg, R_i) \in M^{t'}_{r_i^\beta}$$

Entonces sabemos que para todo sabio $j \neq i$

$$(\mathcal{I}_{ei}, r^\alpha, t') \models K_j((c_j = e))$$

si y sólo si

$enviar_j(\langle e, j\rangle, R_j) \in M^{t'}_{r_j^\alpha}$ si y sólo si $recibir_i(\langle e, j\rangle, j) \in M^{t'}_{r_i^\alpha}$ si y sólo si $recibir_i(\langle e, j\rangle, j) \in M^{t'}_{r_i^\beta}$ si y sólo si $enviar_j(\langle e, j\rangle, R_j) \in M^{t'}_{r_j^\beta}$

si y sólo si

$$(\mathcal{I}_{ei}, r^\beta, t') \models K_j((c_j = e))$$

Para el sabio i

$$(\mathcal{I}_{ei}, r^\alpha, t') \models K_j((c_j = e))$$

si y sólo si

$enviar_i(\langle e, j\rangle, R_i) \in M^{t'}_{r_j^\alpha}$ si y sólo si $enviar_i(\langle e, j\rangle, R_i) \in M^{t'}_{r_j^\beta}$

si y sólo si

$$(\mathcal{I}_{ei}, r^\beta, t') \models K_i((c_i = e))$$

($\Leftarrow$) Supongamos que para todo sabio j se cumple que para toda $t' < t$, $(\mathcal{I}_{ei}, r^\alpha, t') \models K_j((c_j = e))$ si y sólo si $(\mathcal{I}_{ei}, r^\beta, t') \models K_j((c_j = e))$.

Queremos demostrar que $hist_{r_i^\alpha}(2t+1) = hist_{r_i^\beta}(2t-1)$. Esto ocurre si y sólo si $M^{t'}_{r_i^\alpha} = M^{t'}_{r_i^\beta}$ para toda $t' \leq 2t+1$.

Los únicos eventos que pueden estar en $M^{t'}_{r_i^\alpha}$ y en $M^{t'}_{r_i^\beta}$ son

$$recibir_i(\langle e, j\rangle, j), recibir_i(\langle no, i\rangle, j), recibir_j(\langle rey\rangle, rey),$$
$$enviar_i(\langle e, i\rangle, R_i), enviar_i(\langle no, i\rangle, R_i)$$

Veamos los casos:

(a) $recibir_i(\langle e, j\rangle, j) \in M^{t'}_{r_i^\alpha}$ syss $enviar_j(\langle e, j\rangle, R_j) \in M^{t'}_{r_h^\alpha}$ syss $(\mathcal{I}_{ei}, r^\alpha, t') \models K_j((c_j = e))$ syss $(\mathcal{I}_{ei}, r^\beta, t') \models K_j((c_j = e))$ syss $enviar_j(\langle e, j\rangle, R_j) \in M^{t'}_{r_j^\beta}$ syss $recibir_i(\langle e, j\rangle, j) \in M^{t'}_{r_i^\beta}$

(b) $recibir_i(\langle no, i\rangle, j) \in M^{t'}_{r_i^\alpha}$ syss $enviar_j(\langle no, i\rangle, R_j) \in M^{t'}_{r_h^\alpha}$ syss $(\mathcal{I}_{ei}, r^\alpha, t') \models \neg K_j((c_j = e))$ syss $(\mathcal{I}_{ei}, r^\beta, t') \models \neg K_j((c_j = e))$ syss $enviar_j(\langle no, i\rangle, R_j) \in M^{t'}_{r_j^\beta}$ syss $recibir_i(\langle no, i\rangle, j) \in M^{t'}_{r_i^\beta}$

(c) $recibir_i(\langle rey\rangle, rey) \in M^{t'}_{r_i^\alpha}$ syss $recibir_i(\langle rey\rangle, rey) \in M^{t'}_{r_i^\beta}$

porque las acciones del rey siempre son las mismas en todas las ejecuciones.

(d) $enviar_i(\langle e, i\rangle, R_i) \in M^{t'}_{r_i^\alpha}$ syss $(\mathcal{I}_{ei}, r^\alpha, t') \models K_i((c_i = e))$ syss $(\mathcal{I}_{ei}, r^\beta, t') \models K_i((c_i = e))$ syss $enviar_i(\langle e, i\rangle, R_i) \in M^{t'}_{r_j^\beta}$.

(e) $enviar_i(\langle no, i\rangle, R_i) \in M^{t'}_{r_i^\alpha}$ syss $(\mathcal{I}_{ei}, r^\alpha, t') \models \neg K_i((c_i = e))$ syss $(\mathcal{I}_{ei}, r^\beta, t') \models \neg K_i((c_i = e))$ syss $enviar_i(\langle no, i\rangle, R_i) \in M^{t'}_{r_j^\beta}$.

Entonces $(M^{t'}_{r_i^\alpha} = M^{t'}_{r_i^\beta})$, para toda $t' \leq 2t+1$, por lo que $hist_{r_i^\alpha}(2t+1) = hist_{r_i^\beta}(2t+1)$.□

Veamos formalmente la idea de que los sabios con un mismo color se comportan igual.

Lema 31 *Si para dos configuraciones α_0, α_1 si se cumple que $p_a^{\alpha_0} = p_a^{\alpha_1}$ para $a = 1, \ldots, r$ y se cumple $(s_i, d) \in \alpha_0$, $(s_j, d) \in \alpha_1$ entonces para toda t se cumple que $(\mathcal{I}_{ei}, r^{\alpha_0}, 2t+1) \models K_i((c_i = d))$ si y sólo si $(\mathcal{I}_{ei}, r^{\alpha_1}, 2t+1) \models K_j((c_j = d))$.*

Prueba

Hagamoslo por inducción sobre t.

Base

Para $t = 0$. Se cumple que $(\mathcal{I}_{ei}, r^{\alpha_0}, 1) \models K_i((c_i = d))$ si y sólo si ocurre que $p_d^{\alpha_0} = min_d$ y $p_b^{\alpha_0} = \sharp c_b$ para todo color $b \neq d$; si y

sólo si $p_d^{\alpha_1} = min_d$ y $p_b^{\alpha_1} = \sharp c_b$ para todo color $b \neq d$ si y sólo si $(\mathcal{I}_{ei}, r^{\alpha_1}, 1) \models K_j((c_j = d))$.

Hipótesis de inducción

Supongamos que para cualesquiera dos configuraciones α_0, α_1 tales que se cumple que $p_a^{\alpha_0} = p_a^{\alpha_1}$ para $a = 1, \ldots, r$ y se cumple $(s_i, d) \in \alpha_0$, $(s_j, d) \in \alpha_1$, entonces ocurre que para toda $t' < t$, $(\mathcal{I}_{ei}, r^{\alpha_0}, 2t' + 1) \models K_i((c_i = d))$ si y sólo si $(\mathcal{I}_{ei}, r^{\alpha_1}, 2t' + 1) \models K_j((c_j = d))$.

Paso Inductivo

Hagámoslo por contrapuesta. Supongamos que ocurre que $(\mathcal{I}_{ei}, r^{\alpha_1}, 2t + 1) \models \neg K_j((c_j = d))$, entonces debe existir un punto $(r^{\beta_{j,c}}, 2t+1)$ tal que $(r^{\alpha_1}, 2t+1) \sim_j (r^{\beta_{j,c}}, 2t+1)$, entonces $r_j^{\alpha_1}(2t+1) = r_j^{\beta_{j,c}}(2t + 1)$. Entonces $hist_{r_j^{\alpha_1}}(2t + 1) = hist_{r_{j,c}^{\beta}}(2t + 1)$. Además se cumple que $vision_i(\alpha_1) = vision_i(\beta_1)$. Entonces por el lema 7 $\beta_{j,c} = ((\alpha_1)_{j,c})$, $c \neq d$. Entonces todos los sabios se han comportado igual hasta la ronda $2t$, es decir que para todo sabio i_j y toda $t' > t$ se cumple que $(\mathcal{I}_{ei}, r^{\alpha_1}, 2t' + 1) \models K_{i_j}((c_{i_j} = a))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta_{j,c}}, 2t' + 1) \models K_{i_j}((c_{i_j} = a))$.

Sea $\beta_{i,c}$ tal que $p_a^{\beta_{i,c}} = p_a^{\beta_{j,c}}$ para todo color a y que se cumpla que $(s_i, c) \in \beta_{i,c}$ y además que $vision_i(\beta_{i,c}) = vision_i(\alpha_0)$. Entonces $\beta_{i,c} = (\alpha_0)_{i,c}$. Por hipótesis de inducción se cumple que para toda $t' < t$ y toda g, h si se cumple $(s_g, e) \in \beta_{i,c}$ y $(s_h, e) \in \beta_{j,c}$ entonces se cumple que $(\mathcal{I}_{ei}, r^{\beta_{i,c}}, 2t' + 1) \models K_g((c_g = e))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta_{j,c}}, 2t' + 1) \models K_h((c_h = d))$. Es decir que hasta la ronda $2t$ todos los sabios con el mismo color en ambas configuraciones se comportan igual.

Fijémonos en un sabio i_j, supongamos que $i_j \neq i$ y que $(s_{i_j}, e) \in \alpha_0$, sea i_k tal que $(s_{i_k}, e) \in \alpha_1$. Por hipótesis de inducción sabemos que para toda $t' < t$, $(\mathcal{I}_{ei}, r^{\alpha_0}, 2t' + 1) \models K_{i_j}((c_{i_j} = e))$ si y sólo si $(\mathcal{I}_{ei}, r^{\alpha_1}, 2t' + 1) \models K_{i_k}((c_{i_k} = e))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta_{j,c}}, 2t' + 1) \models K_{i_k}((c_{i_k} = e))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta_{i,c}}, 2t' + 1) \models K_{i_j}((c_{i_j} = e))$ porque se cumple $(s_{i_j}, e) \in \beta_{i,c}$. Es decir que en las primeras $2t$ rondas todos los sabios en α_0 y en $\beta_{i,c}$ se comportan igual y entonces el sabio i no distingue entre ambos puntos por lo que siempre responde no. Entonces se cumple que $hist_{r_i^{\alpha_0}}(2t+1) = hist_{r_i^{\beta_{i,c}}}(2t+1)$ y entonces $(r^{\alpha_0}, 2t+1) \sim_i (r^{\beta_{i,c}}, 2t+1)$. Entonces se cumple que $(\mathcal{I}_{ei}, r^{\alpha_0}, 2t + 1) \models \neg K_i((c_i = d))$.

La demostración está terminada porque el regreso consistiría en intercambiar los índices entre i y j.□

A cada configuración α le asociamos a una r-ada de la forma $(p_1^\alpha, \ldots, p_r^\alpha)$, a esta r-ada la llamaremos *r-configuración* y la denotaremos como $[\alpha]$. Podemos ver a las r- configuraciones como clases de equivalencia de las configuraciones, donde

$$[\alpha] = \{\alpha_0 \mid p_a^{\alpha_0} = p_a^\alpha; a = 1, \ldots, r\}$$

Definiremos una gráfica con estas r-configuraciones de manera que habrá un arco dirigido de la r-configuración $[\alpha]$ a la r-configuración $[\beta]$, etiquetado con a_b, si $p_a^\alpha - 1 = p_a^\beta$, $p_b^\alpha + 1 = p_b^\beta$ y $p_e^\alpha = p_e^\beta$, para todos los demás colores. Entonces diremos que $[\alpha] \rightarrow_{a_b} [\beta]$. Es claro que entonces $[\beta] \rightarrow_{b_a} [\alpha]$. Intuitivamente lo que estamos haciendo es una operación similar a la que hacíamos para pasar de la configuración α en la que se cumple $(s_i, a) \in \alpha$, a la configuración $\alpha_{i,b}$. En general si ocurre que $[\alpha] \rightarrow_{a_b} [\beta]$ diremos que pasamos de $[\alpha]$ a $[\beta]$ *intercambiando* en $[\alpha]$ un color a por un color b.

Hay una relación entre la indistinguibilidad en configuraciones y la indistinguibilidad en r-configuraciones.

Lema 32 *$[\alpha] \rightarrow_{a_b} [\beta]$ si y sólo si para todo $\alpha_0 \in [\alpha]$ tal que $(s_i, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $vision_i(\alpha_0) = vision_i(\beta_0)$ y $(s_i, b) \in \beta_0$.*

Prueba

($\Rightarrow$) Supongamos que $[\alpha] \rightarrow_{a_b} [\beta]$. Sea $\alpha_0 \in [\alpha]$; entonces $p_a^{\alpha_0} - 1 = p_a^\beta$; entonces $p_a^{\alpha_0} \geq 1$ y por lo tanto existe i tal que $(s_i, a) \in \alpha_0$. Sea β_0 tal que $vision_i(\alpha_0) = vision_i(\beta_0)$, es decir que a todos los sabios les asignamos el mismo color que en α_0, podemos hacer que $(s_i, b) \in \beta$, porque como $p_b^{\alpha_0} + 1 = p_b^\beta$, entonces ocurre que $p_b^{\alpha_0} < \sharp c_b$ y entonces $f_b^{\alpha_0} > 0$, por el lema 9 entonces β_0 es una configuración posible. Entonces se cumple que $vision_i(\alpha_0) = vision_i(\beta_0)$ donde $(s_i, a) \in \alpha_0$ y $(s_i, b) \in \beta_0$.

($\Leftarrow$) Supongamos que para todo $\alpha_0 \in [\alpha]$ tal que $(s_i, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $vision_i(\alpha_0) = vision_i(\beta_0)$ y $(s_i, b) \in \beta_0$. Entonces ocurre que $p_a^{\alpha_0} - 1 = p_a^{\beta_0}$, $p_b^{\alpha_0} + 1 = p_b^{\beta_0}$ y $p_e^{\alpha_0} = p_e^{\beta_0}$, para cualquier otro color $e \neq a, b$. Como se cumple que $p_c^{\alpha_0} = p_c^\alpha$ y $p_c^{\beta_0} = p_c^\beta$ para todo color c, entonces ocurre que $p_a^\alpha - 1 = p_a^\beta$, $p_b^\alpha + 1 = p_b^\beta$ y $p_e^\alpha = p_e^\beta$ y entonces $[\alpha] \rightarrow_{a_b} [\beta]$. □

Entonces por el lema 7 se cumple que para todo $\alpha_0 \in [\alpha]$ tal que $(s_i, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $\beta_0 = (\alpha_0)_{i,b}$.

Corolario 4 *Para una r-configuración* $[\alpha]$, *existe* $[\beta]$ *tal que* $[\alpha] \rightarrow_{a_b} [\beta]$, *si y sólo si* $p_a^\alpha > 0$ *y* $f_b^\alpha > 0$

Prueba Dada una r-configuración $[\alpha]$ por el lema anterior existe $[\beta]$ tal que $[\alpha] \rightarrow_{a_b} [\beta]$ si y sólo si para todo $\alpha_0 \in [\alpha]$ tal que $(s_i, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $vision_i(\alpha_0) = vision_i(\beta_0)$ y $(s_i, b) \in \beta_0$, si y sólo si $p_a^{\alpha_0} > 0$ y $\beta_0 = (\alpha_0)_{i,b}$ si y sólo si, por el lema 9, $p_a^{\alpha_0} > 0$ y $f_b^{\alpha_0} > 0$, si y sólo si $p_a^\alpha > 0$ y $f_b^\alpha > 0$. □

Supongamos que hay en la caja h colores inevitables, sabemos que $\sharp c_d > k$ para $d = 1, \ldots, h$. Sabemos también que el rey está obligado a asignar el color d al menos a min_d sabios, para $d = 1, \ldots, h$. Pensemos que el rey ya asignó a estos sabios su color, es decir que asignó su color a $\sum_{d=1}^{h} min_d$ sabios. Para cada color inevitable el rey todavía puede asignar el color d a $\sharp c_d - min_d$ sabios. Pensemos en una nueva caja $C' = (\sharp c_1 - min_1, \ldots, \sharp c_h - min_h, \sharp c_{h+1}, \ldots, \sharp c_r)$. Como $min_d = \sharp c_d - k$, para $d = 1, \ldots, h$. Entonces $C' = (\underbrace{k, \ldots, k}_{h}, \sharp c_{h+1}, \ldots, \sharp c_r)$. Podemos pensar que la elección de la configuración inicia aquí; el rey debe escoger a cuántos sabios les asigna cada color teniendo la caja C'. Como $\sum_{d=1}^{h} min_d = \sum_{d=1}^{h} (\sharp c_d - k) = (\sum_{d=1}^{h} \sharp c_d) - hk = n + k - (\sum_{e=h+1}^{r} \sharp c_e) - hk = n - (\sum_{e=h+1}^{r} \sharp c_e) - (h-1)k$, entonces quedan $(\sum_{e=h+1}^{r} \sharp c_e) + (h-1)k$ sabios a los que el rey no les ha asignado su color. Para ello el rey tiene que usar las restricciones dadas por la caja C', a lo más puede asignar cada color inevitable d a k sabios y a lo más puede asignar cada color no inevitable b a $\sharp c_b$ sabios.

Desde el punto de vista del valor de f_a^α para cada color, como $f_a^\alpha = \sharp c_a - p_a^\alpha$, este número representa el número de sabios a los que el rey podría haber asignado el color a en α pero decidió no hacerlo. Dada la caja C' el rey tiene que escoger el número de sabios a los que asigna cada color, además de los que está obligado a asignar, o lo que es lo mismo debe decidir el valor de f_a^α para cada color a.

Entonces una manera equivalente de representar a una r-configuración será como $[\alpha] = (f_1^{[\alpha]}, \ldots, f_r^{[\alpha]})$. Sabiendo que $f_i^{[\alpha]} = \sharp c_i - p_i^{[\alpha]}$ podemos ir facilmente de una representación a otra. Además se cumple que $0 \leq f_d^\alpha \leq k$ para todos los colores inevitables y $0 \leq f_b^\alpha \leq \sharp c_b$ para los demás colores; donde ya sabemos que $\sum_{a=1}^{r} f_a^\alpha = k$.

4.5.2 Tetraedros

Para contar el número total de r-configuraciones, debemos contar las maneras en las que podemos organizar a los términos $f_a^{[\alpha]}$ de manera que $\sum_{a=1}^{r} f_a^{[\alpha]} = k$. Debemos contar de cuántas maneras podemos sumar r términos para que la suma sea k, donde nos importa el orden; esto tiene que ver con lo que se conoce como composiciones de k en r-partes.

Veamos primero un caso que nos ilustrará. Supongamos que para todos los colores se cumple que $\sharp c_b \geq k$. El número de estas r- configuraciones en este caso corresponde exactamente al número de composiciones de k en r-partes, esto es $\binom{r-1+k}{r-1}$.

Si pensamos que cada r-configuración representa una bola, entonces con todas las r configuraciones podemos formar una estructura en forma de un tetraedro $r-1$-dimensional. Diremos que con $\binom{r-1+k}{r-1}$ bolas podemos formar un tetraedro $r-1$-dimensional con $k+1$ niveles.

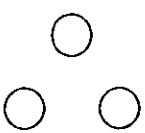
Veamos algunos ejemplos

Si $r = 2$ tenemos un "triángulo" formado por una hilera de $k+1$ bolas.

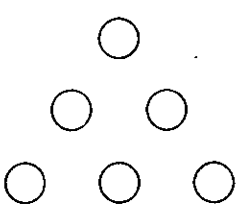
Supongamos que $r = 3$. El tetraedro 2-dimensional es un triángulo. Tenemos entonces $\binom{k+2}{2}$ bolas.

Si $k = 0$, tenemos sólo una bola:

Si $k = 1$, tenemos 3 bolas



Si $k = 2$, tenemos 6 bolas



Así, si $k = m$, agregamos $m + 1$ bolas al triángulo anterior; es decir que agregamos un tetraedro 1-dimensional con $m + 1$ niveles, y las ponemos abajo formando un triángulo con un nivel más.

Esto va a ocurrir en general para ir construyendo el tetraedro $r - 1$ dimensional. Si $k = 0$ tenemos una sóla bola. Para $k = 1$ agregamos el tetraedro $r - 2$ dimensional con 2 niveles. Para $k = 2$ agregamos a lo anterior el tetraedro $r - 2$ dimensional con 3 niveles. Podemos pensar que el tetraedro $r - 1$ dimensional con $k + 1$ niveles está formado por $k+1$ tetraedros $r-2$ dimensionales con $i+1$ niveles, para $i = 0, \ldots, k$.

Esto se confirma porque el número total de bolas en el tetraedro $r - 1$ dimensional con $k + 1$ niveles es $\binom{r-1+k}{r-1}$. Y cada tetraedro $r - 2$ dimensional con i niveles tiene $\binom{r-2+i}{r-2}$ bolas. Ocurre que $\binom{r-1+k}{r-1} = \sum_{i=1}^{k} \binom{r-2+i}{r-2}$

Regresando a las r configuraciones, si nos fijamos en un color d podemos organizar a las r-configuraciones por niveles, dependiendo de cuántos sabios con color d haya; esto nos indicará cómo construir la estructura en forma de tetraedro. En el nivel i estarán todas las r-configuraciones $[\alpha]$ en las que se cumpla $p_d^{[\alpha]} = min_d + i$, para $i = 0, \ldots, k$. En total habrá $k + 1$ niveles.

En general para contar cuántas r-configuraciones hay en el nivel i, debemos contar de cuántas maneras podemos sumar $r - 1$ términos para qwue la suma sea i, donde nos importa el orden; esto es porque el rey debe escoger los valores de $f_b^{[\alpha]}$ de manera que $\sum_{b \neq d} f_b^{[\alpha]} = i$, donde $0 \leq f_b^{[\alpha]} \leq i \leq k$. Tenemos composiciones de i en $r - 1$ partes. El número de estas composiciones es $\binom{r-2+i}{r-2}$. Este es el número de r-configuraciones en el nivel i y esto corresponde al número de bolas que podemos acomodar en un tetraedro $r - 2$ dimensional con i niveles.

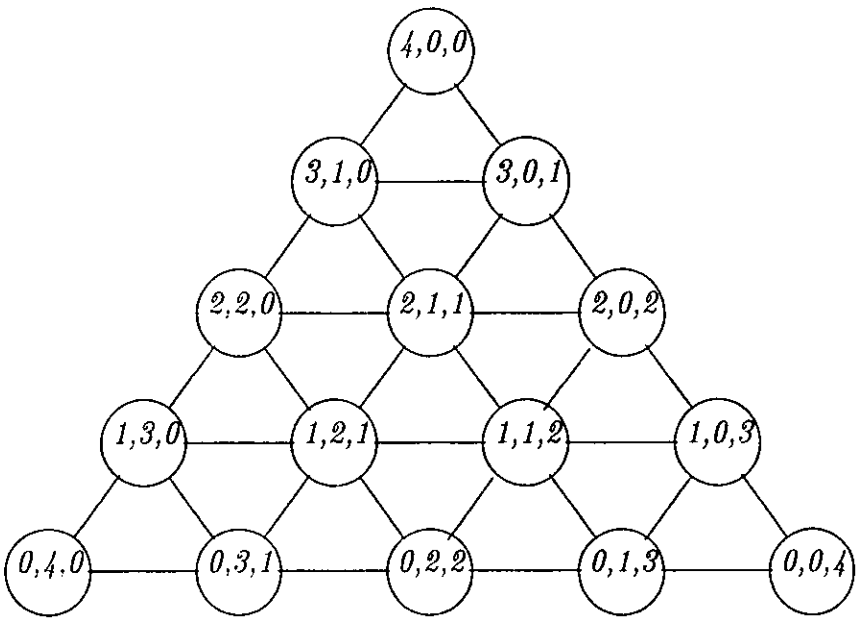
Así que las r-configuraciones formaran un tetraedro $r - 1$ dimensional con $k + 1$ niveles. Hasta ahora hemos supuesto que $\sharp c_b \geq k$ para todo color b.

Veamos un ejemplo

Ejemplo 3 *Supongamos que sabemos que hay tres colores en la caja,* $r = 3$*; que hay* $n = 6$ *sabios y que* $k = 4$.

Supongamos que los tres colores son inevitables; para cada r- *configuración* $[\alpha]$ *el rey debe escoger el valor de* $f_a^{[\alpha]}$ *para cada color* a*, de*

manera que $\sum_{a=1}^{3} f^{[\alpha]} = 4$. *Representamos a cada* 3*-configuración con los valores de* $f_a^{[\alpha]}$.



En este caso tenemos una estructura con $\binom{6}{2} = 15$, 3*- configuraciones en* 5 *niveles.*

Veamos qué pasa si no se cumple que $\sharp c_a \geq k$ para todo color a.

Sea el color a tal que $\sharp c_a < k$. Entonces el color a no es inevitable y no podemos decir que haya un número mínimo de sabios que tengan color a en cada r-configuración. Podemos pensar en la representación de las r-configuraciones con el valor de f_a^{α} para cada color. En el caso anterior sabíamos que $0 \leq f_a^{\alpha} \leq k$. Ahora ocurre que $0 \leq f_a^{\alpha} \leq \sharp c_a$. Podemos pensar que lo que falta ahora son los niveles en los que se cumple $f_a^{\alpha} = \sharp c_a + 1, \sharp c_a + 2, \ldots, k$. Estas r-configuraciones que faltan corresponden a un tetraedro $r-1$ dimensional con $k - \sharp c_a$ niveles. Estamos quitando una esquina al tetraedro $r-1$ dimensional.

En general la estructura que se va a formar es un tetraedro $r-1$-dimensional al que le van a faltar algunas esquinas, si hay colores evitables para los que se cumple $\sharp c_i < k$. Estas esquinas van a corresponder a su vez a tetraedros $r-1$ dimensionales con $k - \sharp c_a$ niveles.

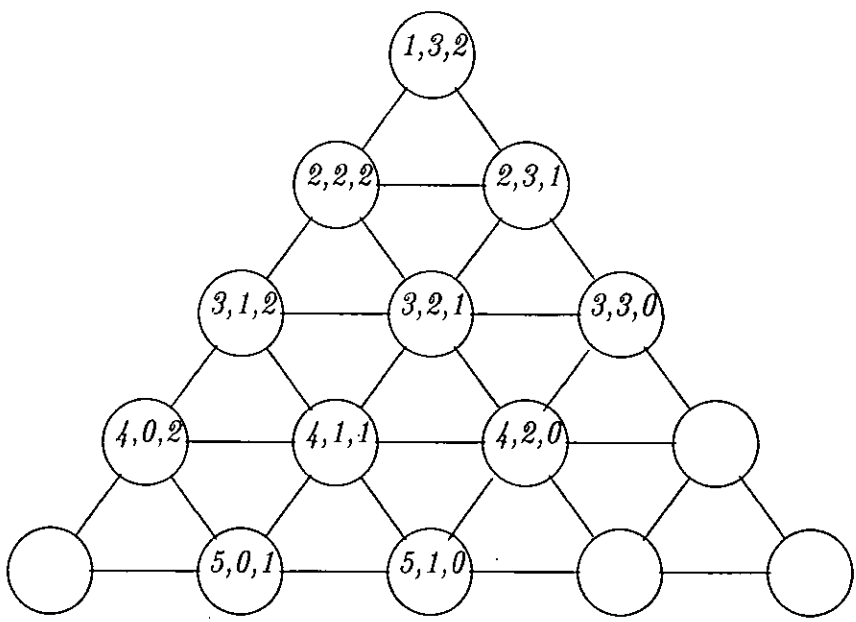
Veamos ahora un ejemplo más concreto:

Ejemplo 4 *Supongamos que tenemos la caja* $C = (5, 3, 2)$ *y hay* $n = 6$ *sabios, entonces* $k = 4$. *Sólo el color* 1 *es inevitable. Puede verificarse que hay* 422 *configuraciones posibles para esta caja, simplemente pensar en dónde poner el dibujo de la gráfica de mundos posibles ya es un problema. Si en vez de las configuraciones posibles ponemos a las r-configuraciones posibles habrá sólo* 11 *de éstas que serían:*

$(1,3,2); (2,2,2); (2,3,1); (3,1,2); (3,2,1); (3,3,0); (4,0,2);$
$(4,1,1); (4,2,0); (5,0,1); (5,1,0).$

Nótese que estamos representando a las configuraciones con el valor de $p_a^{[\alpha]}$.

Veámos la nueva gráfica.



Los nodos que no tienen etiqueta representan a las esquinas faltantes que mencionábamos; lo que ocurre es en el momennto en que llegamos a alguna 3*-configuración en la que hay* 0 *sabios con algún color a ya no podemos intercambiar más algún otro color por este color a. Los nodos faltantes corresponden a tetraedros* 2*-dimensionales; en uno hay* $k - \sharp c_2 = 4 - 3 = 1$ *nivel y un nodo, en el otro hay* $4 - 2 = 2$ *niveles y tres nodos.*

4.5.3 Los colores responden

Lo importante de la gráfica de r-configuraciones es que podemos analizar como razonan los sabios. Nos va a servir, sobre todo, porque es fácil imaginársela y también es más fácil de dibujar. Podemos pensar que las ejecuciones del protocolo $Pg_A^{\mathcal{I}_{\mathcal{C}'}}$ se llevan a cabo sobre la gráfica de r-configuraciones. Hablaremos de la ejecución $r^{[\alpha]}$ y del punto de la ejecución $(r^{[\alpha]}, t)$. La ejecución $r^{[\alpha]}$ representará el comportamiento de los sabios en toda configuración $\alpha_0 \in [\alpha]$. En cada punto $(r^{[\alpha]}, t)$ se guardará el estado local del color a, para $a = 1, \ldots, r$, siendo este estado local igual a la r-configuración $[\alpha]$ y a una historia de respuestas.

Para definir la historia de respuestas hay un detalle importante. De hecho esta historia de respuestas refleja lo que ocurre en la historia de mensajes que habíamos definido anteriormente. Pensemos en una configuración α en la que se cumple $p_a^\alpha = 1$ para algún color a, supongamos que $(s_i, a) \in \alpha$, es decir que el sabio i es el único sabio con color a; en la historia de mensajes del sabio i no se va a registrar ningún mensaje de un sabio con color a porque el sabio i no oye la respuesta de ningún sabio con color a, lo que nos interesa son las respuestas que oye un sabio porque sus respuestas no van a provocar que distinga entre dos puntos. Para ser consistentes con esta situación, en un punto $(r^{[\alpha]}, 2t+1)$, vamos a definir a la historia de respuestas del color a como una sucesión de r-adas de la forma $R(2t'+1) = (res_1, \ldots, res_r)$, $t' < t$.

Definición 52 *La* historia de respuestas *del color a en el punto $(r^{[\alpha]}, 2t+1)$, es una sucesión $R_a^{r^{[\alpha]}}(2t+1) = (R(1), \ldots, R(2t+1))$, tal que para toda $t' \leq t$ $R(2t'+1) = (res_1, \ldots, res_r)$, donde res_b, $b \neq a$, va a ser igual a la respuesta del color b en la ronda $2t'$, si es que $p_b^{[\alpha]} > 0$. Si $p_b^{[\alpha]} = 0$, entonces ningún sabio tiene color b en α y entonces diremos que $res_b = \Lambda$; res_a va a ser igual a la respuesta del color a en la ronda $2t'$ si es que $p_a^{[\alpha]} > 1$, si $p_a^{[\alpha]} \leq 1$ entonces $res_a = \Lambda$.*

Con saber la respuestas de un color en cada ronda podemos saber las respuestas de cada sabio con ese color. Definimos entonces la relación entre los puntos

Definición 53 $(r^{[\alpha]}, 2t+1) \rightarrow_{a_b} (r^{[\beta]}, 2t+1)$ *si* $[\alpha] \rightarrow_{a_b} [\beta]$ *y si* $R_a^{r^{[\alpha]}}(2t+1) = R_b^{r^{[\beta]}}(2t+1)$.

No tendrá sentido definir la historia de respuestas de un color tal que $p_a^{[\alpha]} = 0$. Esto debido a que no puede ocurrir que $[\alpha] \rightarrow_{a_b} [\beta]$ para alguna $[\beta]$ porque si esto pasara, como $p_a^{[\alpha]} - 1 = p_a^{[\beta]}$, entonces $p_a^{[\beta]} < 0$, lo que no es posible. Entonces tampoco puede ocurrir $[\beta] \rightarrow_{b_a} [\alpha]$. Nunca nos vamos a fijar en la historia de respuestas del color a.

El siguiente lema nos dice que los sabios con un mismo color se comportan igual, en todas las configuraciones que pertenecen a la misma clase de equivalencia definida por una r-configuración.

Lema 33 *Si para una configuración α tal que $(s_i, a) \in \alpha$ se cumple $(r^\alpha, 2t+1) \sim_i (r^\beta, 2t+1)$ donde $(s_i, b) \in \beta$ entonces para toda $\alpha_0 \in [\alpha]$ tal que $(s_j, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $(s_j, b) \in \beta_0$ y $(r^{\alpha_0}, 2t+1) \sim_j (r^{\beta_0}, 2t+1)$*

Prueba

Se va a cumplir que $(r^{\alpha_0}, 2t+1) \sim_j (r^{\beta_0}, 2t+1)$ si y sólo si $r_j^{\alpha_0}(2t+1) = r_j^{\beta_0}(2t+1)$, si y sólo si $vision_j(\alpha_0) = vision_j(\beta_0)$, e $hist_{r_j^{\alpha_0}}(2t+1) = hist_{r_j^{\beta_0}}(2t+1)$. Esto último se cumple si y sólo si, por el lema 30, para todo sabio i_j se cumple que para toda $t' < t$, $(\mathcal{I}_{ei}, r^{\alpha_0}, 2t'+1) \models K_{i_j}((c_{i_j} = e))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta_0}, 2t'+1) \models K_{i_j}((c_{i_j} = e))$

Entonces sea β_0 tal que $\beta_0 \in [\beta]$ y $vision_i(\alpha_0) = vision_i(\beta_0)$, $(s_j, b) \in \beta_0$

Sea el sabio $i_j \neq j$ y sea $t' < t$. Supongamos que $(s_{i_j}, e) \in \alpha_0$. Entonces se cumple $(s_{i_j}, e) \in \beta_0$. Sea el sabio i_g tal que $(s_{i_g}, e) \in \alpha$ y $(s_{i_g}, e) \in \beta$. Entonces se cumple que:

$(\mathcal{I}_{ei}, r^{\alpha_0}, 2t'+1) \models K_{i_j}((c_{i_j} = e))$ si y sólo si, por el lema 31, $(\mathcal{I}_{ei}, r^{\alpha}, 2t+1) \models K_{i_g}((c_{i_g} = b))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta}, 2t'+1) \models K_{i_g}((c_{i_g} = e))$ si y sólo si $(\mathcal{I}_{ei}, r^{\beta_0}, 2t'+1) \models K_{i_j}((c_{i_j} = e))$.

Entonces en cada ronda todos los sabios distintos a j responden igual. Esto provoca que el sabio j no distinga entre $(r^{\alpha_0}, 2t'+1)$ y $(r^{\beta_0}, 2t'+1)$ para toda $t' < t$ y entonces la respuesta del sabio en ambos puntos siempre es que no sabe su color.

Entonces $hist_{r_j^{\alpha_0}}(2t+1) = hist_{r_j^{\beta_0}}(2t+1)$ y entonces $(r^{\alpha_0}, 2t+1) \sim_j (r^{\beta_0}, 2t+1)$. □.

Veamos unos resultados sobre las ejecuciones en las r-configuraciones. En particular nos interesa saber cuándo sigue habiendo flechas entre las r-configuraciones. Esto representa que los sabios no distinguen.

Lema 34 $(r^{[\alpha]}, 2t+1) \rightarrow_{a_b} (r^{[\beta]}, 2t+1)$ *si y sólo si para toda* $\alpha_0 \in [\alpha]$ *tal que* $(s_i, a) \in \alpha_0$ *existe* $\beta_0 \in [\beta]$ *tal que* $(s_i, b) \in \beta_0$ *y* $(r^{\alpha_0}, 2t+1) \sim_i (r^{\beta_0}, 2t+1)$

Prueba

Se cumple $(r^{[\alpha]}, 2t+1) \rightarrow_{a_b} (r^{[\beta]}, 2t+1)$ si y sólo si $[\alpha] \rightarrow_{a_b} [\beta]$ y la historia de respuestas del color a en $(r^{[\alpha]}, 2t+1)$ es igual a la historia de respuestas del color b en $(r^{[\beta]}, 2t+1)$. Si y sólo si por el lema 32 para todo $\alpha_0 \in [\alpha]$ tal que $(s_i, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $vision_i(\alpha_0) = vision_i(\beta_0)$ y $(s_i, b) \in \beta_0$ y además lo que oyó en sabio i en α_0 en las primeras $2t$ rondas es lo mismo que lo que oyó el sabio i en β_0 en las primeras $2t$ rondas, si y sólo si $vision_i(\alpha_0) = vision_i(\beta_0)$ e $hist_{r_i^{\alpha_0}}(2t+1) = hist_{r_i^{\beta_0}}(2t+1)$ si y sólo si $(r^{\alpha_0}, 2t+1) \sim_i (r^{\beta_0}, 2t+1)$. □

Veamos ahora la relación entre la nueva gráfica y lo que saben los sabios. También va a ocurrir que un sabio sabe su color si y sólo si en la r-configuración no hay una flecha que salga, etiquetada con ese color.

Lema 35 *Supongamos que se cumple* $(s_i, a) \in \alpha$, *entonces* $(\mathcal{I}_{ei}, r^{\alpha}, 2t+1) \models K_i((c_i = a))$ *si y sólo si* $(r^{[\alpha]}, 2t+1) \not\rightarrow_{a_b} (r^{[\beta]}, 2t+1)$ *para toda r-configuración* $[\beta]$ *y para todo color* $b \neq a$

Prueba

($\Rightarrow$) Si se cumple $(s_i, a) \in \alpha$ y $(\mathcal{I}_{ei}, r^{\alpha}, 2t+1) \models K_i((c_i = a))$ entonces por el lema 14 se cumple que para todo otro punto $(r^{\beta}, 2t+1)$ ocurre que $(r^{\alpha}, 2t+1) \not\sim_i (r^{\beta}, 2t+1)$ entonces por el lema anterior $(r^{[\alpha]}, 2t+1) \not\rightarrow_{a_b} (r^{[\beta]}, 2t+1)$ para toda r-configuración $[\beta]$ y para todo color $b \neq a$

($\Leftarrow$) Demostremos la contrapuesta. Si se cumple $(s_i, a) \in \alpha$ y $(\mathcal{I}_{ei}, r^{\alpha}, 2t+1) \models \neg K_i((c_i = a))$ entonces existe $(r^{\beta}, 2t+1)$ tal que $(r^{\alpha}, 2t+1) \sim_i (r^{\beta}, 2t+1)$, donde $(s_i, b) \in \beta$. Entonces por el lema 33 se cumple que $\alpha_0 \in [\alpha]$ tal que $(s_j, a) \in \alpha_0$ existe $\beta_0 \in [\beta]$ tal que $(s_j, b) \in \beta_0$ y $(r^{\alpha_0}, 2t+1) \sim_j (r^{\beta_0}, 2t+1)$. Entonces por el lema 34 $(r^{[\alpha]}, 2t+1) \rightarrow_{a_b} (r^{[\beta]}, 2t+1)$. □

Esto nos dice que el sabio i va a saber su color en $(r^{\alpha_0}, 2t+1)$ si y sólo si $(s_i, a) \in \alpha$ y en el punto correspondiente $(r^{[\alpha]}, 2t+1)$ no hay alguna flecha que salga de este punto etiquetada con a_b para algún otro color b. De hecho así es como vamos a pensar una ejecución del acertijo de las esposas infieles. A partir de la gráfica de las r-configuraciones, todos los puntos en los que se cumple $p_d^\alpha = min_d$ para algún color inevitable d son los únicos en los que los sabios con color d pueden saber su color en la primera pregunta. Si no lo saben quiere decir que estos puntos no representan a la configuración ρ que el rey escogió y entonces debemos quitarlos. Esto provoca que haya nuevos puntos para los que no haya una flecha etiquetada con d_b y entonces hay nuevos puntos en los que los sabios con color inevitable pueden saber su color. En la siguiente sección veremos con más detalle que ocurre en general.

4.5.4 El conocimiento común se mueve

En la versión de las esposas infieles se cumple el regreso del lema 29, si un sabio sabe que su color es a en la configuración α entonces en ese momento debe ser conocimiento común, entre todos los sabios, que hay al menos p_a^α sabios con color a.

Teorema 4 *Sea α una configuración. Si se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t+1) \models K_i((c_i = a))$ entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t+1) \models C_S((\sharp a \geq p_a^\alpha))$.*

Prueba

Lo haremos por inducción sobre t.

Base

Para $t = 0$. Si $(\mathcal{I}_{ei}, r^\alpha, 1) \models K_i((c_i = a))$ entonces se cumple que $f_a^\alpha = k$; $f_b^\alpha = 0$ para $b \neq a$. Entonces $p_b^\alpha = \sharp c_b$ para todo color $b \neq a$. Entonces $\sum_{b \neq a} p_b^\alpha = dif_a$ sabios con color distinto a a, entonces $dif_a < n$, entonces a es un color inevitable y se cumple y $p_a^\alpha = n - dif_a = min_a$. Por el lema 27 se cumple $(\mathcal{I}_{ei}, r^\alpha, 0) \models C_S((\sharp a \geq min_a))$ y por lo tanto se cumple $(\mathcal{I}_{ei}, r^\alpha, 0) \models C_S((\sharp a \geq p_a^\alpha))$, entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 1) \models C_S((\sharp a \geq p_a^\alpha))$.

Hipótesis de inducción

La hipótesis de inducción es que para toda $t' < t$ se cumple que si $(\mathcal{I}_{ei}, r^\alpha, 2t'+1) \models K_i((c_i = a))$ entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t'+1) \models C_S((\sharp a \geq p_a^\alpha))$.

Entonces por el lema 29 ocurre que para toda $t' < t$, $(\mathcal{I}, r^\alpha, 2t'+1) \models K_i((c_i = a))$ si y sólo si $(\mathcal{I}_{ei}, r^\alpha, 2t'+1) \models C_S((\sharp a \geq p_a^\alpha))$.

Paso Inductivo

Supongamos que se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t+1) \models K_i((c_i = a))$. Entonces por el lema 35 se cumple que $(r^{[\alpha]}, 2t+1) \not\rightarrow_{a_b} (r^{[\beta]}, 2t+1)$ para toda r-configuración $[\beta]$ y para todo color $b \neq a$. En un principio se cumple $(r^\alpha, 0) \sim_i (r^{\alpha_{i,b}}, 0)$ para todo color b tal que $f_b^\alpha > 0$. Si no existe algún color así entonces el color a es inevitable y estamos en una situación igual a la de la base en la que $p_a^\alpha = min_a$ y entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t+1) \models C_S((\sharp a \geq p_a^\alpha))$. Supongamos que sí existe algún color b tal que $f_b^\alpha > 0$. Sabemos que se cumple $(r^{[\alpha]}, 2t+1) \not\rightarrow_{a_b} (r^{[\alpha_{i,b}]}, 2t+1)$. Entonces para cada ejecución de la forma $r^{\alpha_{i,b}}$ tuvo que ocurrir que para algún $t_b < t$, $(r^\alpha, 2t_b - 1) \sim_i (r^{\alpha_{i,b}}, 2t_b - 1)$ pero $(r^\alpha, 2t_b + 1) \not\sim_i (r^{\alpha_{i,b}}, 2t_b + 1)$. Veamos porque pudo pasar esto. La diferencia debe estar en la historia de mensajes, es decir que la historia de mensajes del sabio i en $(r^\alpha, 2t_b - 1)$ es igual que la historia de mensajes del sabio i en $(r^{\alpha_{i,b}}, 2t_b - 1)$, pero esto ya no ocurre al tiempo $2t_b + 1$.

Para todo color $c \neq a, b$ se cumple $p_c^\alpha = p_c^{\alpha_{i,b}}$. Por hipótesis de inducción para todo $t' < t_b < t$ se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t'+1) \models K_j((c_j = c))$ syss $(\mathcal{I}_{ei}, r^\alpha, 2t'+1) \models C_S((\sharp c \geq p_c^\alpha))$ syss $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t'+1) \models C_S((\sharp a \geq p_c^{\alpha_{i,b}}))$ syss $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t'+1) \models K_j((c_j = c))$. Es decir que en las ejecuciones r^α y $r^{\alpha_{i,b}}$ en las primeras $2t_b$ rondas las respuestas de todos los colores distintos a a y b son las mismas.

Como $(r^\alpha, 2t_b - 1) \sim_i (r^{\alpha_{i,b}}, 2t_b - 1)$ entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t_b - 1) \models \neg K_i((c_i = a))$ y entonces, como todos los sabios con el mismo color se comportan igual, todo sabio j tal que $(s_j, a) \in \alpha$ no sabe su color en la ronda $2t_b$.

También se cumple que $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t_b - 1) \models \neg K_j((c_j = b))$ para todo sabio j tal que $(s_j, b) \in \alpha_{i,b}$. Entonces todos los sabios con color b responden que no saben en la ronda $2t_b$

Entonces la diferencia entre $(r^{\alpha_{i,b}}, 2t_b - 1)$ y $(r^\alpha, 2t_b - 1)$ debe ser que los sabios con color b en $(r^\alpha, 2t_b - 1)$ sepan su color o que los sabios con color a en $(r^{\alpha_{i,b}}, 2t_b - 1)$ sepan su color. Sabemos que $p_a^{\alpha_{i,b}} = p_a^\alpha - 1$.

Si los sabios con color a en $\alpha_{i,b}$ responden que sí saben su color ocurre que $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t_b - 1) \models K_j((c_j = a))$ para todo j tal que $(s_j, a) \in \alpha_{i,b}$ y entonces por hipótesis de inducción se cumple $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t_b - 1) \models$

$C_S((\sharp a \geq p_a^{\alpha_{i,b}}))$ y entonces se cumple $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t_b - 1) \models C_S((\sharp a \geq p_a^{\alpha} - 1))$.

Entonces para cualquier punto $(r^\gamma, 2t_b + 1)$ accesible desde $(r^\alpha, 2t_b + 1)$ se cumple que en el tiempo $2t_b - 1$ los sabios con color a no supieron su el color porque por el lema 13 la historia de mensajes en $(r^\gamma, 2t_b - 1)$ y en $(r^\alpha, 2t_b - 1)$ es la misma, es decir que se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models \neg K_j((c_j = a))$. Sabemos que se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models C_S((\sharp a \geq p_a^\alpha - 1))$ porque esto se cumple en $(r^\alpha, 2t_b - 1)$, si suponemos que en γ hay $p_a^\alpha - 1$ sabios con color a, entonces se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models (\sharp a = p_a^\alpha - 1)$ pero entonces por hipótesis de inducción se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t - 1) \models K_j((c_j = a))$ para todo sabio con color a, una contradicción. Entonces en γ hay al menos p_a^α sabios con color a por lo que se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models (\sharp a \geq p_a^\alpha)$ y entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t_b + 1) \models C_S((\sharp a \geq p_a^\alpha))$ y como $t_b < t$ entonces $(\mathcal{I}_{ei}, r^\alpha, 2t + 1) \models C_S((\sharp a \geq p_a^\alpha))$.

Supongamos que para toda ejecución $r^{\alpha_{i,b}}$ se cumple $(\mathcal{I}_{ei}, r^{\alpha_{i,b}}, 2t_b - 1) \models \neg K_j((c_j = a))$ para todo j tal que $(s_j, a) \in \alpha_{i,b}$, entonces debe ocurrir que los sabios con color b en α supieron su color, es decir $(\mathcal{I}_{ei}, r^\alpha, 2t_b - 1) \models K_j((c_j = b))$ para todo sabio j tal que $(s_j, b) \in \alpha$. Entonces por hipótesis de inducción se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t_b - 1) \models C_S((\sharp b \geq p_b^{\alpha_{i,b}}))$.

Sea un punto $(r^\gamma, 2t_b + 1)$ accesible desde $(r^\alpha, 2t_b + 1)$, entonces se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models K_i((c_j = b))$ para todo sabio j tal que $(s_j, b) \in \gamma$ porque el comportamiento de los sabios en r^γ debe ser el mismo que en r^α en las primeras $2t_b$ rondas. También se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models C_S((\sharp b \geq p_b^{\alpha_{i,b}}))$ y se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models \neg C_S((\sharp b \geq p_b^{\alpha_{i,b}} + 1))$ porque $(r^\alpha, 2t_b - 1)$ es un punto accesible en el que hay $p_b^{\alpha_{i,b}}$ sabios con color b. Entonces en γ debe ocurrir que $p_b^\gamma = p_b^{\alpha_{i,b}}$, porque si hubiera más sabios con color b por hipótesis de inducción los sabios con este color no habrían podido saber su color. Entonces se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b - 1) \models (\sharp b = p_b^{\alpha_{i,b}})$ y entonces también se cumple $(\mathcal{I}_{ei}, r^\gamma, 2t_b + 1) \models (\sharp b = p_b^{\alpha_{i,b}})$ por lo que entonces se cumple $(\mathcal{I}_{ei}, r^\alpha, 2t_b + 1) \models C_S((\sharp b = p_b^{\alpha_{i,b}}))$ para todo color b tal que $f_b^\alpha > 0$.

Entonces en todo punto $(r^\gamma, 2t_b + 1)$ accesible desde $(r^\alpha, 2t_b + 1)$ se cumple que hay $p_b^{\alpha_{i,b}}$ sabios con color b, entonces *como* $f_b^\gamma = \sharp c_b - p_b^{\alpha_{i,b}} = f_b^\alpha$ para todo color b tal que $f_b^\alpha > 0$. Sabemos que $\sum_{i=1}^r f_i^\alpha = \sum_{i=1}^r f_i^\gamma = k$. Sean $b_1, \ldots, b_s$ los colores tales que $f_{b_i}^\alpha > 0$ y sean $c_1, \ldots, c_u$ los

colores tales que $f_{c_i}^{\alpha} = 0$. Entonces $f_a^{\alpha} + \sum_{i=1}^{s} f_{b_i}^{\alpha} = k$. Entonces $k = \sum_{i=1}^{r} f_i^{\gamma} = f_a^{\gamma} + \sum_{i=1}^{s} f_{b_i}^{\gamma} + \sum_{i=1}^{u} f_{c_i}^{\gamma} = f_a^{\alpha} + \sum_{i=1}^{s} f_{b_i}^{\alpha}$. Entonces $f_a^{\gamma} + \sum_{i=1}^{u} f_{c_i}^{\alpha} = f_a^{\alpha}$. Entonces $f_a^{\gamma} \leq f_a^{\alpha}$. Como $f_a^{\alpha} = \sharp c_a - p_a^{\alpha}$ entonces $p_a^{\alpha} \leq \sharp c_a - f_a^{\gamma} = p_a^{\gamma}$. Entonces se cumple $(\mathcal{I}_{ei}, r^{\gamma}, 2t_b + 1) \models (\sharp a \geq p_a^{\alpha})$ y entonces se cumple $(\mathcal{I}_{ei}, r^{\alpha}, 2t + 1) \models C_{\mathcal{S}}((\sharp a \geq p_a^{\alpha}))$.□

Cuando los sabios con algún color a saben su color, se vuelve conocimiento común el número de sabios que hay en total con ese color.

Lema 36 *Si ocurre que* $(\mathcal{I}_{ei}, r^{\rho}, 2t - 1) \models K_i((c_i = a))$ *entonces se cumple* $(\mathcal{I}_{ei}, r^{\alpha}, 2t' - 1) \models C_{\mathcal{S}}((\sharp a = p_a^{\rho}))$ *para toda* $t' > t$.

Prueba

Supongamos que se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t-1) \models K_i((c_i = a))$, entonces por el lema 4 se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t-1) \models C_{\mathcal{S}}((\sharp a \geq p_a^{\rho}))$. Para cualquier punto $(r^{\beta}, 2t + 1)$ accesible desde $(r^{\rho}, 2t + 1)$ debe ocurrir que todos los sabios hayan respondido igual hasta ese momento y entonces se debe cumplir $(\mathcal{I}_{ei}, r^{\beta}, 2t - 1) \models K_i((c_i = a))$ y también se debe cumplir $(\mathcal{I}_{ei}, r^{\beta}, 2t - 1) \models C_{\mathcal{S}}((\sharp a \geq p_a^{\beta}))$, entonces no puede pasar que $p_a^{\rho} < p_a^{\beta}$, pero tampoco puede pasar $p_a^{\rho} > p_a^{\beta}$ porque se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t - 1) \models C_{\mathcal{S}}((\sharp a \geq p_a^{\rho}))$, entonces $p_a^{\rho} = p_a^{\beta}$ y entonces se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t + 1) \models C_{\mathcal{S}}((\sharp a = p_a^{\rho}))$. Esto debe cumplirse también para todos los puntos $(r^{\rho}, 2t' + 1)$ donde $t' > t$.□

El siguiente teorema nos indicará cómo se va transformando el conocimiento común sobre el número de sabios que hay con cada color; esto va dependiendo de las respuestas de los sabios. En un inicio hay conocimiento común de que hay al menos min_d sabios con cada color inevitable d, y no hay conocimiento común de que hay al menos un sabio con cada color evitable.

En el teorema suponemos que ya se han realizado $2t - 1$ rondas del acertijo. Ya algunos sabios han adivinado el color que les asignó el rey. Lo que vamos a ver es cómo afecta al conocimiento común las respuestas de los sabios.

Teorema 5 *Supongamos que se cumple* $(\mathcal{I}_{ei}, r^\rho, 2t-1) \models C_{\mathcal{S}}((\sharp a_i = p^\rho_{a_i}))$ *para todo color* a_i, $i = 1, \ldots, s$, *tal que los sabios con color* a_i *ya respondieron que sí saben su color y se cumple* $(\mathcal{I}_{ei}, r^\rho, 2t-1) \models C_{\mathcal{S}}((\sharp b_j \geq m_{b_j})) \wedge \neg C_{\mathcal{S}}((\sharp b_j \geq m_{b_j}+1))$ *para los demás colores* b_j, $j = 1, \ldots, r-s$, *donde* $m_{b_j} \geq 0$.

Supongamos también que cualquier punto $(r^\beta, 2t-1)$ *para el que se cumpla* $(\mathcal{I}_{ei}, r^\beta, 2t-1) \models (\sharp a_i = p^\rho_{a_i}) \wedge (\sharp b_j \geq m_{b_j})$, *para* $i = 1, \ldots, s$, $j = 1, \ldots, r-s$, *entonces* $(r^\beta, 2t-1)$ *es accesible desde* $(r^\rho, 2t-1)$.

Entonces en la siguiente ronda se cumple que:

para todo color a_i, $i = 1, \ldots, s$ *todos los sabios con color* a_i *responden que sí saben su color y se cumple*

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_{\mathcal{S}}((\sharp a_i = p^\rho_{a_i}))$$

para los demás colores se cumple que

(a) Si todos los sabios con color b_j, $j = 1, \ldots, r-s$, *responden que no saben en la ronda* $2t$ *entonces se cumple*

si $m_{b_j} > 0$ *entonces*

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_{\mathcal{S}}((\sharp b_j \geq m_{b_j}+1)) \wedge \neg C_{\mathcal{S}}((\sharp b_j \geq m_{b_j}+2))$$

si $m_{b_j} = 0$ *entonces*

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_{\mathcal{S}}((\sharp b_j \geq 0)) \wedge \neg C_{\mathcal{S}}((\sharp b_j \geq 1))$$

Y si en $(r^\beta, 2t+1)$ *se cumple* $(\mathcal{I}_{ei}, r^\beta, 2t+1) \models (\sharp a_i = p^\rho_{a_i}) \wedge (\sharp b_j \geq l_{b_j})$, *donde* $l_{b_j} = m_{b_j}+1$ *si* $m_{b_j} > 0$, $l_{b_j} = 0$ *si* $m_{b_j} = 0$, *entonces* $(r^\beta, 2t+1)$ *es accesible desde* $(r^\rho, 2t+1)$.

(b) Si para algunos colores d_i *los sabios con color* d_i *responden sí en la ronda* $2t$, *donde* $d_i = b_j$ *para alguna* j, $i = 1, \ldots, u$. *Entonces se cumple*

para todo color d_i, $i = 1, \ldots, u$

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_{\mathcal{S}}((\sharp d_i = p^\rho_{d_i}))$$

para los demás colores se cumple los sabios con ese color respondieron que no saben y

si $m_{b_j} > 0$ *entonces*

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S((\sharp b_j \geq max(m_{b_j}+1, \sharp c_{b_j} - k + \sum_{i=1}^{s} f^\rho_{a_i} + \sum_{i=1}^{u} f^\rho_{d_i}))$$

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models \neg C_S((\sharp b_j \geq max(m_{b_j}+1, \sharp c_{b_j} - k + \sum_{i=1}^{s} f^\rho_{a_i} + \sum_{i=1}^{u} f^\rho_{d_i}) + 1)$$

si $m_{b_j} = 0$ *entonces*

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S((\sharp b_j \geq max(0, \sharp c_{b_j} - k + \sum_{i=1}^{s} f^\rho_{a_i} + \sum_{i=1}^{u} f^\rho_{d_i})))$$

$$(\mathcal{I}_{ei}, r^\rho, 2t+1) \models \neg C_S((\sharp b_j \geq max(0, \sharp c_{b_j} - k + \sum_{i=1}^{s} f^\rho_{a_i} + \sum_{i=1}^{u} f^\rho_{d_i})) + 1)$$

Y si en $(r^\beta, 2t+1)$ *se cumple* $(\mathcal{I}_{ei}, r^\beta, 2t+1) \models (\sharp a_i = p^\rho_{a_i}) \wedge (\sharp d_i = p^\rho_{d_i}) \wedge (\sharp b_j \geq l_{b_j})$, *donde* $l_{b_j} = max(m_{b_j}+1, , \sharp c_{b_j} - k + f^\rho_{a_1} + \ldots + f^\rho_{a_s} + f^\rho_{d_1} + \ldots + f^\rho_{d_u})$ *si* $m_{b_j} > 0$, $l_{b_j} = max(0, \sharp c_{b_j} - k + f^\rho_{a_1} + \ldots + f^\rho_{a_s} + f^\rho_{d_1} + \ldots + f^\rho_{d_u})$ *si* $m_{b_j} = 0$, *entonces* $(r^\beta, 2t+1)$ *es accesible desde* $(r^\rho, 2t+1)$.

Prueba

Por el lema 15 todos los sabios con color a_i, para $i = 1, \ldots, s$, responden que sí saben su color en la ronda $2t$. Por el lema 36 para todo color a_i se cumple $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S((\sharp a_i = p^\rho_{a_i}))$.

Como se cumple $(\mathcal{I}_{ei}, r^\rho, 2t-1) \models C_S((\sharp b_j \geq m_{b_j})) \wedge \neg C_S((\sharp b_j \geq m_{b_j}+1))$ existen puntos $(r^{\beta_{b_j}}, 2t-1)$ accesibles desde $(r^\rho, 2t-1)$ en los que se cumple $(\mathcal{I}_{ei}, r^{\beta_{b_j}}, 2t-1) \models (\sharp b_j = m_{b_j}))$, es decir que $p^{\beta_{b_j}}_{b_j} = m_{b_j}$, por el lema 29 todos los sabios con color b_j saben su color en $(r^{\beta_{b_j}}, 2t-1)$, para toda $j = 1, \ldots, r-s$.

Supongamos que para alguna j, $m_{b_j} > 0$ y que todos los sabios con color b_j en $(r^\rho, 2t-1)$ no supieron su color. Entonces $(r^{\beta_{b_j}}, 2t+1)$ no es accesible desde $(r^\rho, 2t+1)$ porque la respuesta de los sabios con color b_j en estos puntos es distinta. Entonces se cumple $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S((\sharp b_j \geq m_{b_j}+1))$ para todo color b_j tal que $m_{b_j} > 0$.

Por otro lado para los colores b_j tales que $m_{b_j} = 0$ como se cumple $(\mathcal{I}_{ei}, r^\rho, 2t-1) \models C_S((\sharp b_j \geq 0))$ entonces esto también se cumple al tiempo $2t+1$, es decir que se cumple $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S((\sharp b_j \geq 0))$.

Los puntos $(r^{\beta_{b_j}}, 2t-1)$ son accesibles desde $(r^\rho, 2t-1)$ y entonces en cada uno de estos puntos se cumple $(\mathcal{I}_{ei}, r^{\beta_{b_j}}, 2t-1) \models (\sharp a_i = p^\rho_{a_i}) \wedge (\sharp b_l \geq m_{b_d})$ para $i = 1, \ldots, s$ y $d = 1, \ldots, r-s$. Es decir que $p^{\beta_{b_j}}_{a_i} = p^\rho_{a_i}$ y ya sabemos que $p^{\beta_{b_j}}_{b_j} = m_{b_j}$; para $d \neq j$ ocurre que $p^{\beta_{b_j}}_{b_d} \geq m_{b_d}$. Entonces fijémonos en una suma que resultará útil, $n = \sum_{i=1}^{s} p^{\beta_{b_j}}_{a_i} + \sum_{j=1}^{r-s} p^{\beta_{b_j}}_{b_j} = \sum_{i=1}^{s} p^\rho_{a_i} + \sum_{j=1}^{r-s} p^{\beta_{b_j}}_{b_j}$.

Veamos los casos que nos interesan:

(a) Si todos los sabios con color b_j responden no en la ronda $2t$ entonces:

En $(r^\rho, 2t-1)$ todos los sabios con color b_j respondieron que no saben su color, entonces debe ocurrir que $p^\rho_{b_j} \geq m_{b_j}$, si ocurre que $p^\rho_{b_j} = m_{b_j}$ es porque $m_{b_j} = 0$, ya que de otra forma los sabios con color b_j hubieran sabido su color, para $j = 1, \ldots, r-s$.

Fijémonos en lo que ocurre para alguno de estos colores b_j. Sea $b = b_i$ para alguna i, en ρ ocurre que hay $p^\rho_b = m_b + c$ sabios con color b, donde $c \geq 0$.

Si $c = 0$ entonces $p^\rho_b = m_b$ y entonces $m_b = 0$. Ocurre que $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models (\sharp b = 0)$ y entonces se cumple $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models \neg C_S((\sharp b \geq 1))$ y ya vimos que se cumple $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S((\sharp b_j \geq 0))$.

Supongamos entonces que $c > 0$. En ρ se cumple $\sum_{i=1}^{s} p^\rho_{a_i} + \sum_{j=1}^{r-s} p^\rho_{b_j} = n$. Es decir que $\sum_{j=1}^{r-s} p^\rho_{b_j} = \sum_{j=1}^{r-s} p^{\beta_{b_j}}_{b_j}$. Pero además $m_b + c + \sum_{b_j \neq b} p^\rho_{b_j} = p^\rho_b + \sum_{b_j \neq b} p^\rho_{b_j} = \sum_{j=1}^{r-s} p^\rho_{b_j} = \sum_{j=1}^{r-s} p^{\beta_b}_{b_j} = p^{\beta_b}_b + \sum_{b_j \neq d} p^{\beta_b}_{b_j} = m_b + \sum_{b_j \neq d} p^{\beta_b}_{b_j}$. Es decir que $\sum_{b_j \neq d} p^\rho_{b_j} = \sum_{b_j \neq d} p^{\beta_b}_{b_j} - c$. Entonces $c = \sum_{b_j \neq d} f^\rho_{b_j} - \sum_{b_j \neq d} f^{\beta_b}_{b_j}$. Es decir que $\sum_{b_j \neq d} f^\rho_{b_j} \geq c > 0$.

Lo que vamos a hacer es construir una cadena de puntos accesibles a partir de $(r^\rho, 2t-1)$. Iremos intercambiando el color de un sabio por alguno de los colores b_j distintos a b y sabemos que esto podemos hacerlo al menos c veces. Sea $\alpha_0 = \rho$ y tomemos c sabios con color b en ρ, podemos hacerlo porque $p^\rho_b = m_b + c$. Vamos intercambiando el color de estos sabios por algún color b_j tal que $f^\rho_{b_j} > 0$ y vamos formando las configuraciones α_k de manera que en α_k hay $m_b + c - k$

sabios con color b, hay $\sum_{b_j \neq d} p^{\rho}_{b_j} + k$ sabios con color b_j distinto a b y se cumple que $\sum_{b_j \neq d} f^{\rho}_{b_j} \geq c - k$. En α_k siempre hay más de m_b sabios con color b, para los demás colores lo que hicimos fue aumentar el número de sabios que había en ρ con ese color, entonces se cumple $(\mathcal{I}_{ei}, r^{\alpha_k}, 2t+1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp b_j \geq m_{b_j})))$, para $i = 1, \ldots, s$. $j = 1, \ldots, r-s$. Entonces por hipótesis de inducción $(r^{\alpha_k}, 2t-1)$ es accesible desde $(r^{\rho}, 2t-1)$. Para $k = 1, \ldots, c-1$ en cada punto $(r^{\alpha_k}, 2t-1)$ todos los sabios con color b_j no saben su color, los de color $b_j \neq b$ no lo saben porque se cumple $p^{\alpha_k}_{b_j} \geq p^{\rho}_{b_j} \geq m_{b_j}$ y entonces si alguno supiera su color se cumpliría que $(\mathcal{I}_{ei}, r^{\alpha_k}, 2t-1) \models C_{\mathcal{S}}((\sharp b_j \geq p^{\alpha_k}_{b_j}))$ pero entonces se cumpliría $(\mathcal{I}_{ei}, r^{\rho}, 2t-1) \models C_{\mathcal{S}}((\sharp b_j \geq p^{\alpha_k}_{b_j}))$, entonces debe ocurrir que $p^{\alpha_k}_{b_j} = p^{\rho}_{b_j} = m_{b_j}$, pero entonces los sabios con color b_j en $(r^{\rho}, 2t-1)$ debieron saber su color ó $m_{b_j} = 0$ pero entonces no habría sabios con este color. Entonces los puntos $(r^{\alpha_k}, 2t+1)$ son accesibles desde $(r^{\rho}, 2t+1)$ para $k = 1, \ldots, c-1$. Para α_c debemos fijarnos en el valor de m_b. Si $m_b > 0$ entonces en α_c habría m_b sabios con color b y entonces todos los sabios con este color saben su color, por lo tanto $(r^{\alpha_c}, 2t+1)$ no es accesible desde $(r^{\rho}, 2t+1)$. Sin embargo si $m_b = 0$ entonces en α_c no hay sabios con color b y entonces el punto $(r^{\alpha_c}, 2t+1)$ sí es accesible desde $(r^{\rho}, 2t+1)$ porque las respuestas de todos los sabios son las mismas.

Entonces si $m_b > 0$, $(r^{\alpha_{c-1}}, 2t+1)$ es un punto en el que se cumple $(\mathcal{I}_{ei}, r^{\alpha_k}, 2t+1) \models (\sharp d = m_{b_j} + 1)))$, por lo tanto se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models \neg C_{\mathcal{S}}((\sharp d \geq m_b + 2))$. Si $m_b = 0$, $(r^{\alpha_c}, 2t+1)$ es un punto en el que se cumple $(\mathcal{I}_{ei}, r^{\alpha_c}, 2t+1) \models (\sharp d = 0)))$, por lo tanto se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models \neg C_{\mathcal{S}}((\sharp d \geq 1))$.

Entonces se cumple que

si $m_{b_j} > 0$

$$(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp b_j \geq m_{b_j} + 1)) \wedge \neg C_{\mathcal{S}}((\sharp b_j \geq m_{b_j} + 2))$$

si $m_{b_j} = 0$

$$(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp b_j \geq 0)) \wedge \neg C_{\mathcal{S}}((\sharp b_j \geq 1))$$

Sea $(r^{\beta}, 2t+1)$ tal que se cumple $(\mathcal{I}_{ei}, r^{\beta}, 2t+1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp b_j \geq l_{b_j})$, donde $l_{b_j} = m_{b_j} + 1$ si $m_{b_j} > 0$, $l_{b_j} = 0$ si $m_{b_j} = 0$. Se cumple

también $(\mathcal{I}_{ei}, r^{\beta}, 2t-1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp b_j \geq l_{b_j})$. Entonces $(r^{\beta}, 2t-1)$ es accesible desde $(r^{\rho}, 2t-1)$. Ocurre que $\sum_{j=1}^{r-s} p^{\beta}_{b_j} = \sum_{j=1}^{r-s} p^{\rho}_{b_j}$. Debemos fijarnos en los colores $d = b_j$, p.a. j, tales que $p^{\beta}_d > p^{\rho}_d$ y en los colores $b = b_i$, p.a i, para los que ocurra lo contrario, esto es que $p^{\beta}_b < p^{\rho}_b$. Podemos ir de β a ρ por medio de configuraciones β_i, donde $\beta = \beta_0$ de manera que tomamos cada vez un sabio con un color d tal que $d = b_j$ y $p^{\beta_i}_d > p^{\rho}_d$ y le intercambiamos su color por algún color $b = b_i$ tal que $p^{\beta_i}_b < p^{\rho}_b$. Para $i = \sum p^{\beta}_b - \sum p^{\rho}_b$ se cumple que $\beta_i = \rho$. De esta manera ocurre que para toda i se cumple $(\mathcal{I}_{ei}, r^{\beta_i}, 2t-1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp b_j \geq l_{b_j})$ porque estamos restando el número de sabios con color b pero sólo hasta llegar a p^{ρ}_b y sabemos que $p^{\rho}_b \geq l_{b_j}$, para los demás colores o los dejamos en el mismo número o le sumamos sabios con ese color y ya sabemos que β cumple con que $p^{\rho}_{b_j} \geq l_{b_j}$. Entonces todos los β_i cumplen con que $(\mathcal{I}_{ei}, r^{\beta_i}, 2t+1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp b_j \geq l_{b_j})$, donde $l_{b_j} = m_{b_j} + 1$ si $m_{b_j} > 0$, $l_{b_j} = 0$ si $m_{b_j} = 0$ por lo que todos los sabios con color b_j no saben su color en la ronda $2t$ y entonces se cumple que $(r^{\beta}, 2t+1)$ es accesible desde $(r^{\rho}, 2t+1)$ porque todos los puntos $(r^{\beta_i}, 2t+1)$ lo son.

(b) Supongamos que para algunos colores d_i los sabios con color d_i responden sí en la ronda $2t$, donde $d_i = b_j$ para alguna j, $i = 1, \ldots, u$. Debe ocurrir que en ρ hay exactamente m_{d_i} sabios con color d_i, es decir que $p^{\rho}_{d_i} = m_{d_i}$. Para los demás colores b_j todos los sabios respondieron que no saben su color, renombremos a estos colores como $e_k = b_j$ para alguna j, $k = 1, \ldots, r-s-u$

Por el lema 36 se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp d_i = p^{\rho}_{d_i}))$.

Para los demás colores e_j, si $m_{e_j} > 0$ ya vimos que debe cumplirse $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp e_j \geq m_{e_j} + 1))$. Y si $e_{b_j} = 0$ debe cumplirse $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp e_j \geq 0))$.

Veamos que ocurre a partir de que hay un nuevo grupo de sabios que sabe su color. Sabemos que en ρ se cumple $\sum_{i=1}^{r} f^{\rho}_i = k$, entonces $\sum_{i=1}^{r} f^{\rho}_i = \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + \sum_{j=1}^{r-s-u} f^{\rho}_{e_j} = k$.

Entonces $\sum_{j=1}^{r-s-u} f^{\rho}_{e_j} = k - (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})$. Esto quiere decir que para cada $j = 1, \ldots, r-s-u$ se cumple $f^{\rho}_{e_j} \leq k - (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})$

Sabemos que $p^{\rho}_{e_j} = \sharp c_{e_j} - f^{\rho}_{e_j}$. Entonces se cumple que para cada e_j, $\sharp c_{e_j} - p^{\rho}_{e_j} \leq k - (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})$, entonces $p^{\rho}_{e_j} \geq \sharp c_{e_j} - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i}$.

Esto quiere decir que se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp e_j \geq \sharp c_{e_j} - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i}))$

Entonces si $m_{e_j} > 0$ se cumple

$$(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp e_j \geq max(m_{e_j}+1, \sharp c_{e_j} - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})))$$

y si $m_{e_j} = 0$ se cumple

$$(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models C_{\mathcal{S}}((\sharp e_j \geq max(0, \sharp c_{e_j} - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})))$$

Sea $e = e_i$ para alguna i, si $p^{\rho}_{e} = m_e$ entonces ocurre que $m_e = 0$. Entonces ocurre que $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models (\sharp e = 0)$ y entonces se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models \neg C_{\mathcal{S}}((\sharp e \geq 1))$.

Sea l_{e_j} tal que $l_{e_j} = m_{e_j} + 1$ si $m_{e_j} > 0$ y $l_{e_j} = m_{e_j} = 0$ si $m_{e_j} = 0$.

En ρ se cumple $f^{\rho}_{e} \leq k - (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})$ y también se cumple $f^{\rho}_{e} \leq \sharp c_e - l_e$. Veamos dos casos:

(i) Supongamos que $l_{e_j} \leq \sharp c_{e_j} - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i}$ entonces ocurre que $k - (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i}) \leq \sharp c_{e_j} - l_{e_j}$. Pensemos entonces en una configuración γ_e tal que se cumple $p^{\gamma_e}_{a_i} = p^{\rho}_{a_i}$ y $p^{\gamma_e}_{d_i} = p^{\rho}_{d_i}$, para toda i. Además pensemos que se cumple $f^{\gamma_e}_{e} = k - (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i})$ y $f^{\gamma_e}_{e_j} = 0$ para todo color $e_j \neq e$. Entonces se cumple que $p^{\gamma_e}_{e} = \sharp c_e - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i}$ y $p^{\gamma_e}_{e_j} = \sharp c_{e_j}$ para todo color $e_j \neq e$. Resulta entonces que γ_e cumple con que $(\mathcal{I}_{ei}, r^{\gamma_e}, 2t-1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp b_j \geq m_{b_j})$, para $i = 1, \ldots, s$, $j = 1, \ldots, r-s$ y entonces por hipótesis $(r^{\gamma_e}, 2t-1)$ es accesible desde $(r^{\rho}, 2t-1)$.

En ρ ocurre que $p^{\rho}_{e} = \sharp c_e - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + c$ donde $c \geq 0$. También se cumple que $\sum_{j=1}^{r-s-u} p^{\rho}_{e_j} = \sum_{j=1}^{r-s-u} p^{\gamma_e}_{e_j}$. Entonces $\sharp c_e - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + c + \sum_{e_j \neq e} p^{\rho}_{e_j} = p^{\rho}_{e} + \sum_{e_j \neq e} p^{\rho}_{e_j} = \sum_{j=1}^{r-s-u} p^{\rho}_{e_j} = \sum_{j=1}^{r-s-u} p^{\gamma_e}_{e_j} = p^{\gamma_e}_{e} + \sum_{e_j \neq e} p^{\beta_e}_{e_j} = \sharp c_e - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + \sum_{e_j \neq e} p^{\beta_e}_{e_j}$. Es decir que $\sum_{e_j \neq e} p^{\rho}_{e_j} = \sum_{e_j \neq e} p^{\gamma_e}_{e_j} - c$. entonces $\sum_{e_j \neq e} f^{\rho}_{e_j} = c + \sum_{e_j \neq e} f^{\gamma_e}_{e_j}$ y entonces $\sum_{e_j \neq e} f^{\rho}_{e_j} \geq c > 0$.

Vamos a construir una cadena de puntos accesibles a partir de $(r^\rho, 2t-1)$ de manera similar a como lo hicimos en la prueba del inciso (a). Iremos intercambiando el color de algún sabio por un color e_j distintos a e. esto podemos hacerlo al menos c veces. Sea $\alpha_0 = \rho$ y tomemos c sabios con color e en ρ, que sabemos existen porque $p^\rho_\epsilon \geq c$, vamos intercambiando el color de estos sabios por un color e_j y vamos formando las configuraciones α_j de manera que en α_j hay $\sharp c_e - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i} + c - j$ sabios con color e, hay $\sum_{e_j \neq e} p^\rho_{e_j} + j$ sabios con color e_j distinto a e y se cumple $\sum_{e_j \neq e} f^\rho_{e_j} \geq c - j$. En α_j siempre hay más de $\sharp c_e - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i}$ sabios con color e; para los demás colores sólo cambiamos la cantidad de sabios con algunos colores e_j distintos a e y lo que hicimos fue aumentar el número de sabios que había con estos colores en ρ. Ya sabíamos que se cumple $p^\rho_{e_j} \geq max(l_{e_j}, \sharp c_{e_j} - k + (\sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i}))$, entonces se cumple $(\mathcal{I}_{ei}, r^{\alpha_j}, 2t+1) \models (\sharp a_i = p^\rho_{a_i}) \wedge (\sharp d_i = p^\rho_{d_i}) \wedge (\sharp e_j \geq max(l_{e_j}, \sharp C_{e_j} - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i})$, por hipótesis $(r^{\alpha_j}, 2t-1)$ es accesible desde $(r^\rho, 2t-1)$. Para $j = 1, \ldots, c$ en cada punto $(r^{\alpha_j}, 2t-1)$ todos los sabios con color e_j no saben su color y entonces los puntos $(r^{\alpha_j}, 2t+1)$ son accesibles desde $(r^\rho, 2t+1)$ para $j = 1, \ldots, c$. Entonces el punto $(r^{\alpha_c}, 2t+1)$ es accesible desde $(r^\rho, 2t+1)$, como se cumple $(\mathcal{I}_{ei}, r^{\alpha_c}, 2t+1) \models (\sharp e = \sharp c_e - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i})$. Entonces se cumple $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models \neg C_S(\sharp e \geq max(l_j, \sharp c_e - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i}) + 1)$.

(ii) Nos falta considerar el caso en el que se cumple que $l_e > \sharp c_e - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i}$. entonces $k - (\sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i}) > \sharp c_e - l_e$. Esto lo que quiere decir es que $f^\rho_e \leq \sharp c_e - l_e < k - (\sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i})$.

Pensemos entonces en una configuración γ_e tal que se cumple $p^{\gamma_e}_{a_i} = p^\rho_{a_i}$ y $p^{\gamma_e}_{d_i} = p^\rho_{d_i}$, para toda i. Además pensemos que se cumple $p^{\gamma_e}_e = l_e$ y entonces se cumple $f^{\gamma_e}_e = \sharp c_e - l_e$, pero entonces nos faltaría asignar el valor de $f^{\gamma_e}_{e_j}$ para cada color $e_j \neq e$ donde $\sum_{e_j \neq e} f^{\gamma_e}_{e_j} = k - \sum_{i=1}^s f^\rho_{a_i} - \sum_{i=1}^u f^\rho_{d_i} - \sharp c_e + l_e$.

Si para algún color $e_b \neq e$ en γ_e se cumple $\sharp c_{e_b} - k + \sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i} \geq l_{e_b}$ entonces se cumple $f^{\gamma_e}_{e_b} \leq k - \sum_{i=1}^s f^\rho_{a_i} - \sum_{i=1}^u f^\rho_{d_i} - \sharp c_e + l_e < k - \sum_{i=1}^s f^\rho_{a_i} - \sum_{i=1}^u f^\rho_{d_i} \leq \sharp c_{e_b} - l_{e_b}$ y entonces podemos pensar que se cumple $f^{\gamma_e}_{e_b} = k - (\sum_{i=1}^s f^\rho_{a_i} + \sum_{i=1}^u f^\rho_{d_i}) - \sharp c_e + l_e$ y $f^{\gamma_e}_{e_j} = 0$ para todos los demas colores $e_j \neq e, e_l$.

El otro caso es que para todos los colores $e_j \neq e$ en γ_e se cumpla $\sharp c_{e_j} - k + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} < l_{e_j}$. En γ_e hasta ahora ya hemos asignado su color a $s_{\gamma_e} = \sum_{i=1}^{s}(\sharp c_{a_i} - f^{\rho}_{a_i}) + \sum_{i=1}^{u}(\sharp c_{d_i} - f^{\rho}_{d_i}) + \sharp c_e - f^{\gamma_e}_{e}$ sabios, nos hace falta asignar el color de $n - s_{\gamma_e}$ sabios y a éstos queremos asignarles un color $e_j \neq e$. Veamos que podemos hacerlo siempre.

Fijémonos que como se cumple $p^{\gamma_e}_{e} = l_e$ y se cumple $p^{\rho}_{e_j} \geq l_{e_j}$ entonces $\sharp c^{\gamma_e}_{e} - f^{\gamma_e}_{e} \leq \sharp c^{\rho}_{e} - f^{\rho}_{e}$ y entonces se cumple que $f^{\gamma_e}_{e} \geq f^{\rho}_{e}$.

Primero veamos que podemos asignar los colores que queremos a los $n - s_{\gamma_e}$ sabios que nos faltan. Es decir que se cumple que $\sum_{e_j \neq e} \sharp c_{e_j} \geq n - s_{\gamma_e}$. Supongamos que ocurre $\sum_{e_j \neq e} \sharp c_{e_j} < n - s_{\gamma_e}$. Entonces $\sum_{e_j \neq e} \sharp c_{e_j} < n - \sum_{i=1}^{s}(\sharp c_{a_i} - f^{\rho}_{a_i}) - \sum_{i=1}^{u}(\sharp c_{d_i} - f^{\rho}_{d_i}) - \sharp c_e + f^{\gamma_e}_{e}$. Entonces se cumple $\sum_{a=1}^{r} \sharp c_a = \sum_{e_j \neq e} \sharp c_{e_j} + \sum_{i=1}^{s} \sharp c_{a_i} + \sum_{i=1}^{u} \sharp c_{d_i} + \sharp c_e < n + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + f^{\gamma_e}_{e} \leq n + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + f^{\rho}_{e} \leq n + \sum_{a=1}^{r} f^{\rho}_{a}$. Como $\sum_{a=1}^{r} \sharp c_a = n + k$ y $\sum_{a=1}^{r} f^{\rho}_{a} = k$ entonces $n + k < n + k$, una contradicción. Entonces $\sum_{e_j \neq e} \sharp c_{e_j} \geq n - s_{\gamma_e}$. Lo que quiere decir que en principio sí podemos hacer la asignación que deseamos en los sabios que nos quedan.

Sin embargo sabemos que se cumple $p^{\gamma_e}_{e_j} \geq l_{e_j}$ para toda j. Tenemos que ver que se cumple que $\sum_{e_j \neq e} l_{e_j} \leq n - s_{\gamma_e}$, es decir que como estamos obligados a asignar cada color $e_j \neq e$ al menos a l_{e_j} sabios, tenemos que ver que al hacer esto no nos sobran sabios. Supongamos que $\sum_{e_j \neq e} l_{e_j} > n - s_{\gamma_e}$. Sabemos que $l_{e_j} \leq p^{\rho}_{e_j}$, es decir que $l_{e_j} \leq \sharp c_{e_j} - f^{\rho}_{e_j}$ y entonces $\sum_{e_j \neq e} l_{e_j} \leq \sum_{e_j \neq e}(\sharp c_{e_j} - f^{\rho}_{e_j})$. Entonces $\sum_{e_j \neq e}(\sharp c_{e_j} - f^{\rho}_{e_j}) > n - s_{\gamma_e}$ y entonces $\sum_{e_j \neq e}(\sharp c_{e_j} - f^{\rho}_{e_j}) > n - \sum_{i=1}^{s}(\sharp c_{a_i} - f^{\rho}_{a_i}) - \sum_{i=1}^{u}(\sharp c_{d_i} - f^{\rho}_{d_i}) - \sharp c_e + f^{\gamma_e}_{e}$. Por lo tanto se cumple que $\sum_{a=1}^{r} \sharp c_a = \sum_{e_j \neq e} \sharp c_{e_j} + \sum_{i=1}^{s} \sharp c_{a_i} + \sum_{i=1}^{u} \sharp c_{d_i} + \sharp c_e < n + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + \sum_{e_j \neq e} f^{\rho}_{e_j} + f^{\gamma_e}_{e} \leq n + \sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i} + \sum_{e_j \neq e} f^{\rho}_{e_j} + f^{\rho}_{e} = n + \sum_{a=1}^{r} f^{\rho}_{a} = n + k$. Entonces $n + k < n + k$, una contradicción. Entonces $\sum_{e_j \neq e} l_{e_j} \leq n - S_{\gamma_e}$. Lo que quiere decir que no nos sobran sabios.

Todo esto quiere decir que sí hay forma de asignar los colores de los sabios en γ_e de manera que se cumpla $(\mathcal{I}_{ei}, r^{\gamma_e}, 2t+1) \models (\sharp a_i = p^{\rho}_{a_i}) \wedge (\sharp d_i = p^{\rho}_{d_i}) \wedge (\sharp e_j \geq max(l_{e_j}, \sharp c_{e_j} - k + (\sum_{i=1}^{s} f^{\rho}_{a_i} + \sum_{i=1}^{u} f^{\rho}_{d_i}))$. En ρ hay $l_e + c$ sabios con color e. Procediendo como en los casos anteriores podemos ver que $(r^{\gamma_e}, 2t+1)$ es un punto accesible desde $(r^{\rho}, 2t+1)$ y entonces en ρ se cumple $(\mathcal{I}_{ei}, r^{\rho}, 2t+1) \models \neg C_S(\sharp e \geq$

$max(l_j, \sharp c_e - k + \sum_{i=1}^{s} f_{a_i}^{\rho} + \sum_{i=1}^{u} f_{d_i}^{\rho}) + 1)$.

De manera similar al inciso (a) podemos ver que cualquier punto que cumpla con que $(\mathcal{I}_{ei}, r^{\beta}, 2t+1) \models (\sharp a_i = p_{a_i}^{\rho}) \wedge (\sharp d_i = p_{d_i}^{\rho}) \wedge (\sharp b_j \geq l_{b_j})$, donde $l_{b_j} = max(m_{b_j} + 1, \sharp c_{b_j} - k + f_{a_1}^{\rho} + \ldots + \sum_{i=1}^{s} f_{a_i}^{\rho} + \sum_{i=1}^{u} f_{d_i}^{\rho}))$ si $m_{b_j} > 0$, y donde $l_{b_j} = max(0, \sharp c_{b_j} - k + f_{a_s}^{\rho} + f_{d_1}^{\rho} + \ldots + f_{d_u}^{\rho}))$ si $m_{b_j} = 0$, entonces $(r^{\beta}, 2t+1)$ es accesible desde $(r^{\rho}, 2t+1)$.□

Este teorema es el resultado básico para entender lo que ocurre en el acertijo de las esposas infieles. Nos indica cómo cambia el conocimiento común en una ronda. De manera inductiva, ronda por ronda, podemos deducir cómo se va desarrollando cualquier acertijo.

4.5.5 Un protocolo sin conocimiento

Veamos lo que ocurre en el acertijo de las esposas infieles. En un inicio sabemos que se cumple $(\mathcal{I}_{ei}, r^{\rho}, 1) \models C_{\mathcal{S}}(\sharp d \geq min_d) \wedge \neg C_{\mathcal{S}}(\sharp d \geq min_d + 1) \wedge C_{\mathcal{S}}(\sharp b \geq 0) \wedge \neg C_{\mathcal{S}}(\sharp b \geq 1)$; para todo color inevitable d y todo color evitable b. A partir de aquí el teorema 5 nos indica cómo va variando el conocimiento común de cuántos sabios hay con cada color. Por el teorema 4 un sabio sabe que su color es a si y sólo si hay conocimiento común de que hay al menos p_a^{ρ} sabios con color a, esto va a ocurrir, por el teorema 5, si y sólo si en algún momento hay conocimiento común de que hay p sabios con color a, donde $p > 0$.

Analicemos con más cuidado lo que pasa en un acertijo. El rey llega con una caja $C = (\sharp c_1, \sharp c_2, \ldots, \sharp c_r)$ y escoge la r-configuración $[\rho]$. Asigna el color a a $p_a^{[\rho]}$ sabios, para $a = 1, \ldots, r$. Supongamos que hay h colores inevitables, es decir que $\sharp c_h > k$ y $\sharp c_{h+1} \leq k$. Sabemos que el rey está obligado a asignar el color d a min_d sabios, para $d = 1, \ldots, h$. La elección del rey consiste en asignar hasta a k sabios el color d, para $d = 1, \ldots, h$ y asignar hasta a $\sharp c_a$ sabios el color a, para $a = h+1, \ldots, r$. Para los colores inevitables supongamos que en $[\rho]$ el rey asignó a $p_d^{[\rho]} = min_d + q_d^{[\rho]}$ sabios el color d, para $d = 1, \ldots, r$, donde $q_p^{[\rho]} \leq k$.

Reordenemos a los colores inevitables con los índices i_j, $j = 1, \ldots, h$, de manera que se cumpla $q_{i_1}^{[\rho]} \leq q_{i_2}^{[\rho]} \leq \ldots q_{i_h}^{[\rho]}$. Veamos un primer resultado.

Lema 37 *Para* $t = 0, \ldots, q_{i_1}^{[\rho]}$ *se cumple que* $(\mathcal{I}_{ei}, r^\rho, 2t+1) \models C_S(\sharp i_j \geq min_d + t) \wedge \neg C_S(\sharp i_j \geq min_{i_j} + t + 1) \wedge C_S(\sharp b \geq 0) \wedge \neg C_S(\sharp b \geq 1)$ *para todo color inevitable* i_j *y todo color evitable* b.

Prueba

Hagámoslo por inducción sobre t.

Base

Para $t = 0$. Por el lema 27 se cumple $(\mathcal{I}_{ei}, r^\rho, 1) \models C_S(\sharp i_j \geq min_{i_j})$ para todo color inevitable i_j. Hay r-configuraciones $[\alpha]$ en las que se cumple que $p_{i_j}^{[\alpha]} = min_{i_j}$ entonces se cumple $(\mathcal{I}_{ei}, r^\rho, 1) \models \neg C_S(\sharp d \geq min_d + 1)$. Por el lema 28 se cumple $(\mathcal{I}, r^\alpha, 1) \models \neg C_S((\sharp b \geq 1))$ para todo color b evitable. Hay r-configuraciones en las que se cumple $p_b^{[\alpha]} = 0$ entonces se cumple $(\mathcal{I}, r^\alpha, 0) \models C_S(\sharp b \geq 0)$.

Hipótesis de inducción

Supongamos que se cumple para $t-1$, es decir que se cumple que $(\mathcal{I}_{ei}, r^\rho, 2t-1) \models C_S(\sharp i_j \geq min_{i_j}+t-1) \wedge \neg C_S(\sharp i_j \geq min_{i_j}+t) \wedge C_S(\sharp b \geq 0) \wedge \neg C_S(\sharp b \geq 1)$.

Paso Inductivo

Para $t \leq q_{i_1}^{[\rho]}$. Por hipótesis de inducción se cumple $(\mathcal{I}_{ei}, r^\rho, 2t - 1) \models C_S(\sharp i_j \geq min_{i_j} + t - 1) \wedge \neg C_S(\sharp i_j \geq min_{i_j} + t) \wedge C_S(\sharp b \geq 0) \wedge \neg C_S(\sharp b \geq 1)$. Es decir que para cada color inevitable i_j se cumple $(\mathcal{I}_{ei}, r^\rho, 2t - 1) \models \neg C_S(\sharp i_j \geq min_{i_j} + t)$. Como $t \leq q_{i_1}^{[\rho]}$, entonces $t \leq q_{i_j}^{[\rho]}$, entonces se cumple $(\mathcal{I}_{ei}, r^\rho, 2t - 1) \models \neg C_S(\sharp i_j \geq min_{i_j} + q_{i_j}^\rho)$ y entonces $(\mathcal{I}_{ei}, r^\rho, 2t - 1) \models \neg C_S(\sharp i_j \geq p_{i_j}^{[\rho]})$. Para los colores evitables por hipótesis de inducción se cumple $(\mathcal{I}_{ei}, r^\rho, 2t - 1) \models \neg C_S(\sharp b \geq 1)$. Entonces por el teorema 4 los sabios con color a, para todo color $a = 1, \ldots, r$, no saben su color en la ronda $2t$. Entonces por el teorema 5 se cumple $(\mathcal{I}_{ei}, r^\rho, 2t + 1) \models C_S(\sharp i_j \geq min_{i_j} + t) \wedge \neg C_S(\sharp i_j \geq min_{i_j} + t + 1) \wedge C_S(\sharp b \geq 0) \wedge \neg C_S(\sharp b \geq 1)$. □

Este lema nos indica que en la primeras $q_{i_1}^{[\rho]}$ preguntas, todos los sabios responden que no saben su color. Sin embargo a la pregunta $q_{i_1}^{[\rho]} + 1$ todos los sabios que tengan un sombrero de color i_j para el que se cumpla $p_{i_j}^{[\rho]} = p_{i_1}^{[\rho]}$, saben su color.

Lo que va a ocurrir es que todos los sabios que tengan un color inevitable van a saber su color. Esto porque para estos colores desde un principio hay conocimiento común de que al menos hay min_d sabios

con color d, $d = 1, \ldots, h$. Por el teorema 5, en cada ronda, este valor aumenta en al menos uno. Entonces en algún momento va a ocurrir que haya conocimiento común de que hay al menos $p_d^{[\rho]}$ sabios con color d y entonces todos los sabios con color d saben su color. El ordenamiento que dimos nos indica el orden en el que van a ir respondiendo los sabios, los de color i_b nunca van a responder antes que los de color i_a si $a < b$.

Supongamos que los sabios con color i_b acaban de responder que saben su color y que los sabios con color i_j tal que $j > b$ no saben el color de su sombrero. Debe ocurrir que ya han respondido que saben su color todos los sabios con colores inevitables i_a tal que $a < b$. Por el teorema 4 en ese momento hay conocimiento común de cuántos sabios hay con cada color inevitable i_j, $j \leq b$. Para cada color evitable e va a ocurrir que hay conocimiento común de que el número de sabios con color e es al menos $max(0, \sharp c_e - k + \sum_{j=1}^{b} f_{i_j}^{[\rho]})$. Los sabios con algún color evitable van a poder saber el color de su sombrero si y sólo si $\sharp c_e - k + \sum_{j=1}^{b} f_{i_j}^{[\rho]} > 0$, ya que entonces el conocimiento común va a poder ir aumentando. Fijémonos que si para algún color e se cumple $\sharp c_e - k + \sum_{j=1}^{b} f_{i_j}^{[\rho]} > 0$ entonces se cumple también $\sharp c_f - k + \sum_{j=1}^{b} f_{i_j}^{[\rho]} > 0$ para toda $f < e$, es decir para $f = h+1, \ldots, e-1$. Si algún sabio con un color evitable llega a saber su color es seguro que todos los sabios con color $h+1$ van a saberlo.

Esto le puede dar alguna pista al rey para lograr que no haya muchos sabios que sepan su color. Ya vimos que todos los sabios que tengan color inevitable van a saber su color. Es decir que seguro que $\sum_{d=1}^{h} p_d^{[\rho]}$ sabios sabrán su color. Si el rey quiere evitar que haya más sabios que sepan su color, debe asegurarse que se cumpla que $\sharp c_{h+1} - k + \sum_{j=1}^{h} f_{i_j}^{[\rho]} \leq 0$.

Proponemos un protocolo en el que los sabios lo que tienen que ir haciendo es mantener un contador que indicará la evolución del conocimiento común de cuántos sabios hay con cada color. Veremos que las acciones que determina este protocolo son las mismas acciones que determina el protocolo $Pg_A^{\mathcal{I}_{ei}}$. Es decir veremos que este protocolo implementa al programa $\boldsymbol{Pg}_A$ en el contexto (γ_{ei}, π_A).

Para definir este protocolo debemos definir un protocolo para cada sabio y para el rey. El protocolo del rey será el mismo que hemos estado utilizando. Es decir que el rey seguirá ejecutando el protocolo $Pg_{rey}^{\mathcal{I}_{ei}}$.

Para los sabios definimos un nuevo protocolo.

En el nuevo protocolo para el sabio i debemos utilizar el mismo conjunto de estados locales que hemos estado utilizando. Queremos que el sabio i vea sólo a su estado local para determinar si sabe o no sabe su color. Lo primero que hará el sabio i es revisar si el rey le ha preguntado. Es decir revisará si $recibir_i((rey), rey)$ está en el conjunto de su historia de mensajes correspondiente a la última ronda. Luego tiene que ir revisando su historia de mensajes, ronda por ronda, para ir calculando la evolución del conocimiento común. Esto lo tiene que ir haciendo cada vez porque no queremos modificar el estado local del sabio; no podemos pensar en agregar variables a su estado local en las que guarde el estado del conocimiento común para cada color.

El sabio i debe determinar el número de sabios que ve con cada color. Para esto revisa la variable $vision_i(\rho)$ en su estado local. También se va a utilizar un arreglo de enteros $CComun[a]$, donde $a = 1, \ldots, r$ para determinar el valor del estado del conocimiento común para cada color. Lo que hace el sabio i para determinar el valor de $CComun[a]$, es revisar su historia de mensajes. Inicializa $CComun[a] = min_a$ si a es un color inevitable y $CComun[a] = 0$ si no lo es. A partir de aquí va determinando el valor de $CComun[a]$ según lo que se indica en el teorema 5. Es decir que si $CComun[a] > 0$ y en una ronda no hubo nuevos sabios que supieran su color se actualiza $CComun[a] = CComun[a] + 1$. Si $CComun[a] = 0$ se deja este valor. Si hay nuevos sabios que saben su color, ya se sabemos que se vuelve conocimiento común cuántos sabios hay con esos colores. Supongamos que para los colores b_i, $i = 1, \ldots, s$ los sabios con esos colores ya supieron su color. Si $CComun[a] > 0$ entonces $CComun[a] = max(CComun[a] + 1, \sharp c_a - k + \sum_{i=1}^{s} f_{b_i}^{\alpha})$. Si $CComun[a] = 0$ entonces $CComun[a] = max(0, \sharp c_a - k + \sum_{i=1}^{s} f_{b_i}^{\alpha})$

Definición 54 *El sabio i ejecutará el protocolo* P_i

Supongamos que $\ell_i = (t, C, vision_i(\alpha), hist_i(t))$, *donde* $hist_{\ell_i}(t) = (M^1_{\ell_i}, \ldots, M^t_{\ell_i})$

$$P_i(\ell_i) = \begin{cases} enviar_i(\langle 1, i\rangle, R_i) & si \\ & recibir_i(\langle rey\rangle, rey) \in (M^t_{\ell_i}) \\ & y\ CComun[1] > |(s_j, 1) \in vision_i(\alpha)| \\ \ldots & \\ enviar_i(\langle r, i\rangle, R_i) & si \\ & recibir_i(\langle rey\rangle, rey) \in (M^t_{\ell_i}) \\ & y\ CComun[r] > |(s_j, r) \in vision_i(\alpha)| \\ enviar_i(\langle no, i\rangle, R_i) & si \\ & recibir_i(\langle rey\rangle, rey) \in (M^t_{\ell_i}) \\ & y\ CComun[a] \leq |(s_j, a) \in vision_i(\alpha)| \\ & para\ todo\ a = 1, \ldots, r \\ \Lambda & en\ otro\ caso \end{cases}$$

Definimos al protocolo conjunto $P = (Pg^{\mathcal{I}}_{rey}, P_1, \ldots, P_n)$. Veamos que el protocolo P implementa al programa $\boldsymbol{Pg}_A$ en el contexto (γ_{ei}, π_A).

Lema 38 *El protocolo* P *implementa al programa* $\boldsymbol{Pg}_A$ *en el contexto* (γ_{ei}, π_A)

Prueba

Sea el sistema $\mathcal{I} = I^{rep}(P, \gamma, \pi_A)$. Entonces P implementa al programa $\boldsymbol{Pg}_A$ en el contexto (γ_{ei}, π_A) si (1) $\mathcal{I} = I^{rep}(Pg^{\mathcal{I}_{ei}}_A, \gamma_{ei}, \pi_A)$ y (2) P y $Pg^{\mathcal{I}_{ei}}_A$ concuerdan en todos los estados globales que aparecen en $\mathcal{I}$.

$\mathcal{I} = I^{rep}(Pg^{\mathcal{I}_{ei}}_A, \gamma_{ei}, \pi_A)$ si y sólo si $(R^{rep}(P, \gamma_{ei}), \pi_A) = (R^{rep}(Pg^{\mathcal{I}_{ei}}_A, \gamma_{ei}), \pi_A)$ si y sólo si $R^{rep}(P, \gamma_{ei}) = R^{rep}(Pg^{\mathcal{I}_{ei}}_A, \gamma_{ei})$.

Como el contexto es el mismo para ambos sistemas, entonces el conjunto de estados globales iniciales es el mismo. Si demostramos que se cumple el punto (2) demostraremos que $R^{rep}(P, \gamma_{ei}) = R^{rep}(Pg^{\mathcal{I}_{ei}}_A, \gamma_{ei})$ porque las ejecuciones se generan a partir de los estados globales iniciales, entonces si se cumple que P y $Pg^{\mathcal{I}_{ei}}_A$ concuerdan en todos los estados globales que aparecen en I, podemos ir demostrando que el conjunto de todos los t-prefijos en ambos sistemas es el mismo.

En todos los estados globales iniciales se cumple que el rey no ha mandando ningún mensaje a los sabios y entonces los sabios no toman

ninguna acción. Como el protocolo del rey es el mismo en ambos protocolos entonces P y $Pg_A^{\mathcal{I}_{ei}}$ concuerdan en todos los estados globales iniciales. Para cualquier otro estado va a ocurrir que si el tiempo es par el único que realiza alguna acción es el rey y entonces ambos protocolos concuerdan. Para los estados globales con tiempo impar la razón por la que ambos protocolos concuerdan es porque el protocolo P_i de cada sabio sigue el teorema 5

En $Pg_A^{\mathcal{I}_{ei}}$ en cada ronda par el sabio i manda el mensaje $\langle a, i\rangle$, es decir sabe que su color es a si y sólo si hay conocimiento común de que hay al menos p_a^e sabios con color a si y solo si $CComun[a] > |(s_j, a) \in vision_i(\rho)|$ si y sólo si en P el sabio i manda el mensaje $\langle a, i\rangle$.

Es decir que P y $Pg_A^{\mathcal{I}_{ei}}$ concuerdan en todos los estados globales que aparecen en I. □

4.6 Los tres sabios

En la versión de los tres sabios, el rey pregunta uno por uno a los sabios hasta llegar al sabio n, entonces el acertijo se detiene.

Una manera de pensar en este acertijo es que en cada ronda el rey escoga un color al cual "preguntarle". Veremos que podemos hacer un análisis del acertijo utilizando la gráfica de r-configuraciones. La idea básica será que todos los sabios, a los que el rey no ha preguntado, tienen la misma información. Si pensáramos que el rey va decidiendo en cada ronda a qué sabio le pregunta, lo que nos va a interesar es el color que tiene asignado el sabio; podemos olvidarnos de las etiquetas de los sabios y fijarnos solamente en las respuestas que se generan según los colores. Definimos el conjunto G_t como $G_0 = \mathcal{S}$ y para $t > 0$, $G_t = \mathcal{S} \setminus \{1, \ldots, t\}$. G_t representa el conjunto de sabios a los que el rey no ha preguntado después de la ronda $2t$.

Si en un principio el rey le pregunta a un sabio con un color no inevitable b, va a cambiar el conocimiento común de cuántos sabios hay con ese color. El sabio no va a saber su color, pero después de que el sabio responde se vuelve conocimiento común, entre los sabios en G_1, que un sabio con color b respondió. Es decir que se vuelve conocimiento común entre estos sabios que hay al menos un sabio con color b.

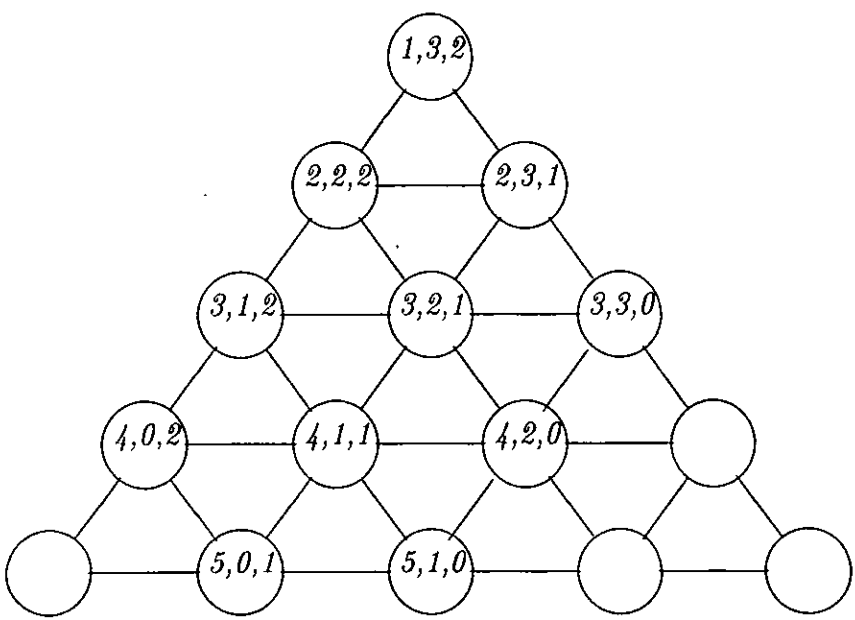
Si el rey le pregunta a un sabio con un color inevitable d y el sabio

no sabe su color, se vuelve conocimiento común, entre los sabios en G_1, que hay al menos $min_d + 1$ sabios con color d. Pero para los demás colores, el conocimiento común de al menos cuántos sabios hay con ese color no cambia.

El conocimiento común, para todos los colores, va a cambiar cuando un sabio responda que sabe su color; supongamos que que acaba de pasar la ronda $2t$, le preguntaron al sabio t y respondió que ya sabe que tiene color d. Entonces se vuelve conocimiento común, entre los sabios en G_t, el número de sabios con color d. Para los demas colores va a pasar que el conocimiento común se actualiza de una manera similar a lo que ocurre en el teorema 5. Para cada color b va a cambiar a $max(m_b, \sharp c_b - k + \sum_{i=1}^{s} f_{a_i}^{\rho})$ donde para los colores a_i ya ha habido un sabio con ese color que responde que sí sabe su color. En la ronda anterior había conocimiento común de que al menos hay m_b sabios con color b.

Presentamos algunos ejemplos de lo que ocurre en esta versión del acertijo.

Ejemplo 5 *Tenemos la caja* $C = (5, 3, 2)$ *y hay* $n = 6$ *sabios;* $k = 4$. *Sólo el color* 1 *es inevitable.*

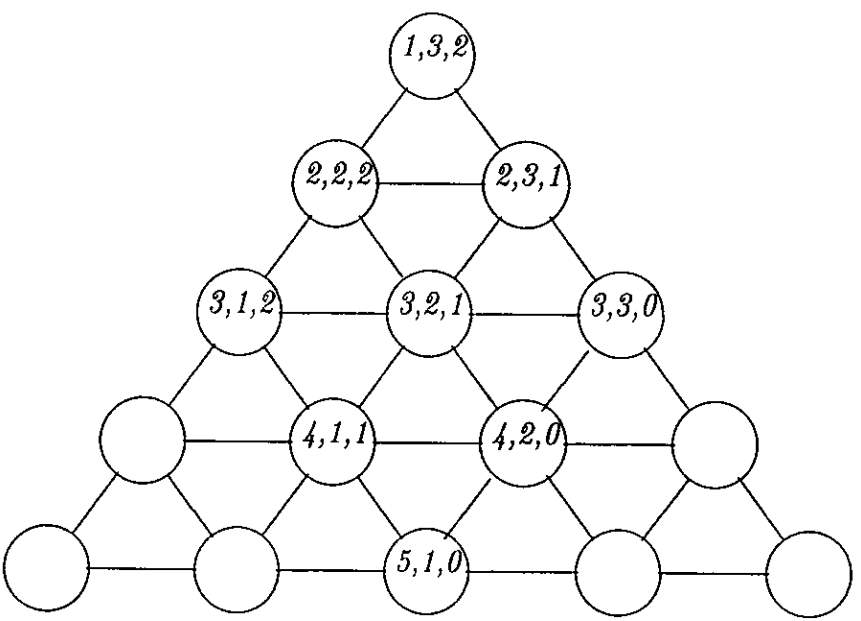


En la primera pregunta el rey puede escoger a un sabio con color 1, 2 *ó* 3. *Veamos los casos:*

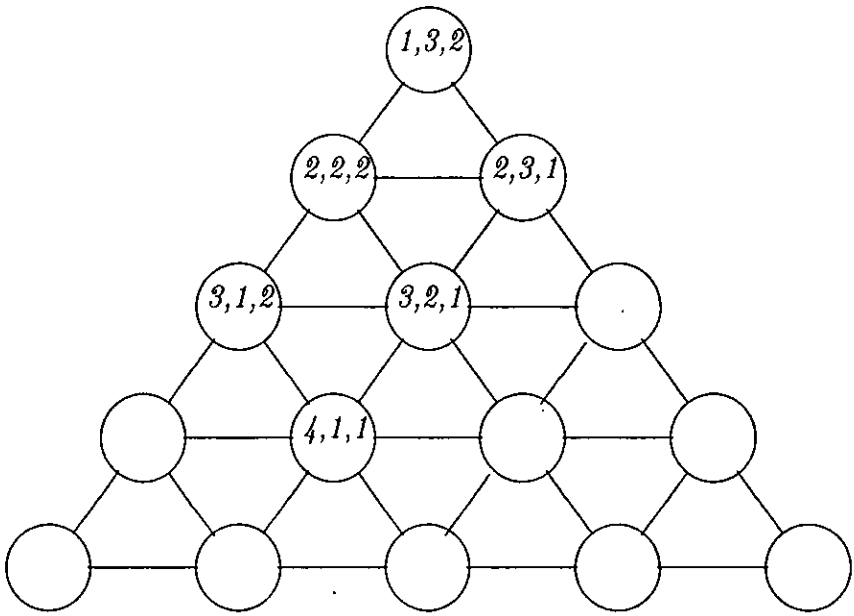
Si le pregunta a un sabio con color 1 *y dice sí entonces* $[\rho] = (1, 3, 2)$ *y entonces se vuelve conocimiento común que* $p_1^\rho = min_1 = 1$; *entonces* $f_1^\rho = 4$. *Para el color* 2 *y* 3 *había conocimiento común de que hay al menos* 0 *sabios con estos colores. Para el color* 2 *el conocimiento común cambia a* $max(0, 3 - 4 + 4) = 3$. *Para el color* 3 *cambia a* $max(0, 2 - 4 + 4) = 2$. *Es decir que se vuelve conocimiento común que al menos hay* 3 *sabios con color* 2 *y* 2 *sabios con color* 3. *Ocurre entonces que en las siguientes preguntas todos los sabios saben su color.*

En este caso la gráfica de r-configuraciones se reduce a un sólo punto.

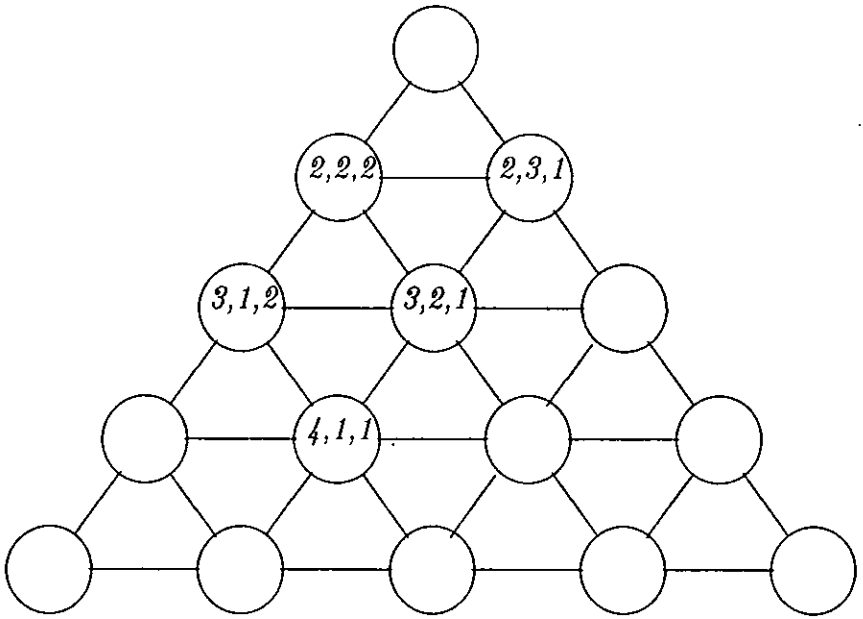
Si le pregunta primero a un sabio con color 2, *este va a decir que no sabe el color de su sombrero, y se vuelve conocimiento común que hay al menos un sabio con color* 2. *La gráfica cambia a*



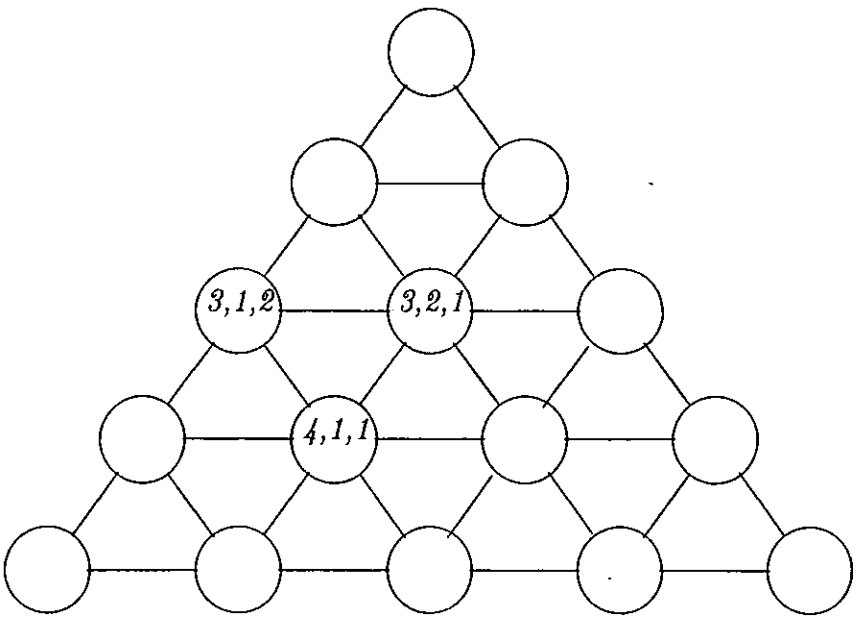
Si en la segunda pregunta el rey escoge a un sabio con color 3, *éste tampoco sabe su color. La gráfica cambia a*



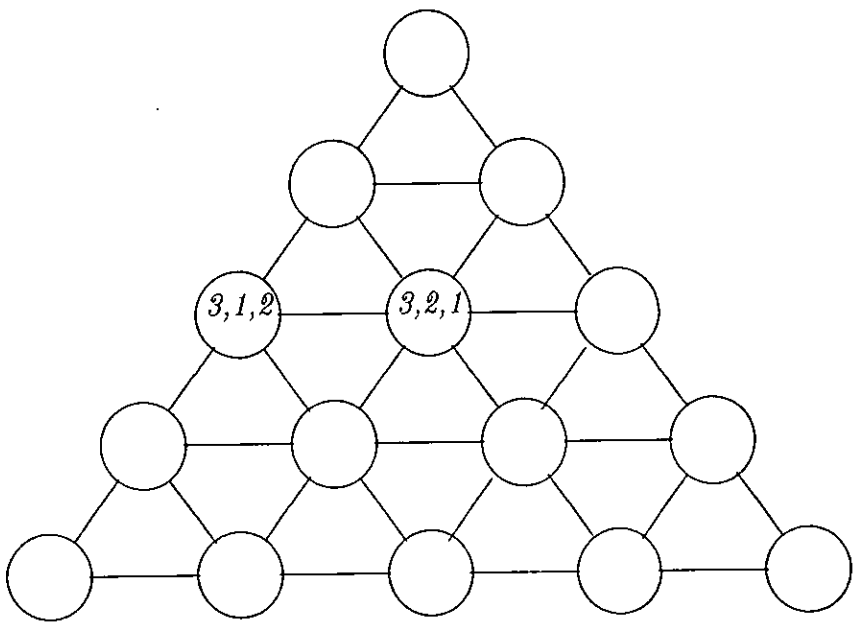
En la tercera pregunta el rey escoge a un sabio con color 1*; supongamos que el sabio responde que no sabe su color. Se vuelve conocimiento común que hay al menos* $min_1 + 1 = 2$ *sabios con color* 1*. Ahora la gráfica cambia a*



En la cuarta pregunta el rey escoge a un sabio con color 1 *y éste responde que no sabe su color. Se vuelve conocimiento común que hay al menos* $min_1 + 2 = 3$ *sabios con color* 1.



En la quinta pregunta el rey escoge a un sabio con color 1 *y éste responde que sí sabe su color. Se vuelve conocimiento común que hay exactamente* 3 *sabios con color* 1 *y se vuelve conocimiento común que* $f_a^\alpha = 2$. *Para el color* 2 *el conocimiento común se actualiza a* $max(1, 3-4+2) = 1$. *Para el color* 3 *el conocimiento común cambia a* $max(1, 2-4+2) = 1$.



En la sexta pregunta el rey puede escoger un sabio con color 2 *o uno con color* 3. *En ambos casos el sabio al que le pregunte va a responder que no sabe su color.*

Capítulo 5

Conclusiones

En este trabajo presentamos un modelo de conocimiento en sistemas distribuidos, basado en la semántica de los mundos posibles. El modelo captura la interacción que existe entre el conocimiento y las acciones de los procesadores del sistema. Utilizamos este modelo para formalizar una versión general de algunos acertijos que han aparecido frecuentemente en la literatura: el acertijo de los tres sabios, el acertijo de las esposas infieles y el acertijo de los niños enlodados. A esta versión general la llamamos el acertijo de los sabios.

Suponemos que el acertijo se desarrolla en un sistema distribuido. Capturamos el comportamiento de los procesadores del sistema, es decir, del rey y de los sabios, con un programa basado en conocimiento. Este programa nos da una especificación del comportamiento de los procesadores a nivel de conocimiento.

Asumimos que los sabios y el rey se comunican por medio de mensajes. El rey le manda un mensaje a un sabio representando la pregunta de si sabe su color. Los sabios responden mandando un mensaje con su color, si es que lo saben y tienen otro mensaje con el que responden que no saben. Con este sistema de mensajes formalizamos a la comunicación entre los sabios y el rey.

Para modelar lo que ocurre fuera del comportamiento de los sabios, se define un contexto. Incluimos en el contexto al protocolo del medio ambiente, que es lo que determina si un mensaje llega a su destino o si un procesador falla. En las versiones del acertijo que estudiamos suponemos que no hay fallos y entonces este protocolo no toma ninguna

acción. En el contexto también guardamos el conjunto de los estados globales iniciales y una función de transición que nos indica los efectos de las acciones de los procesadores sobre los estados globales.

Dado el programa basado en conocimiento y un contexto, encontramos al conjunto de todas las ejecuciones consistentes con el programa en el contexto. Para cada procesador definimos un protocolo asociado a este conjunto de ejecuciones, cada protocolo es una función de estados locales a conjuntos de acciones. Estos protocolos determinan las mismas acciones que el programa basado en conocimiento. Dado que estamos manejando nociones de conocimiento, en estos protocolos las acciones de cada sabio están en función no sólo de su estado local, sino que también debe realizar pruebas de conocimiento que dependen de todas las ejecuciones. Esto hace que sea poco claro pensar como podríamos implementar a estros protocolos directamente, ya que se pierde la intuición de que la única información que tiene un procesador es la que está en su estado local.

Sin embargo, los protocolos de los sabios son una función que va de los estados locales a conjuntos de acciones, aunque en un principio no tenemos bien claro cómo son estas funciones especificamente. Uno de los objetivos de la tesis fue encontrar una especificación que sólo dependa de los estados locales, es decir, encontrar un protocolo que implemente al programa basado en conocimiento, en el que los sabios sólo tengan que ver a su estado local para determinar sus acciones. Lo hicimos en el caso de la versión de las esposas infieles y dimos un esbozo de lo que ocurre en la versión de los tres sabios.

Nos dimos cuenta de que el *conocimiento común* juega un papel importante en el razonamiento que hacen los sabios para determinar qué color tienen asignado; el conocimiento común de un hecho se determina como que todos lo saben, todos saben que todos lo saben, todos saben que todos saben que todos lo saben, etc. En un inicio del acertijo es muy relevante lo que se asume que es conocimiento común. En este caso asumimos que hay conocimiento común de que todos los sabios ven a todos, de que todos oyen a todos y de que el rey les enseña la misma caja a todos. El conocimiento común ocurre porque estando en un punto estos hechos son verdaderos en todos los puntos accesibles para el conjunto de sabios.

Basamos el análisis del acertijo de los esposos infieles en el conocimiento común de los sabios sobre el número de sabios con cada color que hay en la configuración que el rey escogió. Vimos que un sabio sabe su color si y sólo si hay conocimiento común del número de sabios que hay con ese color. Es importante notar que el conocimiento común puede alcanzarse en el sistema porque estamos suponiendo que tenemos un sistema síncrono y que no hay fallas en los canales de comunicación.

Con base en este análisis, presentamos un protocolo que implementa al programa basado en conocimiento; con este protocolo los sabios pueden determinar sus acciones a partir de su estado local. Para no modificar el estado local de cada sabio hicimos que en cada ronda, antes de responder a la pregunta del rey, un sabio calcule el número de sabios que ve con cada color y que calcule cómo ha evolucionado el conocimiento común, respecto al número de sabios que hay con cada color, revisando toda su historia de mensajes. Suponemos que cada sabio ha guardado una historia con todos los mensajes que ha enviado y todos los que ha recibido; es decir que cada sabio recuerda todas las respuestas que ha dado y todas las que ha escuchado. En una implementación más simple podríamos incluir una variable en el estado local de cada sabio, que podría ser un arreglo de enteros, en la que vaya guardando el estado del conocimiento común sobre el número de sabios con cada color. En cada ronda el sabio tiene que ir actualizando esta variable. También podríamos incluir otra variable en la que guarde el número de sabios que ve con cada color, un sabio calcula el valor de esta variable sólo una vez, al inicio de la ejecución.

Para la versión del acertijo de los tres sabios presentamos algunas ideas sobre cómo va cambiando el conocimiento común; aquí resulta útil pensar en el conocimiento común de los sabios que todavía no responden al rey. Hace falta formalizar y escribir bien los resultados que platicamos y también hace falta analizar lo que ocurriría si el rey decide volver a preguntar a los sabios, uno por uno, en el mismo orden en el que ya les preguntó. Quisieramos saber si un sabio que no supo su color en la primera vuelta de preguntas lo logra saber en alguna vuelta posterior.

Por medio del conocimiento fuimos descubriendo aquello que es relevante en el sistema. Logramos entender qué es lo que ocurre en las versiones del acertijo de las esposas infieles y del acertijo de los tres

sabios. Vimos que es muy importante pensar en cuántos sabios tienen cada color, en los colores inevitables y en los mínimos. Los colores inevitables corresponden a aquellos colores para los que sería posible que un sabio tuviera asignado ese color y que lo supiera desde antes que el rey comience a preguntar. Esto depende de la visión de los sabios y también va a depender de lo que saben los sabios sobre la visión de los demás. Pensamos que en nuevas versiones del acertijo seguirá habiendo conceptos equivalentes.

Todavía no queda muy claro la aplicación de los resultados y la relación de los acertijos con los sistemas distrubuidos. En primera instancia sería más fácil que un sabio le pida a algún otro sabio que le mande su visión. Así, con sólo dos mensajes, el sabio puede saber su color. ¿Por qué es necesario todo el intercambio de mensajes en el acertijo? En otras versiones, con mayores restricciones a las condiciones del problema, esta solución no sería posible. Puede ocurrir que se restrinja las condiciones de lo que ven los sabios, de manera que ya no sea cierto que todos se ven a todos. Inclusive los sabios podrían ya no saber cómo es la visión de los demás. También puede ser que deje de ser cierto que todos oyen a todos, es decir que ya no haya una línea de comunicación entre todos los procesadores. En el acertijo de las esposas infieles el rey prohíbe que los esposos se comuniquen entre ellos; también puede ocurrir que los procesadores no se pasan información entre ellos porque no confían en los demás o porque por alguna razón eso no está permitido.

El rey juega un papel especial en el sistema, en cada ronda escoge de manera aleatoria al conjunto de sabios a los que les pregunta. Su función es determinar cuándo un sabio puede tomar una acción. En algún momento se pensó en modelar al sistema de manera que el rey tuviera participación en el medio ambiente, decidiendo cuándo puede actuar un procesador. Sin embargo resultó forzado hacerlo así.

Podríamos modificar el protocolo del medio ambiente para modelar sistemas en los que los canales de comunicación fallen o sistemas en los que algún procesador tiene algún tipo de falla. Poríamos pensar en cambiar la visión de un sabio para obtener versiones del acertijo en las que los sabios no saben el color asignado a otros sabios, serían versiones en las que los sabios no ven a todos lo demás. Si pensamos en una gráfica de visión que represente lo que ven los sabios, la gráfica

de visión dejaría de ser conocimiento común. De hecho podría ocurrir que un sabio sí pueda verse a si mismo y entonces para un acertijo más general cambiaríamos el programa del rey, de manera que en cada ronda el rey escoga un conjunto de sabios a los que les va a preguntar y a cada sabio le pregunte si sabe el color de algún otro conjunto de sabios.

Los resultados obtenidos son bastante interesantes por sí mismos. Pero también nos pareció importante entender bien las dos versiones del acertijo que estudiamos, en las que se tiene una situación de alguna manera ideal. Pensamos que estas ideas nos servirán para ir analizando otras versiones del acertijo en las que haya una relación más clara con los sistemas distribuidos.

Bibliografía

[1] G. Attardi, M. Simi. "Reasoning across viewpoints", *Proceedings of ECAI 84*, T. O'Shea (ed.), North Holland: 315-325, 1984.

[2] R. Aumann. "Agreeing to disagree", *Annals of Statistics*,4(6): 1236-1239, 1976.

[3] J. Barwise. *The Situation in Logic*, Lecture Notes (17), Center for the Study of Language and Information, EUA, 1989.

[4] R. Brafman, J. Latombe y Y. Shoham. "Towards knowledge-level analysis of motion planning", *Proceedings of the National Conference on Artificial Intelligence (AAAI)*: 670-675, 1993.

[5] A. Brandenburger. "The role of common kowledge assumptions in game theory", *The Economics of Information, Games and Missing Markets*, F. Hahn (ed.), Oxford University Press, GB, 1989.

[6] L. Carlucci Aiello, D. Nardi y M. Schaerf. "Yet another solution to the three wisemen puzzle", *Proceedings of the Third International Symposium of Methodologies for Intelligent Systems (ISMIS)*, Z. W. Ras y L. Saitta (eds.), Elsevier Science: 398-407, 1988.

[7] L. Carlucci Aiello, D. Nardi y M. Schaerf. "Reasoning about knowledge and ignorance", *Proceedings of the International Conference of Fifth Generation Computer Systems*: 618-627, 1988.

[8] A. Cimatti y L. Serafini. "Multi-agent reasoning with belief context: The approach and a case study", *Proceedings of ECAI Workshop on Agent Theories, Architectures and Languages*: 71-85, 1994.

[9] K. Chandy y J. Misra. "How processes learn", *Distributed Computing*, 1(1): 40-52, 1986.

[10] H. H. Clark y C. R. Marshell. "Definite reference and mutual knowledge". *Elements of Discourse Understanding*, A. K. Joshi, B. L. Webber e I. A. Sag (eds.), Cambridge University Press, Cambrigde, Gran Bretaña, 1981.

[11] B. Chellas. *Modal Logic. An Introduction*, Cambridge University Press, Cambridge, Gran Bretaña, 1980.

[12] P. Coscia, P. Franchesci, G. Levi, G. Sardu y L. Torre. "Object level reflection of inference rules by partial evaluation", *Meta-level Architectgures and Reflection*, P. Maes y D. Nardi (eds.), North Holland, 1988.

[13] H. Delong. *A Profile of Mathematical Logic*, Addison-Wesley, EUA, 1971.

[14] C. Dwork y Y. Moses. "Knowledge and common knowlede in a byzantine environment: crash failures", *Information and Computation*, 88(2): 156-186, 1990.

[15] J. Elgot-Drapkin. "A real-time solution to the three wise-men problem", *Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 1991.

[16] J. Elgot-Drapkin. "Step-logic and the three-wise-men problem", *Proceedings of the ninth National Conference on Artificial Inteligence (AAAI)*: 412-417, AAAI Press, 1991.

[17] H. Enderton. *Una introducción matematica a la lógica*, UNAM, México, 1987.

[18] R. Fagin y J. Halpern. "Belief, awareness and limited reasoning", *Artificial Intelligence*, 34: 39-76, 1988

[19] R. Fagin, J. Halpern y M. Vardi. "A model-theoretic analysis of knowledge", *Journal of the ACM*, 38(2): 382-428, 1991.

[20] R. Fagin, J. Halpern y M. Vardi. "What can machines know? On the properties of knowledge in distributed systems", *Journal of the ACM*, 39:328-376, 1992.

[21] R. Fagin, J. Halpern , Y. Moses y M. Vardi. *Reasoning about Knowledge*, MIT Press, Cambridge, EUA, 1995.

[22] R. Fagin, J. Halpern , Y. Moses y M. Vardi. "Common knowledge revisted", *Proceedings of the Sixth Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*:283-298, Y. Shoham (ed.), Morgan Kaufmann, 1996.

[23] R. Fagin, J. Halpern , Y. Moses y M. Vardi. "Common knowledge: Now you have it, now you don't", *Proceedings of the International Multidisciplinary Conference on Intelligent Systems: A Semiotics Perspective*,Vol. 1: 177-183, 1996.

[24] R. Fagin, J. Halpern, Y. Moses y M. Vardi. "Knowlgede-based programs", *Distributed Computing*, 10(4): 199-225, 1997.

[25] R. Fagin y M. Vardi. "Knowlgede and implicit knowledge in a distributed environment: Preliminary Report", *Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*: 187-206, J. Halpern (ed.), Morgan Kaufmann, 1986.

[26] M. Fischer y N. Immerman. "Foundations of knowledge for distributed systems", *Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*:171-185, J. Halpern (ed.), Morgan Kaufmann, 1986.

[27] G. Gamow y M. Stern. *Puzzle-Math*, Viking, EUA, 1958.

[28] L. T. F. Gamut. *Logic, Language and Meaning. Volume 2. Intensional Logic and Logical Grammar*, University of Chicago Press, EUA, 1991.

[29] M. Gardner. "The 'jump proof' and its similarity to the toppling of a row of dominoes", *Scientific American*, 236: 128-135, 1977.

[30] M. Gardner. *Juegos y enigmas de otros mundos*, Gedisa Editorial, México, 1984.

[31] M. Gardner. *Penrose tiles to trapdoor ciphers*, W. H. Freeman and Co., NY, EUA, 1989.

[32] P. J. Gmytrasiewicz y E. H. Durfee. "A Logic of knowledge and belief for recursive modeling: Preliminary report", *Proceedings of the National Conference on Artificial Intelligence (AAAI)*: 628-634, 1992.

[33] J. Halpern. "Reasoning about knowlgede: An overview", *Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*: 1-17, J. Halpern (ed.), Morgan Kaufmann, 1986.

[34] J. Halpern. "Using reasoning about knowlgede to analyze distributed systems", *Annual Review of Computer Science* 2: 37-68, 1987.

[35] J. Halpern. "A note on knowledge-bases programs and specifications", *IBM Research Report RJ8454*, 1991

[36] J. Halpern. "Reasoning about knowledge: A survey", *Handbook of Logic in Artificial Inyelligence and Logic Programming*, 4: 1-34, D. Gabbay, C. J. Hogger y J. A. Robinson (eds.), Oxford University Press, 1995

[37] J. Halpern. "A Logical approach to reasoning about uncertainity: A tutorial", *Discourse, Interaction, and Communication*, X. Arrazola, K. Korta y F. J. Pelletier (eds.), Kluwer, 1997

[38] J. Halpern y R. Fagin. "A formal model of knowledge, action, and communication in distributed systems: preliminary report", *Proceedings of the Fourth ACM Symposium on Principles of Distributed Computing*: 224-236, 1985

[39] J. Halpern y R. Fagin. "Modelling knowledge and action in distributed systems", *Distributed Computing*, 3: 159-177, 1989

[40] J. Halpern y M. Vardi. "The complexity of reasoning about knowledge and time. I. Lower Bounds", *Journal of Computer and System Sciences*, 38: 195-237, 1989

[41] J. Halpern y Y. Moses. "Knowledge and common knowledge in a distributed environment", *Journal of the ACM*, 37: 549-587, 1990

[42] J. Halpern y Y. Moses. "A guide to completeness and complexity for modal logics of knowledge and belief", *Artificial Intelligence*, 54: 319-379, 1992

[43] J. Halpern, Y. Moses y M. Vardi. "Algorithmic knowledge ", *Proceedings of the Fifth Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*: 255-266, Morgan Kaufmann, 1994.

[44] J. Halpern y M. Vardi. "Model checking vs: theorem proving: A manifesto", *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning*, J. A. Allen, R. Fikes y E. Sandewall (eds.), Morgan Kaufmann, 1991.

[45] J. Halpern y L. Zuck. "A little knowledge goes a long way: Knowlegde-based derivations and correctness proofs for a family of protocols", *Journal of the ACM*, 39: 449-478, 1992.

[46] J. Halpern, R. van der Meyden y M. Vardi. "Complete axiomatizations for reasoning about knowledge and time", 1997.

[47] M. Herlihy y S. Rajsbaum. "Algebraic topology and distributed computing: A primer", capítulo en *Computer Science Today,* Jan van Leeuwen (ed.), LNCS, 100, Springer Verlag: 203-217, 1995

[48] J. Hintikka. *Knowledge and Belief*, Cornell University Press, EUA, 1962.

[49] R. Hilpinen (ed.). *Deontic Logic: Introductory and Systematic Readings*, D. Reidel Publishing Company, Dordrecht, 1971.

[50] G. Hughes y M. Cresswell. *An Introduction to Modal Logic*, Methuen, Londres, Gran Bretaña, 1968.

[51] K. Konolige. *Belief and Incompleteness*, Technical Report 319. SRI International, 1984.

[52] K. Konolige. *A Deduction Model of Belief*, Morgan Kauffmann, San Francisco, CA, EUA, 1986.

[53] K. Konoloige. "Explanatory belief ascription (Notes and premature formalization)", *Proceedings of the Third Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*: 207-222, Morgan Kaufmann, 1990.

[54] S. Kraus y D. Lehman. "Knowledge, belief and time", Technical Report 87-4, Department of Computer Science, Hebrew University, Jerusalem, Israel, 1987.

[55] R. Kurki-Suonio. "Towards programming with knowledge expressions", *Proceedings 13th ACM Symposium on Principles of Programming Languages*: 140-149, 1986.

[56] S. Kripke. "A semantical analysis of modal logic", *Zeitschr. f. math. Logik und Grundlagen d. Math*, 9: 67-96, 1963.

[57] R. Ladner y J. Reif. "The Logic of distributed protocols. (Preliminary report)", *Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*: 207-222, J. Halpern (ed.), Morgan Kaufmann, 1986.

[58] D. J. Lehman. "Knowledge, common knowledge, and related puzzless", *Proceedings of the Third ACM Symposium on Principles of Distributed Computing*: 62-67, 1984,

[59] D. Lewis. *Convention, a Philosophical Study*, Harvard University Press, EUA, 1969.

[60] N. Lynch y M. Fischer. "On describing the behavior and implementation in distributed systems", *Theoretical Computer Sciene*, 13: 17-43, 1981.

[61] J. L. Martínez. *Nezahualcóyotl*, Colección Lecturas Mexicanas 39, Fondo de Cultura Económica, México, 1984.

[62] J. McCarthy. "Formalization of two puzzles involving knowledge", *Technical Report*, Computer Science Department, Stanford University, 1978.

[63] J. McCarthy, M. Sato, T. Hayashi y S. Igarishi. "On the model theory of knowledge", *Technical Report STAN-CS-78-657*, Computer Science Department, Stanford University, 1979.

[64] L. Morgenstern. "A First order theory of planning, knowledge, and action", *Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*:99-114, J. Halpern (ed.), Morgan Kaufmann, 1986.

[65] S. Morley. *La civilización maya*, Fondo de Cultura Económica, México, 1972.

[66] Y. Moses. "Resource-bounded knowledge", *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*, M. Vardi (ed.), Morgan Kaufmann, EUA, 1988.

[67] Y. Moses, D. Dolev y J. Halpern. "Cheating husbands and other stories", *IBM Research Report RJ4756*, 1985

[68] Y. Moses y M. Tuttle. "Programming simultaneous actions using common knowledge", *Algorithmica*, 3: 121-169, 1988.

[69] A. Nozaki. *Anno's Hat Tricks*, Ilustrado por M. Anno. Philomel, 1985.

[70] G. Neiger y S Toueg. "Simulating real-time clocks and common knowledge in distributed systems", *Journal of the ACM*, 40(2): 334-367, 1993.

[71] G. Neiger y M. Tuttle. "Common Knowledge and consistent simultaneous coordination", *Lecture Notes in Computer Science*, 486: 334-352, 1991.

[72] E. Orlowska. "Logic for reasonong about knowledge", *Zeitschr. f. Math. Logik und Grundlagen d. Math*, 35: 559-572, 1989.

[73] R. Parikh. "Logic of knowledge, games and dynamic logic", *Lecture Notes in Computer Science*, 181: 202-222, 1984.

[74] R. Parikh y R. Ramanujan. "Distributed processes and the logic of knowledge (Preliminary report)", *Lecture Notès in Computer Science*, 193: 256-268, 1985.

[75] Y. Shoham. "Agent oriented programming", *Artificial Intelligence*, 60: 51-92, Elsevier, 1993.

[76] R. Stark. "A Logic of knowledge", *Zeitschr. f. Math. Logik und Grundlagen d. Math*, 27: 371-374, 1981.

[77] V. A. Uspenski. *Lecciones populares de matemáticas. Triángulo de Pascal*, Editorial Mir, Moscú, URSS, 1978.

[78] J. Van Benthem. *A Manual of Intensional Logic*, Lecture Notes (1), Center for the Study of Language and Information, segunda edición, EUA, 1988.

[79] R. Van der Meyden. "Constructing finite state implementations of knowledge-based programs with perfect recall", *Proceedings of PRICAI Workshop on Theoretical and Practical Foundations of Intelligent Agents*, Springer, 1996.

[80] R. Van der Meyden. "Finite state implementations of knowledge-based programs", *Lecture Notes in Computer Science*, 1180: 262-273, 1996.

[81] R. Van der Meyden. "Knowledge-based programs: On the complexity of perfet recall in finite environments", *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, Y. Shoham (ed.), Morgan Kaufmann, 1996.

[82] R. Van der Meyden. "Common knowledge and update in finite environments", por ser publicado en *Information and Computation*, R. Fagin (ed.), Morgan Kaufmann, 1994.

[83] M. Vardi. "Implementing knowledge-based programs", *Proceedings of the Sixth Conference on Theoretical Aspects of Reasoning About Knowledge (TARK)*, Y. Shoham (ed.), Morgan Kaufmann, 1996.