Uma análise da segurança nas comunicações entre agentes inteligentes

Vitor Sobrinho da Fonseca, Nilson Mori Lazarin

[1]Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (Cefet/RJ)
Nova Friburgo, RJ – Brazil

vitor.sobrinho@aluno.cefet-rj.br, nilson.lazarin@cefet-rj.br

Abstract. *Multi-Agent Systems (MAS) have an approach for modeling solutions to complex and distributed problems based on communications between autonomous and cognitive agents. However, some communication protocols between different MAS are based on IoT gateways that do not implement security mechanisms. This article presents the exploitation of a vulnerability in this type of communication, making it possible to capture messages between systems and beliefs, plans and intentions of agents in migration.*

Resumo. *O uso de Sistemas Multiagentes (SMA) possuem uma abordagem para modelagem de soluções para problemas complexos e distribuídos, baseados nas comunicações entre agentes autônomos e cognitivos. Entretanto, alguns protocolos de comunicação entre SMA distintos, estão baseados no uso de gateways IoT que não implementam mecanismos de segurança. Neste artigo é apresentada a exploração de uma vulnerabilidade neste tipo de comunicação, possibilitando a captura de mensagens entre sistemas e de crenças, planos e intenções de agentes em migração.*

1. Introdução

Agentes são softwares autônomos capazes de interagir e tomar decisões para alcançar objetivos individuais e/ou coletivos que compõem um Sistema Multiagentes (SMA) [Wooldridge 2009]. Esses sistemas podem ser abertos ou fechados. Um SMA fechado só permite a interação entre agentes do mesmo sistema, por outro lado, um SMA aberto permite a interação e migração entre agentes móveis de diferentes SMA [Hubner 1995]. O ContextNet é um *middleware* para Internet das Coisas que atua como uma camada intermediária para fornecer suporte a atividades de contexto, permitindo a captura, processamento e disseminação de informações relevantes sobre o ambiente, além de possibilitar uma ampla quantidade de conexões [Endler et al. 2011]. Dessa forma, a aplicação do ContextNet pode prover uma interconexão entre SMA através da internet, permitindo a troca de mensagens e transferência de agentes, possibilitando a criação de um ecossistema digital inteligente [Jesus et al. 2018].

A comunicação é uma parte fundamental no funcionamento de um SMA, pois permite que os agentes compartilhem crenças e planos, coordenem suas ações e realizem movimentações de forma colaborativa, além disso, podem trocar percepções sobre o ambiente em que estão inseridos [Jesus et al. 2021]. Além disso, a arquitetura do ContextNet viabiliza o crescimento da rede garantindo a escalabilidade da comunicação e baixa

latência, pois utiliza do protocolo MR-UDP para fornecer uma comunicação confiável baseada no UDP (*User Datagram Protocol*), com baixa sobrecarga e com recursos focados em mobilidade [David et al. 2012]. Esta é uma opção muito atraente para dispositivos IoT inteligentes, pois possuem recursos limitados de processamento e memória. No entanto, a confiabilidade da implementação desse mecanismo está ligada a integridade da comunicação e não à segurança. Ou seja, mesmo com a implementação desses mecanismos para melhorar o protocolo UDP as comunicações não possuem confidencialidade e autenticidade.

A falta de uma comunicação segura em um SMA gera preocupações e pode levar a consequências negativas, como vazamento de informações sensíveis, sabotagem, manipulação de dados, entre outras. Por exemplo, em um SMA utilizado para gerenciamento de tráfego, os SMA envolvidos podem estar embarcados em veículos autônomos e semáforos inteligentes. Dessa forma, um agente malicioso pode interceptar as comunicações entre os veículos autônomos, enviar informações falsas aos semáforos, criando condições de tráfego caóticas.

Para demonstrar a vulnerabilidade foram utilizadas duas maquinas virtuais (VM) executando a ChonIDE [Souza de Jesus et al. 2023], uma plataforma para programação de SMA com agentes inteligentes que utilizam o modelo BDI (*Belief-Desire-Intention*) [Bratman 1987] que dá suporte à comunicação entre diferentes SMA e migração de agentes baseada em protocolos bio-inspirados [Jesus et al. 2021]. Cada VM executará um SMA, e cada um deles terá um agente Comunicador para realizar a interação entre os sistemas. Essas VMs estarão funcionando em uma rede controlada, onde a comunicação será interceptada com o WireShark[1] a fim de capturar as mensagem e a migração de agentes com suas crenças, planos e intenções trafegados pela rede. Além disso, através do prompt de comando são enviadas falsas mensagens para o SMA destinatário. Dessa forma serão validados os ataques de confidencialidade e autenticidade entre SMA.

O objetivo deste trabalho é explorar uma vulnerabilidade de segurança na comunicação entre SMA que utilizam o ContextNet [Endler et al. 2011], demonstrando a captura de informações trocadas por eles com um ataque de homem-do-meio e classificar quais propriedades de segurança esse ataque inflige. Este trabalho está organizado da seguinte forma, na Seção 2 é apresentada uma fundamentação teórica. na Seção 3 será demonstrado dois cenários de ataques a SMA. Por fim na Seção 4 é apresentada uma discussão e próximos passos nesta pesquisa.

2. Fundamentação Teórica

Nesta seção serão apresentados os conceitos sobre o formato da mensagem entre agentes comunicadores e o ataque de homem no meio, necessários para o entendimento deste trabalho.

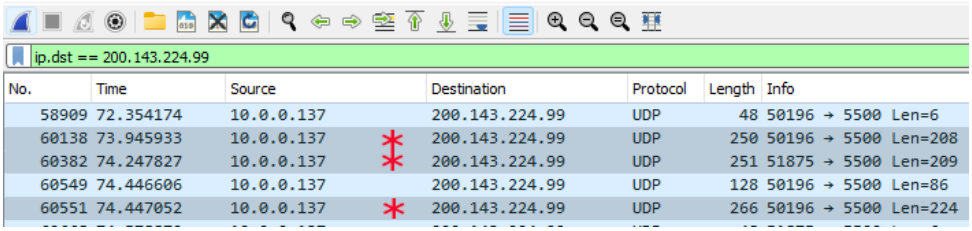
2.1. Formato da mensagem entre agentes Comunicadores

Agentes Comunicadores [Jesus et al. 2021], são uma extensão dos agentes Jason [Bordini and Hübner 2006] responsáveis por facilitar a comunicação entre diferentes SMAs por meio de uma rede IoT utilizando o ContextNet [Endler et al. 2011]. Cada

[1]https://www.wireshark.org/

agente Comunicador é identificado por um *Universally Unique IDentifier* (UUID) e emprega ações internas para enviar mensagens utilizando *Knowledge Query and Manipulation Language* (KQML) [Finin et al. 1994] para outros agentes comunicadores em diferentes SMAs. Além disso, esses agentes são capazes de migrar outros agentes ou a si para diferentes sistemas.

A Figura 1a apresenta a captura dos pacotes de rede utilizando o WireShark. O formato da mensagem enviada na rede, apresentado na Figura 1b é composto pelos seguintes campos: um preâmbulo fixo (*0xFFFE*); um campo para o tamanho em bytes do UUID de destino, seguido pelo endereço do UUID do destino; um campo para o tamanho em bytes da força ilocucionária, seguido pela força da mensagem; um campo para o tamanho em bytes da mensagem, seguido pela mensagem em si. O preâmbulo e os campos de tamanho são todos representados em hexadecimal.



(a) Captura de pacotes de rede utilizando o WireShark.

```
fffe24cdc19c19-5616-4fc5-8d54-372631dd8eff07achieve0cdamageReport"

fffe24788b2b22-baa6-4c61-b1bb-01cff1f5f87804tell0report("Deck 2")"

fffe24cdc19c19-5616-4fc5-8d54-372631dd8eff07achieve1cretransmit(scott,redAlertOn)"
```

(b) Conteúdo dos pacotes de comunicação entre o SMA e o ContextNet.

Figura 1. Captura dos pacotes transmitidos.

2.2. Ataque do homem no meio

O Ataque de Homem no meio é uma forma de ataque cibernético em que um terceiro malintencionado intercepta a comunicação entre dois pontos legítimos. Durante um ataque de homem no meio, o invasor se posiciona entre o remetente e o destinatário da comunicação, interceptando e possivelmente alterando as mensagens trocadas entre eles. O objetivo principal de um ataque de homem no meio é obter acesso não autorizado a informações confidenciais, como senhas, informações bancárias ou dados pessoais [Stallings 2008].

3. Ataque à comunicação entre agentes

Este trabalho demonstra a violação das propriedades de confidencialidade, integridade e autenticidade, através dos ataques de divulgação e de identidade falsa [Bijani and Robertson 2014] nas comunicações entre SMA. No ataque de divulgação conseguimos fazer a interceptação de mensagens dos agentes e capturar suas crenças, planos e intenções, já no ataque de identidade falsa, enviamos mensagens falsas ao SMA de destino para atrapalhar a execução de um plano. Nas subseções a seguir será demonstrado como foram realizados os ataques tanto na comunicação quanto na migração.

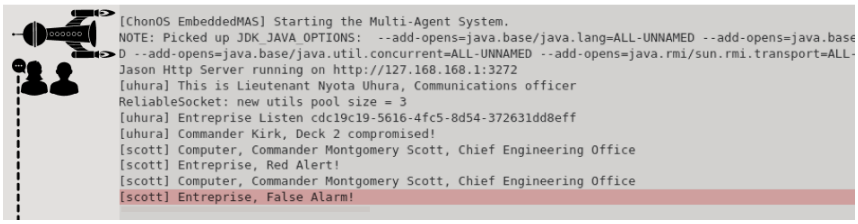
Os ataques foram realizados da seguinte forma: primeiramente executamos os SMAs nas VMs e deixamos os agentes Comunicadores interagirem normalmente na rede,

porém monitorando os envios dos pacotes utilizando a ferramenta WireShark. Logo após é feito uma análise nos pacotes interceptados onde é possível identificar o UUID e a mensagem enviados pelos agentes, dessa forma obtivemos dados suficientes para injetar mensagens falsas na rede e atrapalhar a comunicação. Abaixo são descritos dois cenários de ataque, um de comunicação e outro de migração entre agentes de SMA distintos.

3.1. Cenário 1: Ataque na Comunicação

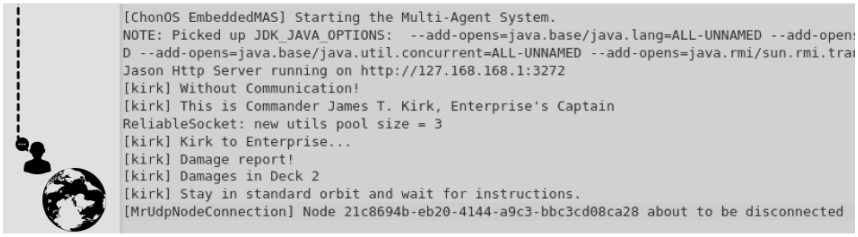
No primeiro cenário, um SMA (*ncc1701.mas2j*) com os agentes Scott e Uhura (Comunicador) representados na Figura 2a. O segundo SMA (*bajor.mas2j*) com o agente Kirk (Comunicador) representado na Figura 2c. O agente Kirk inicia a comunicação enviando mensagens para o agente Uhura que aguarda esse primeiro contato para identificar que foi estabelecido uma conexão. Posteriormente o agente Kirk solicita ao agente Uhura o relatório de danos, que responte que o Deck 2 está comprometido, Kirk pede para aguardar novas atualizações e envia uma crença (*RedAlert*) para ser encaminhada ao agente Scott, que imediatamente executa um plano.

Entretanto, o atacante capturou as comunicações e descobriu o UUID dos comunicadores Uhura e Kirk, uma vez que a comunicação não é criptografada. Através do terminal de comando, é enviada uma mensagem ao agente Uhura, identificando-se como agente Kirk, solicitando que uma crença *RedAlertOff* seja encaminhada ao agente Scott, cancelando a ação anterior do agente como vemos na Figura 2b.



(a) Log do SMA *ncc1701.mas2j*.

(b) Ataque de *Injeção de Mensagem Falsa*.



(c) Log do SMA *bajor.mas2j*.

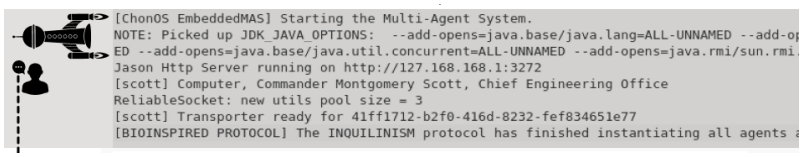
Figura 2. Ataque durante comunicação entre SMAs.

3.2. Cenário 2: Ataque na Migração

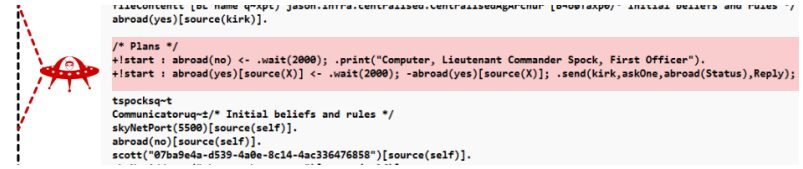
No segundo cenário, temos um SMA (ncc1701A.mas2j) com o agente Scott (Comunicador) representado na Figura 3a. O segundo SMA (andoria.mas2j) com os agentes

Kirk (Comunicador) e Spock representados na Figura 3c. O agente Scott está ativo na rede apenas aguardando mensagens. O agente Kirk inicia a comunicação com o agente Scott informando que irá iniciar uma transferência de agentes. O agente Scott confirma o recebimento da mensagem. Assim o agente Kirk ativa o protocolo Inquilinismo [Jesus et al. 2021] e chega ao SMA destino com sucesso.

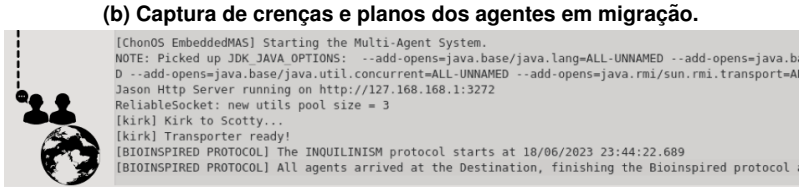
Porém, toda a comunicação foi interceptada e agora não só foi possível obter o UUID e as mensagens transmitidas na rede, pois como os agentes Kirk e Spock migraram, todas a suas crenças, planos e intenções foram capturadas como vemos na Figura 3b. Isso possibilita a criação de agentes clones que podem ser igualmente transmitidos para o SMA de destino.



(a) Log do SMA *ncc1701A.mas2j*.



(b) Captura de crenças e planos dos agentes em migração.



(c) Log do SMA *andoria.mas2j*.

Figura 3. Ataque durante comunicação entre SMAs.

4. Trabalhos futuros

Nesta etapa da pesquisa foi possível provar as vulnerabilidades na comunicação entre SMA que usam o middleware ContextNet. As próximas etapas consistem na formulação de uma abordagem para troca de chaves criptográficas, diretamente na dimensão dos agentes do SMA. Além disso, será necessária uma validação da conformidade dos padrões de segurança a serem adotados, tal como apresentado em [Rocha et al. 2020]. Por fim, uma será necessário criar mecanismos de comunicação segura, sobre meios inseguros de comunicação, tal como apresentado em [Freitas et al. 2021], especificamente para comunicação entre SMA que utilizam o ContextNet.

Referências

Bijani, S. and Robertson, D. (2014). A review of attacks and security approaches in open multi-agent systems. *Artif Intell Rev*, 42:607–636. DOI: 10.1007/s10462-012-9343-1.

Bordini, R. H. and Hübner, J. F. (2006). BDI Agent Programming in AgentSpeak Using Jason. In Toni, F. and Torroni, P., editors, *Computational Logic in Multi-Agent Systems*, pages 143–164, Berlin, Heidelberg. Springer Berlin Heidelberg. DOI: 10.1007/11750734_9.

Bratman, M. (1987). Intention, plans, and practical reason.

David, L., Vasconcelos, R., Alves, L., Andre, R., Baptista, G., and Endler, M. (2012). A large-scale communication middleware for fleet tracking and management. In *Simposio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2012), Salão de Ferramentas*, pages 964–971.

Endler, M., Baptista, G., Silva, L. D., Vasconcelos, R., Malcher, M., Pantoja, V., Pinheiro, V., and Viterbo, J. (2011). Contextnet: Context reasoning and sharing middleware for large-scale pervasive collaboration and social networking. In *Proceedings of the Workshop on Posters and Demos Track*, PDT '11, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2088960.2088962.

Finin, T., Fritzson, R., McKay, D., and McEntire, R. (1994). KQML as an Agent Communication Language. In *Proceedings of the Third International Conference on Information and Knowledge Management*, CIKM '94, page 456–463, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/191246.191322.

Freitas, J., Souza, L., Sardou, P., and Lazarin, N. (2021). Comunicação segura em VANET. In *Anais da XIX Escola Regional de Redes de Computadores*, pages 109–114, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/errc.2021.18551.

Hubner, J. F. (1995). *Migração de agentes em sistemas multi-agentes abertos*. Dissertação (Mestrado), Universidade Federal do Rio Grande do Sul. Instituto de Informática. Curso de Pós-Graduação em Ciência da Computação, Porto Alegre. `https://lume.ufrgs.br/handle/10183/25032`.

Jesus, V., Manoel, F., Pantoja, C. E., and Viterbo, J. (2018). Transporte de agentes cognitivos entre sma distintos inspirado nos princıpios de relaçoes ecológicas. In *Workshop-Escola de Sistemas de Agentes, seus Ambientes e apliCações—XII WESAAC*, pages 179–187.

Jesus, V. S. d., Pantoja, C., Manoel, F., Alves, G., Viterbo, J., and Bezerra, E. (2021). Bio-Inspired Protocols for Embodied Multi-Agent Systems. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART,*, pages 312–320. SciTePress. DOI: 10.5220/0010257803120320.

Rocha, I., Schott, R., Verly, P., and Lazarin, N. (2020). Análise de desempenho e conformidade em bibliotecas criptográficas para internet das coisas. In *Anais da VI Escola Regional de Sistemas de Informação do Rio de Janeiro*, Porto Alegre, RS, Brasil. SBC. `https://sol.sbc.org.br/index.php/ersi-rj/article/view/10118`.

Souza de Jesus, V., Mori Lazarin, N., Pantoja, C. E., Vaz Alves, G., Ramos Alves de Lima, G., and Viterbo, J. (2023). An IDE to Support the Development of Embedded Multi-Agent Systems. In Mathieu, P., Dignum, F., Novais, P., and De la Prieta, F., editors, *Advances in Practical Applications of Agents, Multi-Agent Systems, and Cognitive Mimetics. The PAAMS Collection*, pages 346–358, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-37616-0_29.

Stallings, W. (2008). *Criptografia e segurança de redes princípios e práticas*. Pearson Prentice Hall, São Paulo, 4. ed edition.

Wooldridge, M. (2009). *An introduction to multiagent systems*. John wiley & sons.