Uma proposta de arquitetura para identificação de anomalias em redes IoT utilizando registros de Logs

Jonatham O. Preuss[1], **Bolívar M. Silva**[1], **Raul C. Nunes**[1]

[1]Centro de Tecnologia – Universidade Federal de Santa Maria (UFSM)
Caixa Postal 1000 – 97105-900 – Santa Maria – RS – Brasil
Programa de Pós-Graduação em Ciência da Computação
Santa Maria, R.S.

{bolivar,jonathan.preuss}@redes.ufsm.br, ceretta@inf.ufsm.br

Resumo. *IoT tem sido amplamente empregado em infraestruturas críticas onde auxiliam no monitoramento, gestão e tomada de decisão. Com isso existe a preocupação em relação a segurança e integridade desses ambientes, a falha desses dispositivos IoT em ambientes críticos podem causar uma catástrofe. Devido à grande quantidade e heterogeneidade de dados gerados em ambientes IoT, métodos tradicionais para detecção de anomalias não são eficazes. Este artigo apresenta uma arquitetura para identificação de anomalias em dispositivos IoT, através da análise de logs utilizando ferramentas de Big Data. A arquitetura propostas, apresenta uma metodologia de análise para detecção de anomalias em tempo hábil, para grandes volumes de dados heterogêneos de redes IoT.*

Abstract. *IoT has been widely used in critical infrastructures where they assist in monitoring, management and decision making. With this there is concern regarding the security and integrity of these environments, the failure of these IoT devices in critical environments can cause a catastrophe. Due to the large amount and heterogeneity of data generated in IoT environments, traditional methods for detecting anomalies are not effective. This article presents an architecture for identifying anomalies in IoT devices, through the analysis of logs using Big Data tools. The proposed architecture presents a methodology of analysis for the detection of anomalies in a timely manner, for large volumes of heterogeneous data of IoT networks.*

1. Introdução

A Internet das Coisas (*IoT*) apesar de não ser um tema tão recente, tem sido amplamente estudado e discutido pela comunidade científica, principalmente, nos últimos anos. Um ambiente de *IoT* é composto por dispositivos e tecnologias diferentes, consequentemente é heterogênea e volumosa a geração de dados nesses ambientes [Umar Ahsan 2016]. Segundo [Qian Zhu 2010], em redes *IoT*, são necessários *gateways IoT*, que são encarregados de tornar possível a interação entre elementos de redes e tecnologias heterogêneas. Uma das aplicações do paradigma de *IoT* é o seu emprego em atividades de auxílio e monitoramento de infraestruturas críticas, cuja a falha de um dispositivo pode resultar em perdas financeiras ou catástrofes. Exemplos de uso são demonstrados em [Dan Koo 2015], que faz uso de *IoT* para o sensoriamento de uma estação de água. E

em [Kyle E. Benson 2016] que utiliza uma rede *IoT* para o monitoramento de atividades sísmicas.

Em paralelo ao amplo emprego de sistemas *IoT*, surgem vários problemas no que diz respeito à segurança, como vulnerabilidades de autenticidade e integridade dos dados e dispositivos. Devido às características simplistas dos *hardwares* utilizados em muitos dos dispositivos que compõem a *IoT*, as técnicas de segurança, tradicionais tornam-se inviáveis [Jing et al. 2014], e se tratando de ambientes de infraestruturas críticas, a agilidade na detecção de anomalias e falhas, podem evitar desastres. Nesse contexto, este trabalho apresenta uma proposta de arquitetura que visa prover segurança e integridade de redes *IoT*, através da detecção de comportamentos anômalos de dispositivos *IoT*, essa detecção é realizada através da análise de eventos de *log* gerados por dispositivos conectados aos *gateways*. O presente trabalho está organizado da seguinte forma. Seções 2. Revisão Bibliográfica, tratando os temas abordados ao longo deste trabalho, seções de Trabalhos Relacionados, demonstrando alguns trabalhos os quais autores tratam de um tema semelhante ao deste trabalho, seção 4. Arquitetura Proposta será descrito a arquitetura proposta e seu funcionamento e por fim as seções de 5. Conclusão e 6. Trabalhos Futuros.

2. Revisão Bibliográfica

Essa seção apresenta uma breve descrição dos temas abordadas nesse trabalho, visando fornecer embasamento ao leitor sobre o que se tratam. Internet das coisas é um conceito onde elementos presentes no dia a dia passam a interagir através de uma rede de dados. Em [Pallavi Sethi 2017], os autores definem *IoT* como sendo um aglomerado de diferentes tecnologias que operam em conjunto. Segundo [Janice Canedo 2016], em ambientes de *IoT*, os dois principais desafios quanto a detecção de anomalias, são: a heterogeneidade dos elementos e o grande volume de dados gerados.

Em um contexto onde existem dados diferentes sendo gerados em grande quantidade e com alta velocidade, uma boa abordagem para a análise desses dados é utilizar o conceito de *Big Data*. Em [McKinsey Global Institute 2011], o conceito de *Big Data* refere-se ao grupo de dados que não poderiam ser obtidos, administrados e armazenados pelos tradicionais sistemas de banco de dados. Existe uma variedade de ferramentas e sistemas para trabalhar com esse conceito. Nesse trabalho são utilizadas as ferramentas, Apache Hadoop[1], Hadoop File System[2], Apache Flume[3] e Apache Spark[4].

3. Trabalhos Relacionados

Existe um grande número de trabalhos que propõem as mais variadas técnicas e algoritmos para detecção de anomalias em ambientes *IoT*, como por exemplo técnicas de inspeção de pacotes apresentado em [Douglas H. Summerville and Chen 2015], ou técnicas de *data mining* apresentadas por [Janice Canedo 2016]. Porém, esses trabalhos se direcionam para a detecção de anomalias em aplicações específicas. Diante de ambientes com grandes diversidades tecnológicas, torna-se difícil o emprego dessas técnicas.

[1]Apache Hadoop "https://hadoop.apache.org"
[2]Hadoop File System "https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html"
[3]Apache Flume "https://flume.apache.org/"
[4]Apache Spark "https://spark.apache.org/"

Alguns autores trabalham com técnicas mais genéricas para detecção, como em [Fu et al. 2011], onde o autor ressalta que a heterogeneidade de uma rede *IoT*, dificulta a implantação de sistemas convencionais de detecção de intrusão. Os autores analisam o comportamento padrão dos dispositivos de *IoT* em relação ao tempo, e sugerem cinco características comportamentais: pulso, subida acentuada, descida acentuada, leitura constante e mudança de padrão, que classificam um comportamento como anômalo. Em [Yanbing Liu 2014] os autores propõem um algoritmo para detecção de intrusão em redes *IoT* utilizando análise de similaridade de pacotes. Esse algoritmo utiliza coeficiente *Jaccard* para realizar as medidas de similaridade entre um conjunto de treinamento sem possíveis anomalias e o tráfego de uma rede *IoT*.

Nos trabalhos apresentados nessa sessão, os autores utilizam algoritmos e técnicas genéricas para detecção de anomalias em ambientes *IoT*, baseando-se na análise comportamental do tráfego, porém não são trabalhadas formas de coleta de dados, além de não utilizar nenhuma técnica para execução dos algoritmos de detecção, diante de um grande fluxo de dados. O presente trabalho propõe uma arquitetura que realiza o emprego de uma ferramenta para a coleta das informações dos dispositivos *IoT* e efetua o processamento desses dados, onde é possível realizar análises comportamentais ou de similaridades, de forma ágil, aproximando-se a uma análise em tempo real.

4. Arquitetura Proposta

O escopo desse trabalho engloba a Internet das Coisas e os diversos dispositivos que a compõe. Devido as características já citadas desse tipo de rede, como as diversas fontes heterogêneas, a arquitetura apresentada traz as tecnologias e técnicas necessárias para alcançar o objetivo proposto inicialmente.

O Apache Flume possui a habilidade de operar com múltiplos *Agents* coletando informações de diferentes fontes. Dessa forma, em um cenário onde existam aplicações heterogêneas conectadas a um mesmo *gateway*, é possível criar e configurar agentes individuais para cada aplicação e seus respectivos dispositivos. Uma vez que os *gateways* estão concentrando o fluxo de dados da rede em um único ponto, posicionar agentes Flume nesses pontos garantem acesso aos *logs* de eventos de todos os dispositivos conectados nesse *gateway*. A figura 1 ilustra a arquitetura proposta.

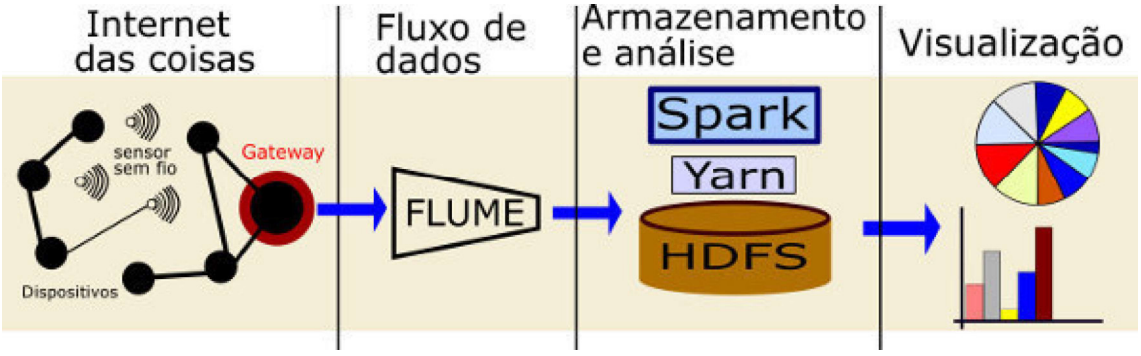

Figura 1. Arquitetura proposta

Da esquerda para a direita, temos os dispositivos que compõem as redes *IoT*, representados por círculos pretos. Os círculos maiores (com borda vermelha) representam

os dispositivos *gateways*, que concentram as informações de determinados conjuntos de sensores. Seguindo a sequência, temos o Apache Flume, que coleta, agrega e transporta os dados dos *gateways*, criando um fluxo de informações e encaminhando para ser processado (Spark) e armazenado (HDFS). Finalmente, à direita da figura, é apresentado a parte da arquitetura referente à visualização. Diz respeito a interface com o usuário. É onde serão apresentados os alertas e demais tipos de informações, pertinentes ao administrador da rede.

Após configurar os agentes Flume em cada ponto de coleta e iniciar o processo de transmissão dos dados de *log* para destino onde está operando o ambiente Spark, inicia-se a fase de tratamento desse fluxo. O primeiro elemento a tratar esses dados é o Spark *Streaming* que irá fragmentar esse bloco de dados em pequenas partes e os encaminhar para um segundo elemento que é o Spark *engine*, onde é executado a aplicação de análise. O fluxograma representado na figura 2 exemplifica as funções que a aplicação necessita para operar adequadamente.

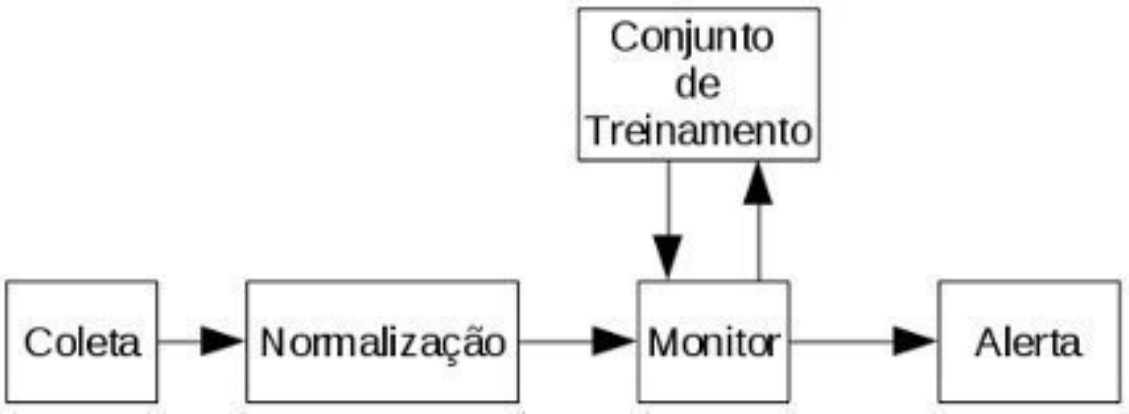

Figura 2. Fluxograma de funções da aplicação utilizada para análise de anomalias

A aplicação de análise representada na figura 2 exemplifica as funções que a aplicação necessita para operar adequadamente. A primeira função é chamada de ”coleta”. Essa função irá delimitar o contexto do Spark, indicando onde estão localizados os fragmentos de dados oriundos do *Streaming*, e delimitará o modo de operação do *cluster* Spark. A segunda função é a de normalização. Essa função irá extrair e normalizar os dados úteis, de acordo com a necessidade de cada tipo de dado. Após isso, os monitores armazenarão na base de dados HDFS.

Para cada aplicação monitorada pela arquitetura proposta nesse trabalho, deve existir um conjunto de treinamento que contenha dados normalizados de comportamento tido como normal para aquela aplicação. Baseado nesse conjunto de treinamento a função do monitor é comparar os dados provenientes de um fluxo e classificá-los como anômalo ou normal. Nesse ponto é executado a análise do comportamento dos dados, baseado nas cinco características definidas por [Fu et al. 2011]. A última função da aplicação é o Alerta, que é responsável por sinalizar a ocorrência de uma anomalia no comportamento de um dispositivo. O Alerta pode ser configurado como uma mensagem em uma página *web* sinalizando em qual dispositivo e o tipo de anomalia que está ocorrendo.

5. Conclusão

O uso de Internet das Coisas para aplicações críticas está crescendo nos últimos anos, com isso surge a preocupação com a integridade e segurança desses ambientes. Devido à grande quantidade e principalmente diferença dos dados gerados pelos dispositivos das redes *IoT*, técnicas de detecção anomalias convencionais não são bem sucedidas. Através do estudo dos trabalhos relacionados e das ferramentas existentes, para processamento e análise rápida de grandes quantidades de dados, foi apresentado uma arquitetura que pode suprir algumas das dificuldades apresentadas de detecção de anomalias no contexto de *IoT*. Dado as características citadas da arquitetura proposta, enfatizando a capacidade de coleta, análise de grandes quantidades de dados heterogêneos, espera-se que a arquitetura proposta realize a análises de possíveis falhas, em tempo de execução ou tão próximo quanto possível. Dessa forma, redes *IoT* com sistemas sensíveis que necessitam ser altamente tolerantes a falhas, podem reduzir o tempo necessário para a identificação de falhas, bem como na prevenção das mesmas.

6. Trabalhos Futuros

Como trabalhos futuros, pretende-se implementar a arquitetura proposta em um cenário de testes, para validação e análise da real viabilidade de utilização do modelo proposto. Para resultados mais precisos, o ideal seria utilizar um cenário com o maior número de dispositivos possível, como sensores de temperatura, sensores de pressão, câmeras de vigilância entre outros.

Referências

Dan Koo, Kalyan Piratla, J. M. C. (2015). Towards sustainable water supply: Schematic development of big data collection using internet of things (iot). *International Conference on Sustainable Design, Engineering and Construction*.

Douglas H. Summerville, K. M. Z. and Chen, Y. (2015). Ultra-lightweight deep packet anomaly detection for internet of things devices. In *Computing and Communications Conference (IPCCC), 2015 IEEE 34th International Performance*. IEEE.

Fu, R., Zheng, K., Zhang, D., and Yang, Y. (2011). An intrusion detection scheme based on anomaly mining in internet of things.

Janice Canedo, Anthony Skjellum, S. G. C. o. E. (2016). Using machine learning to secure iot systems. In *Privacy, Security and Trust (PST), 2016 14th Annual Conference on*. IEEE.

Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., and Qiu, D. (2014). Security of the internet of things: Perspectives and challenges. *Wireless Networks*, 20(8):2481–2501.

Kyle E. Benson, Qing Han, K. K. P. N. N. V. (2016). Resilient overlays for iot-based community infrastructure communications. *Internet-of-Things Design and Implementation (IoTDI), 2016 IEEE First International Conference on*.

McKinsey Global Institute, James Manyika, M. C. B. B. J. B. R. D. C. R. A. H. B. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 1th edition.

Pallavi Sethi, S. R. S. (2017). Internet of things: Architectures, protocols, and applications. In *Journal of Electrical and Computer Engineering Volume 2017 (2017)*. Hindawi.

Qian Zhu, Ruicong Wang, Q. C. Y. L. W. Q. (2010). Iot gateway: Bridging wireless sensor networks into internet of things. *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*.

Umar Ahsan, A. B. (2016). A review on big data analysis and internet of things. In *2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems*. IEEE.

Yanbing Liu, Q. W. (2014). A lightweight anomaly mining algorithm in the internet of things. *Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on*.

