

## Medidas de Qualidade para a Avaliação da Confiança no Funcionamento em Sistemas Ubíquos

Andressa Bezerra Ferreira<sup>1, 2, 3</sup>, Reinaldo Bezerra Braga<sup>1, 3</sup>, Rossana Maria de Castro Andrade<sup>1, 2, \*</sup>

<sup>1</sup>Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat),

<sup>2</sup>Mestrado e Doutorado em Ciência da Computação (MDCC), Universidade Federal do Ceará (UFC)

<sup>3</sup>Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

{rossana, andressaferreira}@great.ufc.br,  
reinaldo.braga@ifce.edu.br

**Abstract.** *Ubiquitous systems are, in principle, autonomous and capable of perceiving the environment, and user activities and needs. However, this perception and decision autonomously demands user confidence in the system operation what makes difficult the acceptance and use of this type of system. To ensure that trust is a difficult and even more challenging task in ubiquitous systems due to their peculiarities (e.g., context-awareness). One of the possible techniques for the assessment of this feature is to apply measurements. This article presents then a set of thirteen quality measures to evaluate the confidence in the functioning (i.e., the dependability characteristic) in ubiquitous systems. To evaluate this set of measurements, three case studies are used and also described in this paper.*

**Resumo.** *Sistemas ubíquos são, por princípio, autônomos e capazes de perceber o ambiente, e as atividades e necessidades do usuário. Entretanto, esta percepção e tomada de decisão de forma autônoma faz com que a aceitação e utilização desse tipo de sistema demande confiança no funcionamento por parte do usuário. Garantir essa confiança é uma tarefa difícil e ainda mais desafiadora nesses sistemas devido às suas peculiaridades (e.g., mobilidade, sensibilidade ao contexto). Uma das possíveis técnicas para a avaliação da confiança no funcionamento em sistemas ubíquos é o uso de medições e este artigo apresenta um conjunto de treze medidas de qualidade para essa avaliação bem como estudos de caso.*

### 1. Contextualização

Sistemas ubíquos são, por princípio, autônomos e capazes de perceber o ambiente, as atividades e necessidades do usuário, sem que seja necessário determiná-las explicitamente.

\*Bolsista do CNPq de Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora (DT) 2

A autonomia desses sistemas é possível devido ao uso da sensibilidade ao contexto, pois o usuário não precisa realizar uma operação explícita (e.g., clicar em um botão) para que o sistema entenda isso como uma entrada. Geralmente, ações cotidianas (e.g., entrar em um ambiente) são suficientes para que o sistema reaja provendo serviços (e.g., ligar o ar-condicionado de maneira automática) (DEY 2001) (LIMA 2011) (DEBASHIS e AMITAVA 2003). Em resumo, as aplicações ubíquas devem perceber o contexto no qual o usuário está inserido e tomar uma decisão que seja a mais adequada possível. Entretanto, esta tomada de decisão de forma autônoma pelo sistema requer que seja assegurado ao usuário final a confiança no seu funcionamento<sup>1</sup> (característica chamada em inglês de *dependability*), garantindo assim uma melhor aceitação do mesmo e com isso uma maior utilização desse tipo de sistema.

*Dependability* (i.e., confiança no funcionamento) é um termo definido por (AVIZIENIS et al 2004) como a capacidade de desenvolver um sistema no qual se pode justificadamente confiar. Esta confiança pode ser justificada com base nos atributos de disponibilidade (*availability*), confiabilidade (*reliability*), proteção (*safety*), confidencialidade (*confidentiality*), integridade (*integrity*) e manutenibilidade (*maintenability*). O termo é proposto como representação dos esforços das comunidades de confiança no funcionamento e segurança (*security*) buscando unificar vários aspectos dessas duas áreas, servindo como uma espécie de guia de termos e conceitos para pesquisadores das áreas de engenharia de software e redes de computadores.

Garantir a confiança no funcionamento de sistemas é uma tarefa difícil, pois esta característica depende de diversos atributos de qualidade. Além disso, devido às peculiaridades (e.g., sensibilidade ao contexto, mobilidade e heterogeneidade) inerentes aos sistemas ubíquos, essa questão se torna ainda mais desafiadora do que em sistemas convencionais.

Uma das possíveis técnicas para a avaliação da confiança no funcionamento de um software é a realização de medições (SCHOLTZ e CONSOLVO 2004) e (SANTOS et al 2013). Uma medição consiste no processo contínuo de definição, coleta e análise de dados sobre o processo de desenvolvimento de software e seus produtos, fornecendo informações significativas com o objetivo de melhorá-los (ISO 25000). O resultado da medição é chamado de medida.

Nesse cenário, este artigo apresenta um conjunto de treze medidas de qualidade de software, definidas por meio da metodologia *Goal-Question-Metric* (GQM) (BASILI et al 1994), para avaliação da confiança no funcionamento em sistemas ubíquos que executam em dispositivos móveis.

Como contribuição, espera-se que este trabalho seja utilizado como apoio à atividade de avaliação da confiança no funcionamento em sistemas ubíquos com foco em aplicações rodando em dispositivos móveis. Com isso, pretende-se auxiliar o analista de qualidade na identificação de quais atributos estão impactando a confiança do usuário no funcionamento desse tipo de sistema e, em consequência, que os

---

<sup>1</sup> Nesta pesquisa, o conceito *Dependability* é traduzido como *Confiança no Funcionamento*, baseado em discussões de pesquisadores da área de Engenharia de Software que tentaram chegar a um consenso sobre a tradução correta do termo para o português (em <http://www.cs.kent.ac.uk/people/staff/rdl/CoF>).

desenvolvedores usem os resultados dessas avaliações para melhorar a qualidade dessas aplicações, aumentando a aceitação e utilização das mesmas.

Para avaliar o conjunto de medidas apresentado, três aplicações foram utilizadas nos estudos de casos<sup>2</sup>: um guia de visitas móveis, um sistema para controle de volume de chamada do celular e um sistema de detecção e alerta de quedas. Sete usuários foram convidados a utilizar estas aplicações e as informações necessárias para a coleta das medidas foram extraídas e armazenadas com o uso de questionários e logs de interação. Os resultados dos estudos de caso identificam e apresentam pontos de melhoria nestes sistemas. Além disso, de acordo com relatos dos usuários, é possível levantar evidências de que a coleta das medidas aumenta o nível de confiança do usuário em relação ao funcionamento das aplicações avaliadas.

## 2. Medidas Existentes para a Avaliação da Confiança no Funcionamento

Dentre os trabalhos apresentados na literatura, é possível encontrar medidas de qualidade de software para a avaliação da confiança (neste caso, consideramos confiança, traduzida do inglês *trust*, e confiança no funcionamento, traduzida do inglês *dependability*)<sup>3</sup> em (SCHOLTZ e CONSOLVO 2004), (RANGANATHAN 2005), (SANTOS et al 2013), (LEE e YUN 2012) e (ISO 25010). Essas medidas foram omitidas aqui por falta de espaço, mas podem ser encontradas no trabalho de (FERREIRA 2015).

Os principais problemas relacionados à maioria das medidas encontradas nesses trabalhos são que os mesmos desconsideram as peculiaridades inerentes aos sistemas ubíquos, mencionadas na Seção 1 ou, quando consideram, por exemplo, a taxa de erros de contexto inferidos em (RANGANATHAN 2005), não utilizam um formato padrão para a definição da medida (i.e., medidas sem função de medição ou método de coleta).

Dessa forma, utilizando as medidas propostas pelos autores citados anteriormente, pode não ser possível obter uma medição repetível, pois a falta de função de medição possibilita resultados com um nível de abstração e ambiguidade significativos.

É importante também mencionar que há divergências na classificação que cada um dos trabalhos apresenta quanto aos atributos considerados como impactantes para a avaliação da confiança. De acordo com (SCHOLTZ e CONSOLVO 2004) e (SANTOS et al 2013), a confiança pode ser avaliada considerando questões relativas à privacidade, controle das informações pessoais do usuário e consciência do usuário sobre o uso de suas informações por parte do sistema.

---

<sup>2</sup> Os estudos de caso para avaliação das medidas deste trabalho são experimentos, usando as questões da metodologia GQM e não seguem o rigor dos estudos de casos da engenharia de software experimental.

<sup>3</sup> Não faz parte deste trabalho discutir sobre as controvérsias geradas por parte da literatura, por exemplo, se confiança no funcionamento (*dependability*) é sinônimo de confiança (*trust*) (FOGG e TSENG 1999) (GOILAU 2003) (SOLNER 2012), mas sim fazer uso dos conceitos apresentados no trabalho de (AVIZIENIS et al 2004) para a área de qualidade de software e medições.

Essas mesmas limitações podem ser encontradas no trabalho de (RANGANATHAN 2005) e (LEE e YUN 2012), porém, nesse caso, o foco das medidas definidas está na confiabilidade e proteção, desconsiderando os demais atributos. Além disso, segundo o autor, o conjunto de medidas definidas em seu trabalho auxilia a avaliação da confiança no funcionamento, porém, não é o suficiente para o domínio dos sistemas ubíquos, acarretando na necessidade da definição de outras medidas.

Por fim, apesar de ser possível encontrar na (ISO 25010) medidas para a avaliação da qualidade em uso (i.e., medidas associadas a visão do usuário sobre o produto de software quando utilizado em ambiente e contexto de uso estabelecido pelo avaliador), estas não levam em consideração características particulares dos sistemas ubíquos (e.g., sensibilidade ao contexto).

Uma análise comparativa desses trabalhos pode ser encontrada, de forma sintetizada, na Tabela 1, levando em consideração os principais diferenciais do nosso trabalho.

**Tabela 1: Comparação dos Trabalhos Relacionados**

	(SCHOLTZ e CONSOLVO 2004)	(RANGANATHAN 2005)	(SANTOS et al 2013)	(LEE e YUN 2012)	(ISO 25010)
Considera o domínio dos sistemas ubíquos	SIM	SIM	SIM	SIM	NÃO
Avalia por meio de medidas de qualidade	NÃO, apenas há medidas e um framework de características	SIM	SIM	SIM	SIM
As medidas são definidas seguindo formato padrão	NÃO	NÃO	SIM	NÃO	SIM
Considera a disponibilidade como fator de impacto na confiança	NÃO	NÃO	NÃO	NÃO	NÃO
Considera a integridade como fator de impacto na confiança	NÃO	SIM	NÃO	NÃO	NÃO
Considera a manutenibilidade como fator de impacto na confiança	NÃO	NÃO	NÃO	NÃO	NÃO
Considera a proteção como fator de impacto na confiança	SIM	SIM	SIM	SIM	SIM
Considera a confiabilidade como fator de impacto na confiança	NÃO	SIM	NÃO	SIM	NÃO

Conforme observado na Tabela 1, não existem medidas para avaliar a confiança e todos os seus atributos ou fatores de impacto em sistemas ubíquos. É possível encontrar medidas que ainda estão incompletas (e.g., não possuem qualquer apoio na sua coleta ou

interpretação) ou não avaliam todos os atributos da característica confiança. Diante desse cenário, este trabalho apresenta medidas que permitem coleta, interpretação e repetição (caso haja necessidade).

### 3. O GQM e as Medidas Propostas

De acordo com (ROCHA et al 2012), ao definir medidas de qualidade de software é necessário ter clareza sobre a adequação e utilidade destas, para que seja possível garantir que as informações obtidas com a sua coleta apoiem a tomada de decisão, respeitando os objetivos definidos para o sistema que está sendo avaliado. Para tanto, segundo (SANTOS 2014), é possível utilizar diferentes metodologias para a definição de medidas, dentre elas o GQM, o GQiM e o PSM (BASILI et al 1994) (HUGHES 2000).

O *Goal-Question-Metric* (GQM) é uma abordagem orientada a objetivos para a mensuração de produtos e processos de software. Essa abordagem é baseada na premissa de que as medições de software devem ser orientadas aos objetivos de medição, que por sua vez devem ser orientados aos objetivos da organização (BASILI et al 1994).

Devido à dificuldade em se derivar objetivos de medição no GQM, surgiu, como auxílio, a abordagem *Goal-Question-Indicator-Metric* (GQiM). A metodologia é composta por dez passos e procura alcançar os objetivos de negócio preparando um plano de medições. O *Practical Software Measurement* (PSM), por sua vez, é uma abordagem para especificar formalmente medidas de qualidade de software e como aplicá-las em um processo de medição (HUGHES 2000).

#### 3.1. Etapas da Metodologia GQM

Neste trabalho, a metodologia utilizada para a definição de medidas é o GQM. A escolha dessa metodologia é atribuída a experiência que o nosso grupo de pesquisa já possuía no seu uso bem como a sua ampla utilização na comunidade para definição de medidas de qualidade.

Segundo (BASILI et al 1994), a metodologia GQM é dividida em quatro etapas: planejamento, definição, coleta de dados e interpretação de resultados.

Na etapa de planejamento é necessário definir a equipe que participará do GQM, a área de melhoria e os projetos que farão parte da aplicação do método. Para esta pesquisa, a equipe GQM foi definida de acordo com as áreas de interesse desse trabalho. Foram selecionados então especialistas com conhecimento nas áreas de sistemas ubíquos, qualidade de *software* e confiança em sistemas de *software*. Um participante no nosso trabalho é considerado especialista se possui mestrado (completo ou em andamento) ou doutorado (completo ou em andamento) em pelo menos uma das áreas de interesse. No total, oito especialistas participaram (2 mestrandos e 3 doutorandos na área de sistemas ubíquos, 2 mestres e 1 doutorando na área de segurança e confiança).

Para cada especialista, as atividades realizadas foram: (i) apresentação do problema, motivação e objetivos da pesquisa; (ii) análise da lista de medidas

encontradas na literatura (ou propostas com base nas medidas encontradas); e, (iii) seleção de medidas, adaptação e/ou sugestões de medidas por parte dos especialistas.

Para cada foco de qualidade apresentado foi utilizado um abstract sheet (Figura 1) com o intuito de facilitar as entrevistas e documentação das medidas.

<b>Foco de qualidade</b> Qual é o conhecimento atual dos especialistas com respeito ao foco de qualidade?	<b>Fatores de variação</b> Quais fatores dos sistemas ubíquos os especialistas acham que influenciam esse fator?
<b>Hipóteses de baseline</b> Quais são as possíveis medidas para mensurar o objeto de estudo de acordo com os especialistas? Qual é o conhecimento atual dos especialistas com respeito a essas medidas?	<b>Impacto nas hipóteses de baseline</b> Quais tipos de dependências existem entre as medidas e os fatores de variação?

**Figura 1: Abstract Sheet**

Na fase de definição, por sua vez, é preciso elucidar os objetivos do GQM, as questões a serem respondidas, além de definir e refinar as medidas (as quais serão apresentadas na Sub Seção 3.1). No caso deste trabalho de pesquisa, o objeto sob medição é o sistema ubíquo para dispositivo móvel, o propósito é avaliar e os focos de qualidade são os cinco atributos de qualidade relacionados à confiança no funcionamento. Por fim, o ponto de vista é o do usuário e o do desenvolvedor.

Assim como indica (BASILI et al 1994), após a definição do objetivo de medição, é necessário realizar as entrevistas GQM com o time estabelecido na fase de planejamento. Essas entrevistas são fundamentais para identificar quais são as questões relevantes, com respeito ao objetivo, e quais as possíveis medidas capazes de respondê-las.

Com base nas informações obtidas com a revisão da literatura e com as entrevistas GQM<sup>4</sup> foi possível derivar as seguintes questões:

Q1.1- Qual o grau de disponibilidade do serviço provido pelo sistema?

Q1.2 - Qual o grau de integridade dos dados armazenados ou transmitidos pelo sistema?

Q1.3 - Qual o grau de confiabilidade do serviço provido pelo sistema?

Q1.4 - Qual o esforço necessário para realizar manutenção ou evolução do sistema?

---

<sup>4</sup> Por limitações de espaço, não foi possível acrescentar no texto deste artigo os abstract sheets gerados com as entrevistas GQM. Eles estão disponíveis e podem ser encontrados no endereço [https://github.com/GREatPesquisa/Medidas\\_de\\_Qualidade\\_Sistemas\\_Ubiquos](https://github.com/GREatPesquisa/Medidas_de_Qualidade_Sistemas_Ubiquos), no diretório "Dependability", dentro do repositório "Avaliação de Medidas". O mesmo site vale para os questionários e formulários utilizados como métodos de coleta, no repositório "GQM".

Q1.5 – Qual o grau de consciência e controle do usuário sobre suas informações utilizadas pelo sistema?

A questão Q1.1 busca identificar se o sistema está disponível sempre que requisitado. A questão Q1.2, por sua vez, busca identificar a ausência de alterações inadequadas no sistema. A questão Q1.3 busca avaliar se o sistema funciona conforme o especificado. A questão Q1.4 tem por objetivo avaliar o grau de eficácia e eficiência com o qual o sistema pode ser modificado. A questão Q1.5, por sua vez, busca identificar se o usuário é ciente do uso de suas informações pessoais por parte do sistema e se tem controle sobre o uso destas (i.e. subcaracterística proteção, do inglês *safety*). Com isso, é possível obter informações úteis para entender se o sistema pode levar a um estado no qual a vida humana, a saúde, a propriedade ou o ambiente está em perigo e se há proteção da informação contra acessos não autorizados.

Por fim, nas fases de coleta e interpretação é necessário coletar os dados, com base nas medidas definidas, e interpretá-los buscando responder as questões estabelecidas. Neste trabalho, para coletar as medidas estabelecidas, apresentadas na Seção 3.2, foi necessário utilizar três métodos diferentes:

- Coleta Automática: a coleta de dados automática é feita com base em dados obtidos durante o uso da aplicação (em testes) e armazenados em log.
- Análise do Usuário: a análise com o usuário é feita por meio de formulários e questionários e procura avaliar a confiança do usuário durante a utilização do sistema.
- Análise do Desenvolvedor: a análise com o desenvolvedor é feita por meio de formulários e questionários e busca auxiliar o desenvolvedor na construção de um sistema no qual o usuário possa justificadamente confiar.

### 3.2. As Medidas

As medidas apresentadas neste artigo seguem o formato da norma ISO 25000 e fazem parte de um conjunto de vinte e cinco medidas que podem ser encontradas no trabalho (FERREIRA 2015) O subconjunto apresentado neste trabalho é composto por novas medidas em relação ao que é encontrado na literatura. As medidas são compostas por identificador, nome, definição, função de medição, interpretação e método de coleta, conforme apresentada na Tabela 2.

A medida GFB busca avaliar se o serviço é entregue ao usuário conforme o especificado e sempre que requisitado, mesmo quando o dispositivo estiver bloqueado (em descanso). Essa medida retrata aspectos específicos dos sistemas desenvolvidos para dispositivos móveis e deve ser coletada utilizando como método de coleta o questionário de Análise do Usuário.

A medida GCEA busca avaliar aspectos do sistema que possibilitem garantir que o serviço é entregue ao usuário de forma adequada, representando a execução correta dos mecanismos ou regras de adaptação proveniente de mudança de contexto que, por sua vez, implicam em um provimento de serviço conforme o esperado. Essa medida deve ser coletada pelo questionário de Análise do Desenvolvedor.

**Tabela 2: O conjunto de Medidas**

ID	Nome	Função de Medição	Valores de Interpretação
GFB	Grau de Funcionamento em Background	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: o sistema funciona normalmente em background (2) Médio: Algum aspecto do sistema é comprometido quando o dispositivo está bloqueado e a aplicação rodando, consequentemente, em background (1) Baixo: O sistema não funciona em background
GIdB	Grau de Independência da Bateria	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: o sistema funciona conforme o especificado e sempre que necessário em qualquer nível de bateria (2) Médio: Algum aspecto do sistema é comprometido quando o dispositivo está com pouca bateria (menos de 25%) (1) Baixo: O sistema não funciona quando o dispositivo está com pouca bateria (menos de 25%).
GPUD	Grau de Permissão no uso dos Dados	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: o usuário tem controle sobre 75% ou mais dos dados pessoais que são utilizados pelo sistema e pode alterar as permissões a qualquer momento. (2) Médio: existem recursos disponíveis para o usuário controlar e alterar a permissão de alguns (75% - 50%) dos seus dados pessoais que são utilizados pelo sistema. (1) Baixo: Existem recurso disponíveis para o usuário controlar o uso de menos de 50% dos seus dados que são utilizados por parte do sistema.
GICE	Grau de Independência de Componentes Externos	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: O sistema funciona conforme o especificado e sempre que necessário mesmo que qualquer componente externo tenha algum problema. (2) Médio: Algum aspecto do sistema é comprometido quando algum componente externo, facilmente identificável, tem algum problema. (1) Baixo: O sistema não funciona quando qualquer componente externo tem algum problema.
GIDR	Grau de Integridade dos Dados Transmítidos em Rede	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: Implementa técnicas de integridade com verificação e correção. (2) Média: Implementa técnicas de integridade, porém não garante verificação e correção. (1) Baixo: Não implementa técnicas de integridade.
GCUD	Grau de Compreensão no Uso dos Dados	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: o usuário é ciente de mais de 75% das suas informações que são utilizadas pelo sistema. (2) Médio: o usuário é ciente de apenas algumas (75% - 50%) das suas informações que são utilizadas pelo sistema. (1) Baixo: o usuário é ciente de menos de 50% das suas informações que são utilizadas pelo Sistema.
PPS	Presença de Políticas de Segurança	X = A, onde A é (1) Nenhuma (2) Permissão (3) Interação	(1) Nenhuma (2) Presença de Política de Concessão Baseada em Permissão (3) Presença de Política de Interação
GDoc	Grau de Documentação	X = A, onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: Mais de 75% dos componentes estão documentados. (2) Médio: Apenas uma parte (75% - 50%) dos componentes estão documentados. (1) Baixo: Menos de 50% dos componentes estão documentados.

**Tabela 2 (Continuação): O conjunto de Medidas**

ID	Nome	Função de Medição	Valores de Interpretação
GEMS	Grau de Esforço para Realizar Manutenção do Sistema	$X = A$ , onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: O tempo para realizar manutenção é superior em até 25% ao tempo que foi necessário para implantação. (2) Médio: O tempo para realizar manutenção é equivalente ao tempo que foi necessário para implantação. (1) Baixo: O tempo para realizar manutenção é inferior em até 25% ao tempo que foi necessário para implantação.
GEES	Grau de Esforço para Realizar Manutenção ou Evolução do Código	$X = A$ , onde A é (3) Alto (2) Médio (1) Baixo	(3) Alto: O tempo para realizar manutenção ou evolução é superior em até 25% ao tempo que foi necessário para codificação. (2) Médio: O tempo para realizar manutenção ou evolução é equivalente ao tempo que foi necessário para codificação. (1) Baixo: O tempo para realizar manutenção ou evolução é inferior em até 25% ao tempo que foi necessário para codificação.

A medida GIDB tem como objetivo garantir que o serviço é entregue ao usuário conforme o especificado e sempre que requisitado, mesmo quando o dispositivo estiver com bateria baixa. A ideia chave desta medida é avaliar como o sistema se comporta quando a bateria do dispositivo está baixa (menos de 25%). Algumas funções nativas do desenvolvimento para sistemas *Android* (o comando *broadcast*, por exemplo) não são ativadas (ou ativadas em intervalos de tempo maiores do que o adequado) quando o dispositivo está com pouca bateria. Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Usuário.

A medida GInB busca garantir que o serviço é entregue ao usuário conforme o especificado e sempre que requisitado sem comprometer significativamente a bateria do dispositivo (e.g., 15% a mais de consumo quando utilizando o sistema), buscando garantir a característica de mobilidade, inerente aos sistemas ubíquos. Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Usuário.

A medida GPUD busca avaliar se há ou não recursos disponíveis ao usuário para controle sobre suas informações pessoais: privacidade do conteúdo, identidade, localização e outros. Por exemplo, é possível alterar permissões? É possível restringir conteúdo? Quanto maior o número recursos disponíveis mais próximo de ‘Alto’ o valor da medida, quanto menor, mais próximo de ‘Baixo’. Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Usuário.

A medida GICE objetiva garantir que o serviço é entregue ao usuário conforme o especificado e sempre que requisitado independente de componentes externos. Um exemplo de questionamento que essa medida pode ajudar a sanar é: em um sistema que coloca o celular no mudo de acordo com os compromissos da *Google* agenda do usuário, caso ocorra algum problema de indisponibilidade com o sistema de agendas da *google*, o sistema presente no dispositivo, que interage com essa agenda, também deixará de funcionar? Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Usuário.

As medidas GIDR e GIDA objetivam avaliar e garantir que os dados utilizados pelo sistema, que são transmitidos via rede ou armazenados, não sejam acessados por

quem não tem autorização. Ambas as medidas devem ser coletadas utilizando como método de coleta o questionário de Análise do Desenvolvedor.

A medida GCUD busca avaliar se há ou não recursos disponíveis ao usuário para entender e compreender como os dados gravados são utilizados. Por exemplo, em sistemas que utilizam informações sobre a localização do usuário, o usuário é ciente desta utilização? Quanto maior o número de recursos disponíveis mais próximos de ‘Alto’ está o valor da medida, quanto menor, mais próximo de ‘Baixo’. Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Usuário.

A medida PPS busca avaliar a presença de políticas de segurança utilizadas pelo sistema ou pela aplicação. De acordo com (KILINC et al 2012), referência da qual foram retiradas informações para definição desta medida, as políticas de segurança em sistemas ubíquos são: Políticas de Concessão Baseada em Permissão e Políticas de Interação. Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Desenvolvedor.

A medida GDoc tem como objetivo garantir que o sistema desenvolvido possui documentação, auxiliando no seu entendimento e utilização. Essa medida deve ser coletada utilizando como método de coleta o questionário de Análise do Desenvolvedor.

As medidas GEMS e GEES, por sua vez, buscam garantir que o tempo necessário para manutenção ou evolução do sistema e do código sejam adequadas, não implicando em perdas ou atrasos para o provimento do serviço conforme o especificado. Ambas as medidas devem ser coletadas utilizando como método de coleta o questionário de Análise do Desenvolvedor.

#### **4. Avaliação das Medidas**

Com o intuito de avaliar o uso das medidas apresentadas na Seção 3, estudos de casos foram realizados. Esse método de investigação foi escolhido, pois costuma ser utilizado por projetos que utilizam a abordagem GQM e permite a coleta de dados tanto quantitativos como qualitativos. Para tanto, sete participantes, todos graduandos, com boa experiência no uso de smartphones, foram convidados a utilizar as aplicações. Alguns deles, além de atuarem como usuários, também participaram do processo de desenvolvimento dos sistemas avaliados.

Os sistemas considerados para a realização dos estudos de caso foram: o *fAlert*, encontrado em (PIVA et al 2014), o *GREat Mute* apresentado em (SANTOS et al 2013), e o *GREat Tour* presente em (MARINHO et al 2012). Essas aplicações fazem parte de um conjunto de aplicações que vem sendo avaliadas também em outros trabalhos do nosso grupo de pesquisa. Porém, em nenhum outro trabalho, a característica de confiança no funcionamento havia sido avaliada, por isso e pela importância dessa característica para sistemas ubíquos, a escolha dessas aplicações do grupo.

O *fAlert (Fall Alert)* é um protótipo de aplicação de *software*, para *smartphones* com sistema operacional *android*. A aplicação atua monitorando (detectando quedas) de forma inteligente de pessoas que necessitam de cuidados especiais, como idosos e deficientes físicos (PIVA et al 2014).

A GREat Mute consiste em uma aplicação para *Smartphones* com sistema operacional *Android*, capaz de silenciar o dispositivo móvel do usuário em determinadas condições nas quais isso é adequado. Executando em *background*, a aplicação monitora o *Google Calendar* do usuário. Ao encontrar eventos nos quais o usuário considere que não é adequado receber ligações (e.g., reunião), o dispositivo móvel aciona de modo autônomo o modo silencioso do dispositivo.

O GREat Tour, por sua vez, é uma aplicação *Android* que serve como um guia de visitas do laboratório GREat. A aplicação, sensível ao contexto, captura a posição do visitante em tempo real e exibe informações do local onde ele se encontra, bem como realiza adaptações com base nessa localização.

#### 4.1 Resultados Obtidos

Para cada uma das aplicações, os resultados obtidos com a coleta das treze medidas podem ser encontrados na Figura 2.

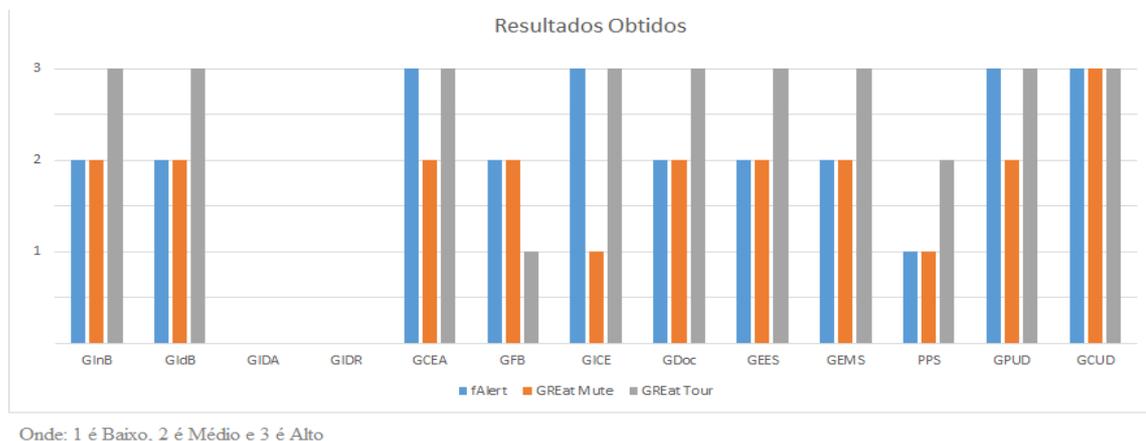


Figura 2: Resultados Obtidos nos Estudos de Caso

Como pode-se observar na Figura 2, a aplicação com melhores resultados foi a *GREat Tour*. Os piores resultados, por sua vez, estão concentrados na aplicação *GREat Mute*. A análise dos resultados obtidos de acordo com cada medida e atributo da confiança no funcionamento encontra-se, em detalhes, na Seção 4.2.

Maiores informações sobre o perfil dos participantes envolvidos nos estudos de caso, bem como os dispositivos utilizados por cada um, podem ser encontradas no endereço [https://github.com/GREatPesquisa/Medidas\\_de\\_Qualidade\\_Sistemas\\_Ubiquos](https://github.com/GREatPesquisa/Medidas_de_Qualidade_Sistemas_Ubiquos), no diretório “Dependability”, dentro do repositório “Avaliação das Medidas”. Nesse mesmo repositório podem ser encontrados também todos os resultados obtidos com a coleta das medidas, para cada uma das aplicações.

#### 4.2 Análise dos Resultados Obtidos

A análise dos resultados tem como principal objetivo utilizar os resultados obtidos com a coleta das medidas para responder as questões definidas na Seção 3. A Tabela 3 apresenta as respostas para cada uma das aplicações.

**Tabela 3: Respostas obtidas com a coleta e análise dos dados**

Aplicação	Q1.1	Q1.2	Q1.3	Q1.4	Q1.5
fAlert	Médio	Ausência de Resultados	Alto	Médio	Alto
GREat Mute	Médio	Ausência de Resultados	Médio	Médio	Alto
GREat Tour	Alto	Ausência de Resultados	Alto	Alto	Alto

As principais medidas utilizadas para responder a questão Q1.1 foram GFB, GIdB e GInB. Com elas, foi possível avaliar a disponibilidade dos três sistemas levando em consideração, principalmente, o seu comportamento em relação a bateria do dispositivo. Para as aplicações *fAlert* e *GREat Mute*, a disponibilidade, segundo os usuários, não é a melhor possível, pois o sistema demora e, por vezes, falha no provimento do serviço quando o nível é inferior a 25%. Além disso, o mau funcionamento em *background* do *GREat Mute* foi detectado com os resultados das medidas, impulsionando o desenvolvimento de uma nova versão do aplicativo, onde esse problema foi solucionado.

Com relação a questão Q1.2, na aplicação *fAlert* não há armazenamento de dados relativos ao processo de detecção. As verificações são feitas de maneira contínua, sem aprendizado ou armazenamento de dados. Logo, não há aplicabilidade para a medida GIDA. O mesmo vale para a medida GIDR. Em relação ao *GREat Mute* e *GREat Tour*, em ambas as aplicações não existe a necessidade da implementação de mecanismos de integridade, devido a existência dos mesmo na pilha TCP-IP, por tratar-se de um serviço web. Logo, as medidas também não foram coletadas nessas aplicações, ocasionando, portanto, em uma ausência de respostas para a questão Q1.2.

Para a questão Q1.3, as principais medidas consideradas foram GCEA, GIdB e GICE. Apesar dos resultados medianos obtidos com a coleta das medidas de disponibilidade, as respostas para as questões relativas à confiabilidade foram animadoras. As aplicações obtiveram resultados próximos do melhor possível, para a maioria das medidas. Isso indica que, mesmo não provendo serviço sempre que requisitado, a conformidade com o esperado, quando não ocorrem problemas de disponibilidade, pode proporcionar o aumento do grau de confiança no funcionamento do sistema.

A questão Q1.4, por sua vez, teve como principais medidas utilizadas a GEES, GEMS e GDoc. A aplicação *GREat Tour* obteve os melhores resultados possíveis para as três medidas porém, as aplicações *GREat Mute* e *fAlert* obtiveram resultados medianos. Segundo os participantes, o principal responsável por esses resultados não tão bons foi a ausência parcial de documentação nos sistemas avaliados. Após esses resultados alcançados com a coleta das medidas, melhorias ocorreram nesse aspecto.

Por fim, para responder a questão Q1.5, as principais medidas utilizadas foram a PPS, GPUD e GCUD. Nessa questão, todas as aplicações obtiveram ótimos resultados. É importante ressaltar que, no *GREat Tour*, além da implementação do mecanismo de login, a aplicação é altamente satisfatória em relação ao uso dos dados do usuário e a permissão para a utilização destes. É importante destacar aqui o uso não apenas das

informações do usuário da aplicação, mas, também, dos ambientes e pessoas envolvidas nele. Uso este não constatado pelas medidas, mas sim pelas pessoas envolvidas no momento da coleta.

Como pode-se observar na Tabela 3, a aplicação *fAlert* obteve como resposta para as questões Q1.1 e Q1.4 o valor “Médio” pois, as medidas utilizadas para responder essas perguntas obtiveram como resultados das funções de medição o valor 2 que, segundo as descrições das medidas, correspondem ao valor de interpretação “Médio”. O mesmo vale para as questões Q1.3 e Q1.5, porém, nesse caso, as medidas coletadas obtiveram como resultado o valor de interpretação “Alto”.

A aplicação *GREat Tour*, por sua vez, obteve como resposta para as questões Q1.1, Q1.3, Q1.4 e Q1.5 o resultado “Alto” pois, as medidas coletadas obtiveram como resultado o valor de interpretação “Alto”. Como exceção, a medida GFB obteve como resultado “Baixo”, porém, esse valor não influenciou na resposta pois, visto que a aplicação exige a interação do usuário no momento do uso (e.g. obriga o usuário a interagir, por meio da leitura de QRCode) esta nunca é utilizada em *background*.

Por fim, para a aplicação *GREat Mute*, as medidas relativas às questões Q1.1, Q1.3 e Q1.4 obtiveram como resultado, durante a coleta, o valor “Médio”. Logo, as respostas para essas questões também são “Médio”. Para a questão Q1.5, a medida GPUD obteve como resultado o valor de interpretação “Médio” porém, como as permissões no uso dos dados estão associadas ao *Google Calendar* do usuário e não ao *GREat Mute*, esse resultado não foi associado a aplicação.

Como limitações dos resultados alcançados, aponta-se o fato de que parte dos participantes envolvidos nos testes relativos aos estudos de caso, foram também desenvolvedores dos sistemas avaliados. Apesar disso, considera-se útil e importante a participação dos mesmos, pois, como mencionado na Seção 3, algumas das medidas apresentadas devem ser coletadas e analisadas sob o ponto de vista do desenvolvedor e não apenas do usuário final. Entretanto, os desenvolvedores que participaram dos testes foram orientados que o objetivo da avaliação estava relacionado com a evolução e melhoria dos sistemas. Logo, resultados positivos e negativos eram igualmente importantes. É relevante mencionar também que, para o sistema *fAlert* não foi possível utilizar nos testes pessoas que fazem parte do público alvo da aplicação.

## 5. Conclusões e Trabalhos Futuros

Objetivando oferecer serviços em todo lugar e a qualquer momento de forma transparente, invisível e natural, os sistemas ubíquos estão cada vez mais presentes no nosso dia-a-dia. Para serem aceitos e utilizados, é aconselhável que esse tipo de sistema possibilite ao usuário confiança no seu funcionamento.

Sabendo que uma das possíveis técnicas para a avaliação dessa característica é a realização de medições (SCHOLTZ e CONSOLVO 2004), (SANTOS et al 2013), esse trabalho apresentou um conjunto de medidas de qualidade de software para a avaliação da confiança no funcionamento de sistemas ubíquos executando em dispositivos móveis.

O conjunto de medidas apresentados neste artigo foi coletado em três estudos de caso e, com a coleta e análise dos resultados obtidos, foi possível proporcionar uma avaliação da confiança no funcionamento desses sistemas. Isso revelou assim, indícios de que as medidas apresentadas são capazes de avaliar características de qualidade em aplicações ubíquas executando em dispositivos móveis.

Entretanto, ainda é preciso investigar desafios que ficaram em aberto, gerando perspectivas para trabalhos futuros, dentre elas, é possível citar: analisar se o conjunto de medidas apresentado pode ser utilizado para a avaliação da confiança no funcionamento em outros tipos de sistemas ubíquos, não apenas os que executam em dispositivos móveis; coletar as medidas em projetos reais de desenvolvimento de software, visto que as medidas foram aplicadas apenas em sistemas desenvolvidos em projetos de pesquisa, não refletindo detalhes particulares e eventuais do processo de desenvolvimento; buscar aplicações onde o atributo integridade possa ser avaliado, visto que nos sistemas utilizados nos estudos de caso a integridade dos dados armazenados ou transmitidos era garantida por outras camadas diferente da de aplicação; e, por fim, realizar um estudo mais aprofundado sobre a característica de confiança no funcionamento (i.e., *dependability*) em comparação a confiança (i.e., *trust*) e, eventualmente, integrar no modelo de características apresentado em (SANTOS et al 2013).

## 7. Agradecimentos

Agradecemos ao time do GREat pelo apoio na realização deste trabalho bem como a FUNCAP (Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico) e CNRS pelo apoio financeiro através do projeto Maximum, de cooperação internacional. Este trabalho é fruto do CACTUS, projeto universal financiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

## Referências

- Basili, V. R., Caldiera, G., e Rombach, H. D., The Goal Question Metric Paradigm, Encyclopedia of Software Engineering, v.2, p 527-532, 1994.
- Debashis S., Amitava M., Pervasive Computing: A Paradigm for the 21st Century, Computer, p. 25-31, 2003.
- Dey A. K. Understanding and Using Context. Personal and Ubiquitous Computing, v.5, n.1, p. 4-7, 2001.
- Ferreira, A. B., Medidas de Qualidade de Software para a Avaliação da Confiança no Funcionamento em Sistemas Ubíquos. Dissertação de Mestrado - Universidade Federal do Ceará, 2015.
- Huges, B., Practical Software Measurement. McGraw - Hill Maidenhead, 2000.
- ISO/IEC 2500, Software Engineering - Software Product Quality Requirements and Evaluation (SQuARE) – Guide to SQuARE, 2005.
- Kilinc C., Booth T., Andersson K., WallDroid: Cloud Assisted Virtualized Application Specific Firewalls for the Android OS, 11th IEEE International Conference on Trust,

- Security and Privacy in Computing and Communications (TrustCom), p. 877–883, 2012.
- Lee, J., Yun, M. H. ., Usability Assessment for Ubiquitous Services: Quantification of the Interactivity in Inter-Personal Services, IEEE International Conference on Management of Innovation & Technology (ICMIT), 2012.
- Marinho, F. G., Andrade, R. M. C., Werner, C., Viana, W., Maia, M. E. F., Rocha, L. S., Teixeira, E., Ferreira, J. B.; Dantas, V. L. L.; Lima, F.; Aguiar, S. , MobiLine: A Nested Software Product Line for the domain of mobile and contextaware applications., Science of Computer Programming, 2012.
- Lima, F. F. de P. ., SysSu - Um Sistema de Suporte para Computação Ubíqua. Dissertação de Mestrado - Universidade Federal do Ceará, 2011.
- Piva, L. S., Ferreira, A. B., Braga, R. B., Andrade, R. M. C. , fAlert: Um Sistema Android para Monitoramento de Quedas em Pessoas com Cuidados Especiais, Workshop de Ferramentas e Aplicações (WFA), XX Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), 2014.
- Ranganathan, A., Al-Muhtadi, J., Biehl, J., Ziebart, B., Campbell, R. H., Bailey, B., Towards a Pervasive Computing Benchmark, 3th IEEE International Conference on Pervasive Computing and Communications Workshops, 2005.
- Santos R. M., Oliveira K. M., Andrade R. M. C., Santos I. S., Lima E. R.; A Quality Model for Human-Computer Interaction Evaluation in Ubiquitous System, 6th Latin American Conference on Human Computer Interaction (CLIHC), 2013.
- Scholtz, J., Consolvo, S., Toward a Framework for Evaluating Ubiquitous Computing Applications. IEEE Pervasive Computing, 2004.